http://icct.iacst.org
ISSN : 2508-6561

The 6th International Conference on Culture Technology 2023

December 1 (Friday) ~ 4 (Monday), 2023
Sunway University, Sunway City, Malaysia

http://icct.iacst.org/

Opening Address

Dr. Rattasit Sukhahuta

General Co-Chair of ICCT2023

It is my great pleasure to extend a warm welcome to each and every one of you gathered here at the 6th International Conference on Culture Technology (ICCT2023) hosted by Sunway University, Malaysia. On behalf of the IACST committee, I would like to express my immense pleasure in welcoming all participants to this esteemed gathering. We are grateful to the dedicated committee members and directors from different countries who have worked tirelessly to bring this event to fruition.

Established in 2015, the International Conference on Culture Technology has emerged as a premier knowledge-sharing, ideas and innovations platform for the culture technology industry. For six years, the ICCT has brought together scholars and practitioners from around the world to exchange ideas and showcase innovations in this dynamic field. This year's conference is a testament to the unwavering commitment of professionals at the forefront of culture technology.

International conferences like ICCT are vital in fostering collaboration, facilitating the sharing of professional knowledge, and connecting top researchers from around the world. They provide a platform for scholars and industry professionals to discuss recent progress, address challenges, and explore opportunities in the ever-evolving realm of culture technology.

Our gratitude extends to the researchers who have submitted a multitude of papers, contributing to the richness and diversity of discussions we anticipate over the coming days. The ICCT has always been a melting pot of cultures, ideas, and perspectives, and this year is no exception.

As we gather here in the vibrant setting of Sunway University, Malaysia, let us embrace the spirit of collaboration and intellectual exchange. Our goal is to bridge the gap between academia and industry, bringing together researchers to share insights, address challenges, and explore the multifaceted aspects of culture technology.

The diverse range of topics to be covered, from cultural applications of technology to the latest advancements in the field, reflects the interdisciplinary nature of our discussions. We invite authors to present original papers and works, encouraging a dialogue that spans the breadth of culture technology and its applications.

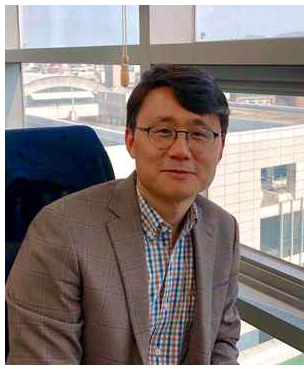
In conclusion, I extend my deepest appreciation to all participants, speakers, and contributors who make this conference a truly global and enriching experience. May the ICCT2023 be a catalyst for new ideas, collaborations, and advancements in the fascinating world of culture technology.

Thank you, and I wish you all a fruitful and inspiring conference.

Dr. Rattasit Sukhahuta

General Co-Chair of ICCT 2023 and President of IACST

TPC Chair Message



Dr. Dongkyun Kim

TPC Chair of ICCT 2023

Greetings from the Technical Program Committee (TPC)! It is our pleasure to unveil a dynamic technical program encompassing a diverse array of topics in culture technology and its applications. Our focus spans across Advanced Technology, Digital Content, Art and Design, and Convergence Technology. This year's program features engaging sessions on cutting-edge subjects like new methods for traffic monitoring, machine intelligence, augmented reality, digital twins, and the metaverse.

The technical program is elevated by our prestigious keynote speaker: Prof. Kwangyun Wohn (Chair Professor at Kunyang University, Korea). He will share his visionary insights on the importance of storytelling.

With a global perspective, we are delighted to announce the acceptance of 36 oral and online papers and 17 posters from 11 different countries (India, Korea, Malaysia, Thailand, China, Pakistan, Malta, Turkey, Japan, UK, USA). The accepted papers were organized into 8 technical oral sessions and a poster session. Besides the paper contribution from all over the world, this successful program was made possible by the devoted service of technical program committee members. We would like to express many thanks to all the TPC members as well as to the Organizing Committee Chairs for their active support and guidance.

We look forward to a conference filled with enlightening discussions, valuable insights, and meaningful connections. Welcome to an event that promises to inspire and advance the frontiers of culture technology. We hope that all of participants enjoy the excellent program of this ICCT 2023.

Dr. Dongkyun Kim

TPC Chair of ICCT 2023 and Chief Vice President of IACST

Organization

Organizing Committee

Honorable Chairs

- **Dr. Sibrandes Poppema**, President of Sunway University, Malaysia
- **Dr. Rattasit Sukhahuta**, President of IACST, Thailand
- **Dr. Pyeoungkee Kim**, CEO of IACST, Korea

Organizing Chairs

- **Dr. Lau Sian Lun**, Sunway University, Malaysia
- **Dr. Dongkyun Kim**, Kyungpook National University, Korea

Registration Chairs

- **Dr. Yong Yoke Leng**, Sunway University, Malaysia
- **Dr. SeHyun Park**, Daegu University, Korea

Local Arrangement Chair

- **Dr. Richard Wong**, Sunway University, Malaysia
- **Wincen Fung**, Sunway University, Malaysia

Publicity Chair

- **Helen Guek**, Sunway University, Malaysia
- **Dr. Bong Hwan Kim**, Daegu Catholic University, Korea

Publication Chair

- **Dr. Chai Wai Chong**, Sunway University, Malaysia
- **Prof. KyungSu Kwon**, Dongseo University, Korea

Technical Program Committee

TPC Chairs

- **Dr. Dongkyun Kim**, Kyungpook National University, Korea
- **Dr. Lee Yun Li**, Sunway University, Malaysia
- **Dr. Nicole Fu**, Sunway University, Malaysia
- **Dr. Sungpil Lee**, Dongseo University, Korea
- **Dr. Hyeyoung Ko**, Seoul Women's University, Korea
- **Dr. Tae Young Byun**, Daegu Catholic University, Korea

Industry Session Committee

- **Mr. Chungkyo Jung**, President of PIPA, Korea

International Advisory Committee

- **Dr. Kwangyun Wohn**, Kunyang University, Korea
- **Dr. Chonggi Kim**, Shanghai University of Science & Technology, China
- **Dr. Tongjin Kim**, Purdue University, USA

Steering Committee

- **Professor, Rattasit Sukhahuta**, Chiang Mai University, IACST
- **Professor, Pyeoungkee Kim**, Silla University, IACST
- **Professor, Dongkyun Kim**, KyungPook National University, IACST
- **Professor, EunYi Kim**, Konkook University, Korea
- **Professor, Guydeuk Yeon**, CHRIST University, India
- **Professor, Intiraporn Mulasastra**, Kasetsart University, Thailand
- **Professor, Rolly Intan**, Petra Christian University, Indonesia
- **CEO, SangHyo Lee**, Namu Edutech, IACST
- **Professor, SeHyun Park**, Daegu University, Korea
- **Principal Researcher, Thepchai Supnithi**, NECTEC, Thailand
- **Professor, YongUk Lee**, Tokyo Polytechnic University, Japan
- **Chief, Adam Chee**, Smart Health Leadership Centre, National University of Singapore, Singapore
- **President, Bill Chen**, Zhejiang Creative Textile Industry Research Institute, China
- **Rector, Dam Quang Minh**, Western University, Vietnam
- **Professor, Donghwa Lee**, Daegu University, Korea
- **Professor, Elena Tsomko**, Dongseo University, Korea
- **Professor, Fei Hao**, Shanxi Normal University, China
- **Professor, Hisaki Nate**, Tokyo Polytechnic University, Japan
- **Professor, Hyeyoung Ko**, Seoul Women's University, Korea
- **Professor, Hyunjin Chun**, Nanjing University of Aeronautics and Astronautics, China
- **Professor, I Putu Agung Bayupati**, Udayana University, Indonesia
- **Professor, Jaeho Pyeon**, San Jose State University, USA
- **Professor, Jiman Hong**, Soongsil University, Korea
- **Professor, Kiesu Kim**, Silla University, Korea
- **Professor, KyungSu Kwon**, Dongseo University, IACST
- **Professor, Lee Yun Li**, Sunway University, Malaysia
- **Professor, Mustafa Eren Yildirim**, Bahcesehir University, Turkey
- **Professor, Seungpok Choi**, Silla University, Korea
- **Professor, Sian Lun Lau**, Sunway University, Malaysia
- **Professor, Sungpil Lee**, Dongseo University, Korea
- **Professor, Ted Shin**, Denver Metropolitan University, USA
- **Professor, Xia Yingchong**, Zhijiang College of Zhejiang University of Technology, China
- **Professor, Yulia M. Kom**, Petra Christian University, Indonesia
- **Professor, Soonki Jung**, Kyungpook National University, Korea
- **Professor, Weeraphan Chanhom**, Chiang Mai University, Thailand

Keynote Speech

Dr. Kwangyun Wohn

Chair Professor, Kunyang University, Korea
Honorary Professor, KAIST, Korea

A Brief History of Storytelling

Abstracts

One of the differences between humans and other animals is that we humans create fictional stories and tell them in a way that persuades or motivates others. We call this storytelling. What's interesting is that storytelling took different forms depending on the era.

This keynote lecture begins with a look at the evolution of storytelling over time. The important point here is that storytelling is greatly influenced not only by the world view (i.e. cultural manifestation) of the time, but also by the media that conveys the story. Additionally, media is determined by the new technologies of that era. In the end, the development history of storytelling is in line with the development history of worldview and technology.

Keywords: Storytelling, media, technologies.

Biography

Kwangyun Wohn is an honorary professor of KAIST, Korean and served as a Chairperson of National Research Council of Science and Technology (NST in short). NST is the governing organization that oversees the 25 major national research institutes, including KIST, ETRI, and KAERI. Before he accepted the Chairship of NST, he has been at KAIST as professor for 27 years. He started his professional career at Agency of Defense Development in Korea. Having completed the master's and Ph.D. studies at University of Wisconsin and University of Maryland, respectively, he had been with Harvard University (USA) as Lecturer, and with University of Pennsylvania (USA) as Assistant Professor. Having returned back to his home country, Korea, he had been with Computer Science Department for fifteen years, and founded a new graduate school, Graduate School of Culture Technology (GSCT) in 2005, and served as Dean. Having retired from KIAST, he still holds the Professor Emeritus of KAIST. Major activities and accomplishments include: Director of Virtual Reality Research Center which is a national center of research excellence, Founding President of Korean Society of Human-Computer Interaction (HCI), Founding President of Korean Society of Performing Art, and Editorial Board of British Computer Society. While his research interests span a broad range of the intersection between art and science – from theoretical aspects to practicalities – he focuses his research efforts to the ap-plication of virtual reality technology to various cultural artifacts such as stage performances, museum exhibitions and educational contents.

Dr. Tong Jin Kim

Professor of Industrial Design, Purdue University, USA

Technology Integration in Design Education

Abstracts

In the swiftly evolving realm of technology, we observe significant shifts in various industries and societies. These changes introduce fresh experiences to our community and prompt us to integrate them into pre-existing economic and cultural systems rapidly.

Responding to this dynamic landscape, there is a growing acknowledgment of the necessity for innovative educational changes. Numerous scholars are actively exploring the fusion of new technologies with educational methods. In this study, our endeavors focused on applying diverse artificial intelligence platforms to traditional design processes, with the aim of uncovering more efficient and imaginative design methodologies. Building upon the insights derived from this research, we have documented a novel educational approach and established a website to disseminate this information, offering a valuable repository for scholars and students alike.

Keywords: Technology, Artificial intelligence, innovative education, education platform.

Biography

Tong Jin Kim is a full professor of industrial design at Purdue University. Before joining academia, he worked as a design leader for consultancies and corporations for more than 18 years, delivering leading-edge products and business solutions, designing over 300 products, and holding more than 70 US patents. Also, he has won over 30 awards, including design competitions, business plan competitions, and international design awards, and the works are published through many magazines, newspapers, and design blogs.

He has encouraged many future generations of innovators to follow the path and make a meaningful contribution to the world, being involved with various universities and design communities, giving lectures on design thinking, and coordinating creative workshops for non-designers and community organizations.

His design experiences and expertise includes, Design Strategy, Design Platforms, and Design Technology.

Schedule

Time	Program
	December 1(Friday) – Day for Friendship Place: Sunway University, Sunway City, Malaysia
14:00~17:00	**Experiencing Local Culture** @Sunway City Area
17:00~18:30	**IACST Regional Director Meeting** (IACST directors and Regular members) TBD @ Sunway University
18:30~20:30	**Reception for Directors** (IACST directors and Regular members) TBD @ Sunway University

	December 2 (Saturday) – Day for Excellence Place: Sunway University, Sunway City, Malaysia
08:30~09:30	**Registration** TBD @ Sunway University
09:30~11:30	**Opening Ceremony & Keynote Speech I** Dr. Kwangyun Wohn Room: JC3 Hall @ Sunway University
11:30~11:40	*Tea Break*
11:40~12:50	**Oral Paper Presentation** Session 1A, Session Room: LT5 Session 1B, Session Room: LT7 @ Sunway University
12:50~13:50	*Lunch Time* - Cafeteria @ Sunway University
13:50~15:00	**Oral Paper Presentation** Session 2A, Session Room: LT5 Session 2B, Session Room: LT7 @ Sunway University
15:00~15:10	*Tea Break*
15:10~16:20	**Oral Paper Presentation** Session 3A, Session Room: LT5 Session 3B, Session Room: LT7 @ Sunway University
16:20~16:30	*Tea Break*
16:30~17:40	**Oral Paper Presentation** Session 4A, Session Room: LT5 Session 4B, Session Room: LT7 @ Sunway University

Time	Program
	December 3 (Sunday) – Creativity & Convergence Place: Sunway University, Sunway City, Malaysia
08:30~09:00	**Registration** TBD @ Sunway University
09:00~09:30	**Setting Up for Poster Paper Presentation** Session Flyer 1F, 2F Session Room: Gallery @ Sunway University
09:30~10:20	**Keynote Speech II** Prof. Tongjin Kim TBD @ Sunway University
10:20~11:10	**Poster Paper Presentation** Session Flyer 1F, Session Room: Gallery @ Sunway University
11:10~11:20	*Tea Break*
11:20~12:30	**Poster Paper Presentation** Session Flyer 2F, Session Room: Gallery @ Sunway University
12:30~13:30	*Lunch Time* - Cafeteria @ Sunway University
14:00~18:00	**Industry Meeting** Session Room *TBD* @Sunway University
	Open Discussion Session Room *TBD* @Sunway University
18:30~ 20:00	**ICCT 2023 Awarding Ceremony** **Banquet & Performance** TBD @ Sunway University

	December 4 (Monday) – Day for Culture Experience Place: Sunway University, Sunway City, Malaysia
09:30~11:30	**OC Meeting** TBD @ Sunway University
11:30~12:30	**Evaluation (by the Organizing Committee)** TBD @ Sunway University

Program

Oral Presentation, 2nd December 2023

Session 1A: Art and Design – Creative Design Innovations

- **Session Chair: Dr. TongJin Kim, Purdue University, USA**
- **Date: December 02, 2023**
- **Room: LT5, Time: 11:40 ~ 12:50**

1A.1: The Application of Peking Opera Elements in Illustration
Xiaotong Zhang, Sojin Kim, Yong-Ki Lee (Dongseo University, Korea)

1A.2: Rethinking Design Education: The Biodesign Approach
Roofia Farheen, Punyotai Thamjamrassri, Yong-Ki Lee (Dongseo University, Korea)

1A.3: Shaping The Future Entertainment and Media Industries of Bangladesh Through OTT Platforms
Nazmus Galib, Sunghoon Cho (Dongseo University, Korea)

1A.4: A Study on Storytelling AI Utilizing Minhwa Elements
Sang-Nam Kim, Young-Suk Lee (Dongguk University, Korea)

Session 1B: Art and Design – Film Analysis and Identity

- **Session Chair: Dr. Guydeuk Yeon, CHRIST University, India**
- **Date: December 02, 2023**
- **Room: LT7, Time: 11:40 ~ 12:50**

1B.1: A Study on Cultural Memory and Identity of Chinese Myth Animation
Yuwen Shao, Hae Yoon Kim (Dongseo University, Korea)

1B.2: Mimetic and poetic visualization of war in film as a means of raising awareness based on the analysis of the visual style of "Come and See"
Olena Lastivka, Sunghoon Cho, Yong-Ki Lee (Dongseo University, Korea)

1B.3: Compressed Character Arcs in Short Film: Exploring Character Development within Time Constraints
Yang Wanye, Sunghoon Cho, Yong-Ki Lee (Dongseo University, Korea)

1B.4: Interdisciplinary Development Trends and Research Hotspots in Digital Humanities and Film: A Bibliometric Analysis from 2000 to 2023
Fang Ru Lin, Hae Yoon Kim (Dongseo University, Korea)

Session 2A: AI-enabled IoT (Safety Technologies)

- **Session Chair: Dr. TaeYoung Byun, Daegu Catholic University, Korea**
- **Date: December 02, 2023**
- **Room: LT5, Time: 13:50 ~ 15:00**

2A.1: Machine Learning-Based Detection of Hazardous Road Objects
Tee Yi Loong, Athirah Mohd Ramly (Sunway University, Malaysia)

2A.2: Unlocking Anomaly Dynamics: A Masked Autoregressive Model and GCN Fusion for Openset Video Anomaly Detection
Trisha Prasad, Harsha B M, Guydeuk Yeon (CHRIST University, India)

2A.3: Intrusion Detection and Prevention System for IoT Networks
Ajay Kumar (Galgotias University, India); Loknath Sai Ambati (Oklahoma City, USA); Kumar Abhishek (National Institute of Technology, India)

2A.4: A Study of the Relationship Between Current Educational Status and Virtual Teacher Appearance Orientation in China
Qianru Liu, Ki-hong Kim (Dongseo University, Korea)

2A.5: Thai Document Recognition using Yolov8
Nattapong Phu-utha, Paphatsara Thiniwong, Wannasa Pomjun (Kasetsart University, Thailand); Pyeoungkee Kim (Silla University, Korea); Somchoke Ruengittinun (Kasetsart University, Thailand)

Session 2B: Connectivity and Communication Technologies

- **Session Chair: Dr. Bonghwan Kim, Daegu Catholic University, Korea**
- **Date: December 02, 2023**
- **Room: LT7, Time: 13:50 ~ 15:00**

2B.1: Enhance Connectivity: RSMA Application and Unveiling the Power of CF-mMIMO Deployment
Ayesha Siddiqa, Mahnoor Ajmal, Malik Muhammad Saad, Muhammad Ashar Tariq, Dongkyun Kim (Kyungpook National University, Korea)

2B.2: Wi-Fi Aware based V2P communication for VRU Safety Service
Donghyun Jeon, Euiri Jo, Seri Park, Junho Seo, Dongkyun Kim (Kyungpook National University, Korea)

2B.3: Predictive Analytics and Performance Optimization in Mobile Networks Using Machine Learning
Shiplu Das (Adamas University West Bengal, India); Rutvij H. Jhaveri (Pandit Deendayal Energy University, India); Buddhadeb Pradhan (University of Engineering and Management West Bengal, India); Rajkumar Singh Rathore (Cardiff Metropolitan University, United Kingdoms); Chunhua Su (University of Aizu Fukushima, China)

2B.4: Performance Optimization of React Applications: A Comparative Analysis of AWS EKS and EC2 Deployments for Enhanced Web Metrics
Haider Rizvi, Asad Mehboob (National Information Technology Board, Pakistan)

Session 3A: AI and Machine Learning Applications

- **Session Chair: Dr. YongJun Jon, Daegu Catholic University, Korea**
- **Date: December 2, 2023**
- **Room: LT5, Time: 15:10 ~ 16:20**

3A.1: Learn to Unlearn: Targeted Unlearning in ML
Muhammad Shaheryar, Junhyeok Jang, Lamyanba Laishram, Jong Take Lee, Soon Ki Jung (Kyungpook National University, Korea)

3A.2: Machine Learning Approaches for Effective Energy Forecasting and Management: Case Study Building in Chiang Mai University, Thailand
Natdanai Keivwath, Pattaraporn Khuwuthyakorn (Chiang Mai University, Thailand)

3A.3: Thai Cyberbullying Classification for Specific Domain using Machine Learning Analysis
Supattanawaree Thipcharoen (The Far Eastern University, Thailand); Teerawich Wongsa (Rajamangala University of Technology Lanna, Thailand)

3A.4: Anomaly Detection in Video through Semi-Supervised Learning with Graph Convolutional Networks
Harsha B M, Guydeuk Yeon (CHRIST University, India)

3A.5: Binary Image Shape Classification by LeNet-5
Mustafa Eren Yildirim (American University of Malta, Malta); Yucel Batu Salman Bahcesehir University, Turkey)

Session 3B: Education and Business Process Management

- **Session Chair: Dr. Thepchai Supnithi, NECTEC, Thailand**
- **Date: December 2, 2023**
- **Room: LT7, Time: 15:10 ~ 16:20**

3B.1: Empowering Education: AI-Driven Customized Learning for Inclusive Education of Differently-Abled Students in India
Anjali Gupta, Guydeuk Yeon (CHRIST University, India)

3B.2: Revolutionizing Rural Education: A Comprehensive Review of AI-Enabled Education Initiatives in India
Anjali Upadhyay, Guydeuk Yeon (CHRIST University, India)

3B.3: The Effect of Institutional Ownership on the Relationship between CEO tenure and Firm Performance
Duangnapa Sukhahuta (Maejo University, Thailand)

3B.4: A Study on the Optimization of LED Light Source and Nutrient solution in Plant Factory Model Using Hydroponic Cultivation Technology
Hyunjin Chun, Jiahao Xu (Nanjing University of Aeronautics and Astronautics, China); Xiaonan Zhan (Jiangsu University, China)

3B.5: Transformative Learning: A Systematic Exploration of AI-Driven Tools in Workplace Education
Saloni Arora, Guydeuk Yeon (CHRIST University, India)

Session 4A: Digital Twin/Metaverse and Smart Technologies

- **Session Chair: Dr. Sian Lun Lau, Sunway University, Malaysia**
- **Date: December 2, 2023**
- **Room: LT5, Time: 16:30 ~ 17:40**

4A.1: Simulating Sensor Data and Modeling Human Activity with Markov Chain for Energy Optimization in Apartments
Malik Muhammad Saad, Muhammad Ashar Tariq, Seri Park, Dongkyun Kim (Kyungpook National University, Korea)

4A.2: Analyzing the development of intelligence in non-player characters in digital games
KaiXing Wang, Kihong Kim (Dongseo University, Korea)

4A.3: Character-based Dataset Generation for Thai Document OCR
Rujisaya Srirattana, Nongnapat Trangarnrueng, Nuttapong Nadee, Kerdphum, Cholathorn Chuto (Kasetsart University, Thailand); Pyeoungkee Kim (Silla University, Korea)

4A.4: School Violence Prevention Through Immersive Behaviour Learning 3D Metaverse Platform
Joanna Michelle, Yong-Ki Lee (Dongseo University, Korea)

4A.5: Integrating 3D Image Generation Techniques into Virtual Reality: A Synthesis of Review and Practical Application
Amirthavarshini V, Harsha B M, Guydeuk Yeon (CHRIST University, India)

Session 4B: Information Retrieval and Data Generation

- **Session Chair: Dr. Lee Yun Li, Sunway University, Malaysia**
- **Date: December 2, 2023**
- **Room: LT7, Time: 16:30 ~ 17:40**

4B.1: Generating a Knowledge Graph-based Infobox for Thai Encyclopedia Historical Reports
Akkharawoot Takhom (Thammasat University, Thailand); Taneth Ruangrajitpakorn, Kanchana Saengthongpattana, Tharathon Utasri, Vorapon Luantangsrisuk, Rattapoom Kedtiwerasak, Thepchai Supnithi (National Science and Technology Development Agency, Thailand)

4B.2: Explainable AI for Image Segmentation: A review on the types of explainability
Teck Wai Chin, Sian Lun Lau (Sunway University, Malaysia)

4B.3: Classification of Wagyu Grade using Deep Learning Methods
Ratanon Treewannakul, Chalothon Chootong, Aurawan Imsombut, Somchoke Ruengittinun (Kasetsart University, Thailand)

4B.4: Catalyzing Efficiency and Effectiveness: How Explainable AI Elevates Customs Brokerage
Supamas Sitisara, Wanchai Rattanawong, Varin Vongmanee (University of the Thai Chamber of Commerce, Thailand)

Poster Presentation, 3rd December 2023

Session 1F: Information Retrieval and Data Generation
- Session Chair: Mr. Somchoke Ruengittinun, Kasetsart University, Thailand
- Date: December 3, 2023
- Room: Gallery, Time: 09:30 ~ 10:50

1F.1: Mascot Image Generation using StyleGAN2-ADA
Je Kyung Lee, Jeoung Gi Kim, Kyung-Ae Cha (Daegu University, Korea)

1F.2: Virtual Idol Design Based on Existing Game Characters - Focus on Character Setting and Personality
Yin Lyu, Yong-Ki Lee (Dongseo University, Korea)

1F.3: Interactive Virtual Museum Exhibit Design
Le Xuan Tan (Sunway University, Malaysia)

1F.4: Analysis of Artistic Concept and Expression in the Film 'Man Jiang Hong'
Chuxuan Wang, HaeYoon Kim, Yong-Ki Lee (Dongseo University, Korea)

1F.5: Portable Monitoring System for Fall Detection using YOLOv7
JaeYeon Byun, YunSeok Yang, BoKyeong Kim, Kyung-Ae Cha, Se-Hyun Park (Daegu University, Korea)

1F.6: Gas Monitoring and Disinfection System in an Agricultural Farm Prototype
Md. Jahirul Islam, Hyeji Lee, Kihak Lee (Daegu Catholic University, Korea); Kihak Lee (SEMS Co. Ltd., Korea); Wolyoung Kim (SOLARLIGHT Korea, Korea); Bonghwan Kim (Daegu Catholic University, Korea)

1F.7: Efficient business process management and achievement of management objectives through the development of a process for the integration of ERP and PMS systems
Junyeop Kim, Dongkyun Kim (Kyungpook National University, Korea)

Session 2F: Information Retrieval and Data Generation

- **Session Chair: Mr. Junyeop Kim, Kyungpook National University, Korea**
- **Date: December 3, 2023**
- **Room: Gallery, Time: 11:00 ~ 12:20**

2F.1: A Study on Hybrid Approach Combining Agriculture and Education through Digital Twin
JinJu Lee, Kwanseon Hong (Dongseo University, Korea)

2F.2: Implementation of Digital Twins in Smart Homes and Generation of Remote Meter Data
Jun Hyeok Jang (Kyungpook National University, Korea); Rinrada Tirasirichai (Kasetsart University, Thailand); Dong Hyeon Kim, Jin Ho Lee, Soon Ki Jung (Kyungpook National University, Korea)

2F.3: Exploring Street view API connection to improve user experience in digital twin environment
Dong Hyeon Kim (Kyungpook National University, Korea); Rinrada Tirasirichai Kasetsart University, Thailand); Jun Hyeok Jang, Hyeon Hoo Hwang, Soon Ki Jung (Kyungpook National University, Korea)

2F.4: PolyGoat: Interactive Manglish Learning Game
Wui Ling Chai, Lee Yunli (Sunway University, Malaysia)

2F.5: Design of a Mobile Application Using Two Numbers for eSIM-based Mobile CRM
Sang-Soo Park (NANOIT Ltd. Inc., Korea); Yong-Jun Jon, Bonghwan Kim, Tae-Young Byun (Daegu Catholic University, Korea)

2F.6: Two Number X: Mobile Application Using eSIM for Life and Work Balance
Kwan-Su Kim, Sang-Soo Park (NANOIT Ltd. Inc., Korea); Bonghwan Kim, Tae-Young Byun (Daegu Catholic University, Korea)

2F.7: A.I. Data Sourcing Platform 'Sullivan RealEyes' based on Public BlockChain
ChanYong Park, YoungMin Im , YoungHeuk Cho (TUAT Corporation, Korea); KyungSu Kwon (Dongseo University, Korea); JinWook Kim (Kyungpook National University, Korea)

2F.8: Comparison of gaze activity between HMD and Theater Screen
Takamura Tamaki, Nate Hisaki (Tokyo Polytechnic University, Japan)

2F.9: Development of a Computer Vision-Based Medical Display Monitoring System
Hyeon Hoo Hwang, Jin Ho Lee, Jun Hyeok Jang, Min Jang, Soon Ki Jung (Kyungpook National University, Korea)

Guide Line for Authors/Chairs

Guideline for Authors

Oral Sessions

1. Duration of the Presentation

The allotted time for each speaker is 12 minutes to present and 6 minutes for Q&A.

2. Equipment in Presentation Room

Each presentation room will have a projector,a screen and a laptop computer running PowerPoint under MS Windows, equipped with USB port.

3. Preparation for Your Presentation Session

Bring a USB memory with your PowerPoint presentation and make sure that your file is copied on the laptop computer before your session starts. Please show up 15 minutes before the actual session starts and introduce yourself to the session chair. Be prepared to give some bibliographic details about yourself to the chairperson so that he/she can introduce you before your presentation.

Poster Sessions

1. Duration of the Presentation

The poster session has 60 minutes, requiring all presenters to be available at their posters during the session.

2. Poster Specification

Posters must be designed to fit a 841mm wide x 1189mm tall board. Posters may be prepared as a single poster or as several smaller sections mounted together. The heading of the poster should list the paper title, author(s) name(s), and affiliation(s).

3. Poster Set-Up

Posters may be attached to the boards by push pins or tapes, which will be provided. Posters must be set up by presenters 10 minutes before the session starts. Posters must be removed by presenters right after the session is over. Posters not removed by 10 minutes after the session will be removed by volunteers (session organizers not responsible for posters left after this deadline).

Guideline for Chairs

Before Your Session

1. Check the Program

Prior to departure for the meeting, check the program on our website (http://icct.iacst.org/index.php) to find the time slot for the session that you are chairing.

2. Pick Up the Materials for Session Chair from Registration Desk

Please arrive at the registration desk about 20minutes prior to the start of the session and pick up the material prepared for a session chair.

3. Check the Meeting Room

Please arrive at the session room about 10 minutes prior to the start of the session and familiarize yourself with the controls for lights, microphones, a pointer, and a projector.. If you encounter problems, immediately alert the session staff who is serving your session in the session room. Meanwhile, you have to check the presence of individual speaker in your session.

During Your Session

1. Introduction

At the start of the session, briefly introduce yourself and explain the timing system to the audience, and as often during the session as you think necessary.

2. Time Allotment

The allotted time for each speaker is 12 minutes to present and 6 minutes for Q&A. If possible, you may give a brief introduction of the speaker to the audience, including his or her affiliation and position, at the beginning of each presentation.

3. Absent Speakers

Should a speaker fail to appear, you may recess the session until it is time for the next scheduled abstract. If you are notified of the absence of any speaker before the session starts, please announce it to the audience. You have to report the absence of any speaker to the secretariat for conference administration at the registration desk.

Oral Presentation (Session 1A ~ 4B)

2nd December 2023

Session 1A: Art and Design - Creative Design Innovations
- **Session Chair: Dr. TongJin Kim, Purdue University, USA**
- **Date: December 02, 2023**
- **Room: LT5, Time: 11:40 ~ 12:50**

Session 1B: Art and Design - Film Analysis and Identity
- **Session Chair: Dr. Guydeuk Yeon, CHRIST University, India**
- **Date: December 02, 2023**
- **Room: LT7, Time: 11:40 ~ 12:50**

Session 2A: AI-enabled IoT (Safety Technologies)
- **Session Chair: Dr. TaeYoung Byun, Daegu Catholic University, Korea**
- **Date: December 02, 2023**
- **Room: LT5, Time: 13:50 ~ 15:00**

Session 2B: Connectivity and Communication Technologies
- **Session Chair: Dr. Bonghwan Kim, Daegu Catholic University, Korea**
- **Date: December 02, 2023**
- **Room: LT7, Time: 13:50 ~ 15:00**

Session 3A: AI and Machine Learning Applications
- **Session Chair: Dr. YongJun Jon, Daegu Catholic University, Korea**
- **Date: December 2, 2023**
- **Room: LT5, Time: 15:10 ~ 16:20**

Session 3B: Education and Business Process Management
- **Session Chair: Dr. Thepchai Supnithi, NECTEC, Thailand**
- **Date: December 2, 2023**
- **Room: LT7, Time: 15:10 ~ 16:20**

Session 4A: Digital Twin/Metaverse and Smart Technologies
- **Session Chair: Dr. Sian Lun Lau, Sunway University, Malaysia**
- **Date: December 2, 2023**
- **Room: LT5, Time: 16:30 ~ 17:40**

Session 4B: Information Retrieval and Data Generation
- **Session Chair: Dr. Lee Yun Li, Sunway University, Malaysia**
- **Date: December 2, 2023**
- **Room: LT7, Time: 16:30 ~ 17:40**

The Application of Peking Opera Elements in Illustration

Xiaotong Zhang
dept. of Visual Content
Dongseo University
Busan, South Korea
lumiaomiao233@gmail.com

Sojin Kim
dept. of Visual Content
Dongseo University
Busan, South Korea
sojinkim@gdsu.dongseo.ac.kr

Yong-Ki Lee
dept. of Design
Dongseo University
Busan, South Korea
ykl77@gdsu.dongseo.ac.kr

***Abstract*—Peking Opera, a cornerstone of China's rich traditional culture and a revered intangible cultural heritage, faces diminishing attention in the wake of societal development. This paper endeavors to address the imperative need for safeguarding and revitalizing this cherished aspect of Chinese heritage. The study delves into the incorporation of Peking Opera elements within illustration design, employing an analysis of creative instances and an exploration of pivotal factors in their integration. The findings underscore the merits of infusing Peking Opera elements into illustration design, shedding light on future prospects for its continued development and preservation. This research underscores the significance of bridging tradition and contemporary creativity to ensure the enduring relevance of Peking Opera in modern society.**

Keywords—Peking Opera, Illustration, Traditional Culture, Element Application

I. INTRODUCTION

Peking Opera, a renowned emblem of China's traditional theatrical heritage, traces its origins to the Peking region. Its nomenclature, "Peking Opera," was officially embraced during the second year of the Guangxu reign in 1876, and it has since acquired the status of an esteemed art form, recognized as the "national quintessence" of Chinese opera.

At its core, Peking Opera showcases a distinctive performance style distinguished by four fundamental elements: singing, speaking, acting, and combat. The art form's allure extends beyond these pillars, incorporating a rich tapestry of artistic disciplines, including music, dance, ancient literature, traditional martial arts, and acrobatics. In this fusion, Peking Opera stands as a condensed representation of diverse artistic expressions.

However, in the face of modern societal evolution, the traditional art of Peking Opera has experienced a gradual wane in public interest. To revitalize this cultural treasure and introduce it to contemporary audiences, the integration of Peking Opera with illustrations emerges as a promising avenue for promotion and preservation. This paper explores the potential synergy between Peking Opera and illustration, offering insights into how this fusion can reignite interest and appreciation for this cultural gem in the modern era.

II. THE COMBINATION OF ILLUSTRATIONS AND PEKING OPERA ELEMENTS

A. The Applicability of Peking Opera Elements in Illustrations

As a traditional art, Peking Opera has usability in various aspects such as costumes, facial makeup, color matching, and performance postures.

The costumes and shapes of the Peking Opera are very exquisite. Peking Opera costumes are mainly Ming Dynasty costumes, which have been refined and sublimated by artists to form the Peking Opera costumes that people see today. As Peking Opera is a stage performance art, Peking Opera costumes are mostly characterized by bright colors, high purity, strong color contrast, and eye-catching.

Peking opera costumes are very good at using contrast effects, using light and dark, and warm and cold effects to enhance visual impact. For example, contrasting color combinations such as red and black, gold and purple, and blue and white often appear in Peking Opera costumes, making the characters more full and three-dimensional. Therefore, these bright colors and clothing features are also very practical in illustrations.

In addition to costumes and colors, facial makeup is also very useful in illustrations. Peking opera facial makeup is a unique makeup art of Peking opera. From the perspective of art, it is patterned; from the perspective of opera, it reflects the character's personality. The function of Peking Opera's facial makeup is to show the characters' character, morality, and character traits. This kind of expression uses the characters as a carrier and spreads the good and evil views and moral values of traditional Chinese culture to the audience through clear and solidified visual symbols. and values[1]. Yellow, pink, gray, and gold and silver. Among them, red represents loyalty and courage, black represents loyalty and virtue, and white represents cruelty, ungratefulness, ambition, and domineering.

In addition to color, Peking Opera facial makeup art also uses other surface features such as face shape and facial features to express the nature and inner world of drama characters. These facial makeups of various shapes can provide many elements of support in terms of illustrations and are very usable.

[1] L. Guanxing, "Research on visual expression of traditional peking opera themes——taking Peking Opera illustrations from the late Qing dynasty to the end of the 20th century as an example". Opera Art, vol.43(03), pp.104-115, 2022.

Fig. 1. Peking opera face makeup (2023.09.26)
Source: Image. https://huaban.com/pins/2915879239

B. The advantages of illustrations in conveying emotions and stories

As a historically significant art form, illustration possesses its unique forms of expression, research value, and aesthetic worth. The advent of digitalization has made illustration creation more efficient, expanding the creative possibilities a means to convey emotions and information, leaving viewers with highly intuitive experiences. The composition of an illustration greatly influences its emotional portrayal. There are various forms of composition in illustration, such as centered, surrounding, and symmetrical compositions, which are the main composition styles employed in contemporary illustration.

Fig. 2. Horizontal composition illustration (2023.9.26)
Source: Image. https://huaban.com/search?q=%E6%9E%84%E5%9B%BE

Fig. 3. Slash composition illustration(2023.9.26)
Source:Image.
https://huaban.com/search?q=%E5%AF%B9%E8%A7%92%E7%BA%BF%E6%9E%84%E5%9B%BE

Different compositions evoke different emotional responses. For example, in Fig 2, a horizontal composition is employed, which gives a sense of solemnity and stability, particularly suited for portraying flat and expansive scenes, effectively reducing visual agitation. In Fig 3, the diagonal composition creates a sense of instability and dynamism, conveying a feeling of speed. Therefore, illustration is not merely a technical skill, but also a means of emotional expression.

In addition to composition, the use of color in illustration can also convey different emotions. Color is a highly influential artistic expression that plays a vital role in illustration design. Colors are symbolic, and different hues can evoke various associations and imaginations. For example, red generally symbolizes passion, power, and love. However, besides these positive connotations, red also carries negative meanings. It can represent danger, warning, violence, and anger. In Fig 4, the red color reflects the passion typically associated with red, while Fig 5 represents the danger symbolized by red. As painting has evolved, colors have acquired subjective interpretations, thus giving them symbolic meanings.

Fig. 4 & 5. Slash composition illustration(2023.9.26)
Source: Image.
https://huaban.com/search?q=%E7%BA%A2%E8%89%B2%E6%8F%92%E7%94%BB

In addition, color is also decorative and common, which can convey cultural value and emotional value, and facilitate people's communication and communication. Therefore, color can reflect the emotions of illustrations, convey the artist's inner world, express the theme of illustrations, and convey the information of illustrations.

Illustrations have an intuitive advantage in emotional expression, allowing viewers to easily receive the central idea of illustrations.

III. Analysis of Works Adding Peking Opera Elements

In recent years, as people pay more attention to traditional culture, works with Peking Opera as the theme or elements of Peking Opera have also emerged. The elements of Peking Opera are like an all-encompassing treasure house of traditional culture. Its colors, shapes, facial makeup, etc. are widely used in illustration design. At the same time, it collides with many different artistic styles and concepts and also produces a unique artistic spark.

For example, in the design re-creation, designers can not only combine different characters with Peking Opera shapes but also combine animals with Peking Opera shapes to make the shape design full of wit and wit. The character image in

the animation "Peking Opera Cat" combines the appearance of Peking Opera with cats, and is very popular among young people. Fig 6 is an illustration poster in the animation "Peking Opera Cat". On the left is the facial makeup of the Overlord in the Peking Opera "Farewell My Concubine", combined with the characteristics of the cat, while retaining the brave and mighty characteristics of the Overlord in Peking Opera. On the right is Yu Ji, who is dancing gracefully in her sleeves, which is also combined with the image of a cat, but retains the softness of the role of Yu Ji in Peking Opera. This combination is intuitive and interesting, allowing people to see the theme of Peking Opera at a glance. By combining Peking Opera with animals, illustration creation can show unique imagination and creativity, and integrate elements of traditional culture and nature.

Fig. 6. Illustration "Overlord's Fold"(2023.9.27)

Source: Image. https://www.so.com/s?ie=utf-8&src=360se7

In addition to the combination with animals, the combination of Peking Opera and cyberpunk styles. Cyberpunk art has the characteristics of high color contrast. Cyberpunk art focuses on the display of high-tech content such as network technology and artificial intelligence. Most of the worldview structures are imaginary future worlds[2]. However, the color characteristics of Peking Opera itself have the characteristics of strong color contrast and bright colors. The collision of these two elements will bring people a different visual experience and give people a new look. Fig 7 is a combination of the cyberpunk style and the female roles in Peking Opera. The costumes of the characters in Peking Opera are used, and the mechanical decoration of future elements is added. At the same time, the use of color is also very clear, with strong contrast. The combination of cyberpunk elements and Peking opera elements increases the visual impact and excitement and conveys a fashionable, bold, and jumping atmosphere. Such avant-garde works can better attract the attention of young people, inject new vitality into traditional culture, and create a dialogue that spans time and space. Express the form and artistic conception of Peking Opera through modern illustration language, so that the audience can feel the commonality between different historical backgrounds and cultures while appreciating the works.

Fig. 7. Illustration - Cyberpunk Peking Opera Huadan (2023.9.27)
Source: Image.
https://huaban.com/search?q=%E8%B5%9B%E5%8D%9A%E8%8A%B1%E6%97%A6

In addition to the direct use of costume modeling elements, some illustrations start with points, lines, and planes, deconstruct Peking Opera elements, and reconstruct and apply them in illustrations. Deconstruction further breaks the original internal structure, extracts and retains its essential features, and then forms new styling elements with visual beauty through artistic techniques, laying the foundation for redesign[3]. Then reconstruct these elements, that is, delete or add other elements based on maintaining the natural state after deconstruction; reorganize, that is, use methods such as superposition, dislocation, rotation, and continuity to achieve spatial transformation and combination for reprocessing.

For example, the 2019 FIBA icon design in Fig 8 uses this method to deconstruct and reconstruct the shape and color of Peking Opera to form a basketball-shaped icon and uses the symbolism of colors in facial makeup to symbolize Wisdom, persistence, strength, and perfection[4]. This kind of design combines the theme of the basketball game and also shows strong Chinese elements, which is an excellent use of Peking Opera elements.

Fig. 8. 2019 FIBA icon design (2023.9.27)

Source: Image.
https://www.google.com/search?q=fiba+world+cup+2019&tbm

[2] Y. Juan, “Innovative application of Peking Opera character modeling in contemporary illustration design”, Chongqing Normal University, p.4, 2012.

[3] L. Sike, C. Yelei, “The application of traditional Peking Opera cultural elements in illustration creation under the background of new media”, Art Education Research, (03), p.086-p.087, 2022.

IV. Problems needing attention when applying Peking Opera elements in illustration design

It should be noted that when applying the traditional elements of Chinese Peking Opera to modern visual communication design, designers must not blindly follow the trend, but should combine the characteristics of the times and use traditional Peking Opera elements flexibly and reasonably. In specific design applications, designers need to pay attention to the following points.

1. Conduct sufficient research on Peking Opera: If you want to use the elements of Peking Opera well, you first need to conduct sufficient research on Peking Opera and learn the relevant knowledge of Peking Opera to avoid misuse and misuse. Master the artistic style and characteristics of Peking Opera for better use.

2. Respect Peking Opera culture: Peking Opera is a traditional Chinese culture. When using Peking Opera elements, we must have some respect for traditional culture, so as not to distort and disrespect Peking Opera culture.

3. Creativity and innovation: When creating illustrations, it is necessary to pay attention to the appropriate innovation of Peking Opera elements, and combine modern art with traditional Peking Opera to create unique illustrations. Don't copy it mechanically. Not all Peking Opera elements in traditional Peking Opera art are suitable for direct application in modern visual communication design works. Modern designers should understand traditional cultural concepts, combine design creativity with Chinese Peking Opera elements, and create contemporary design art vitality. s work.

Applying Peking Opera elements in illustration design is a challenging task, which requires a full understanding and respect of the cultural traditions of Peking Opera, respect for Peking Opera culture, and at the same time incorporating creative and innovative ideas.

V. Summary

The application of Peking Opera elements in illustration design has vast potential for development. As a visual form of expression, illustration can convey emotions and narratives. Peking Opera, with its rich history and cultural connotations, provides abundant elements and inspiration for illustration.

The diverse elements embodied in Peking Opera can play a positive role in the innovative development of modern illustration design. In contemporary illustration design, it is important to carefully balance the traditional Peking Opera elements with the distinctiveness of modern illustration, so that the resulting works can not only reflect the artistic charm of traditional Peking Opera culture but also harmoniously integrate with contemporary artistic trends.

However, it is crucial to respect and preserve the traditional characteristics of Peking Opera when applying its elements. As a traditional art form of China, Peking Opera should be respected for its cultural heritage, accurately portraying character images, costumes, and stage elements, and avoiding any distortion or disrespect towards Peking Opera culture. Designers should embrace a sincere and reverent attitude towards traditional culture while seeking innovation, to create works that possess intellectual and emotional depth as well as cultural cohesion. [4]

There is still untapped potential and challenges in the application of Peking Opera elements in illustration design. Further research and practice in the field of illustration design will contribute to the inheritance and innovation of Peking Opera culture, injecting new vitality and creativity into visual arts. The application of Peking Opera elements in illustrations not only brings aesthetic enjoyment to viewers but also promotes and disseminates traditional culture, attracting more attention and appreciation for Peking Opera.

[1] L. Guanxing, "Research on visual expression of traditional peking opera themes——taking peking opera illustrations from the late qing dynasty to the end of the 20th century as an example". Opera Art, vol.43(03), pp.104-115, 2022.

[2] H. Jing, "The Power of Illustrations—A Study on the Subordination and Independence of Illustrations", China Academy of Art, p.7, 2016.

[3] Y. Juan, "Innovative application of Peking Opera character modeling in contemporary illustration design", Chongqing Normal University, p.4, 2012.

[4] L. Sike, C. Yelei, "The Application of Traditional Peking Opera Cultural Elements in Illustration Creation under the Background of New Media", Art Education Research, (03), pp.86-87, 2022.

[5] Z. Hanchen, Z. Jun, "Research on the Application of Peking Opera Elements in Illustration Design", Art Education Research, (17), pp.124-126, 2022.

[6] L. Jiale, "Exploring the Application of Chinese Traditional Cultural Elements in Visual Communication Design", China National Expo, pp.165-166, 2022.

[7] Z. Yichen, "Research on Overseas Communication of Digital Illustrations of Traditional Chinese Culture", Hunan University of Technology, 2022.

[8] L. Ying, "Research on the Application of Traditional Culture in Illustration Design of National Fashion", Footwear Technology and Design, 2(06), pp.111-113, 2022.

[9] L. Jie, "The application of "deconstruction" and "reconstruction" in the basic teaching of decoration", Art Education Research, (7), pp.122-123, 2016.

[10] H. Fulian, "Creative Application of Peking Opera Elements in Illustration Design", Screen Printing, (02), pp.77-79, 2023.

[4] Z. Hanchen, Z. Jun, "Research on the application of Peking Opera elements in illustration design", Art Education Research, (17), pp.124-126, 2022.

Rethinking Design Education: The Biodesign Approach

Roofia Farheen
Graduate School Department of Design
Dongseo University
Busan, South Korea
roofia.farheen@gmail.com

Punyotai Thamjamrassri
Graduate School Department of Design
Dongseo University
Busan, South Korea
punyotai@gmail.com

Yong-Ki Lee
Graduate School Department of Design
Dongseo University
Busan, South Korea
ykl377@gmail.com

Abstract—Sustainability in design is a significant problem that can be addressed through collaboration between biology and design. Currently, these fields are not well-connected and their potential for synergy remains largely untapped. Designers have traditionally drawn inspiration from nature and its materials. However, a recent trend has emerged where they are incorporating and modifying living organisms in their work. While some activities are taking place in this direction, a solid framework is missing. To achieve sustainability in the field of design, it is crucial to bridge the gap between design and biology and foster collaboration between these disciplines. We are proposing a framework to connect design and biology through service design methods with a focus on problem definition rather than problem-solving. This framework consists of a set of service design methods that guides design teams for the initial step involving defining the overarching purpose of the project and identifying the specific domain within biology where design intervention is required .These methods include various service design tools such as Value Proposition, Persona, Customer Journey Map, Service Blueprint, and UX Research. By using these methods, we acquire valuable insights, establish clear value propositions, understand user requirements, plan user journeys, analyze service processes, and design user experiences. Five design teams utilized our comprehensive framework and presented detailed concepts for their Biodesign projects. Through analysis of their design results, we propose the future direction of Biodesign education. Our contribution lies in proposing stepping stones for an actionable framework that effectively bridges the gap between design principles and the intricate field of biology.

***Keywords*—**Design Education, Biodesign, Service Design, Sustainability

I. INTRODUCTION

In the ever-evolving domain of design and innovation, a profound and consequential transformation is currently unfolding. Historically, designers and artists have traditionally sought inspiration from the natural world, drawing upon its diverse forms, vibrant colors, and intricate patterns to inform and shape their creative endeavors. However, the contemporary landscape of design has witnessed a remarkable and pivotal evolution, that transcends conventional inspiration. In this new era, designers are not merely taking inspiration from nature instead, they are actively engaging in collaboration with the natural world. This monumental paradigm shift signifies the inception of an extraordinary and transformative field known as biodesign.

Biodesign constitutes a synthesis of artistic creativity and biological principles, challenging the demarcation between living organisms and human-created artifacts. It encompasses not only the aesthetic integration of nature into the domain of design but also the deliberate modification and manipulation of living entities to serve creative and utilitarian purposes. This burgeoning discipline is not merely altering our perception of design but is also reshaping our approach to addressing some of the most profound and pressing challenges of our time.

In this context, the role of designers has transcended the confines of aesthetics. They have emerged as catalysts of change, deploying their talents and expertise to confront issues that were once considered external to the traditional realm of design. These multifaceted challenges span an array of domains, encompassing scientific visualization, the development of human-computer interfaces, the exploration of sociological theories, and the ethical ramifications of emerging technologies. Designers now find themselves at the vanguard of multidisciplinary endeavors, pioneering solutions that bridge the gap between artistic creativity and scientific inquiry.

What renders biodesign particularly compelling is its intrinsic commitment to sustainability. In a world characterized by finite resources and a burgeoning global population, the integration of design principles that emulate the innate efficiency and resilience of the natural world has emerged as both an innovative imperative and an ethical necessity. Biodesign transcends the mere creation of aesthetically pleasing and functional objects; it reimagines our relationship with the natural world, inviting us to harness its enduring wisdom to forge a more enlightened and sustainable future.

At its core, biology serves as the bedrock upon which biodesign flourishes. As a venerable scientific discipline, biology provides profound insights into the intricate

This research was supported by BK21 Four Service Design driven Social Innovation Educational Research Team in Dongseo University.

workings of the natural world—ranging from the molecular intricacies of living organisms to the intricate dynamics of the ecosystems that sustain life on Earth. This comprehensive body of knowledge offers a fertile foundation from which designers can draw inspiration, enabling them to craft solutions that are not only innovative and sustainable but also deeply ingrained in the timeless wisdom of nature.

Within the pages of this narrative, we embark upon a journey to explore the dynamic and ever-evolving realm of biodesign a world where creativity, scientific rigor, and unwavering commitment to sustainability harmoniously converge to shape a future.

A. Collaboration Between Design and Biology

Sustainability in Design is a significant problem and must be solved with collaboration. The term "biodesign" specifically refers to the integration of living organisms as essential elements, enhancing the functionality of the final creation. Designers and biologists are increasingly collaborating, driven by a shared purpose [1].

Collaboration between designers and experts is common, with the experts seeking designers' assistance in bridging the gap between theory and practical application. Design has become a means of exploring topics such as society, politics, eroticism, food, and even design itself. Ultimately, it serves as a tool for constructing figurative utopias or metaphors about life. This collaboration opens up thrilling possibilities for design and art projects.

B. The Role of Design in Addressing Environmental Issues

The presence of non-renewable resources or the generation of harmful by-products during usage can pose a threat to the environment through a product. The act of exploring the potential environmental effects of a product and making necessary modifications to the design to minimize or eliminate those impacts is referred to as examining the possible environmental repercussions of a product and refining its design as needed. On the other hand, In order to delve into biology, it is essential to address the living surroundings and the interconnection between living organisms and their environment. The study of biology not only equips students with a vast amount of fundamental knowledge pertaining to the environment, but also nurtures a sense of care towards the environment and cultivates the skills necessary for its preservation. So both these disciplines play a critical role in reducing negative social and environmental impacts, as their decisions significantly influence the impact of the goods and services we consume. The early stages of the design process hold particular importance, as around 80% of a product's environmental impact is determined during this phase [1].

C. Examples of Innovative and Sustainable Biodesign Solutions by Designers

Biodesign is revolutionizing design fields by integrating biological principles and materials. For instance:

1) Fashion: Biodesign is transforming fashion with sustainable fabrics like mycelium and bacterial cellulose, reducing environmental impact. Natural dyes from microorganisms offer eco-friendly alternatives to synthetic chemical dyes. BioCouture explores bacteria-produced, eco-friendly materials for clothing [1].

2) Architecture: Biodesign is enhancing architecture through living walls, green roofs, and innovative materials. BioBricks, made from sand, bacteria, and alternatives, aim to replace traditional bricks. BioConcrete utilizes bacteria for self-repair, potentially extending building lifespans [1].

3) Packaging: Traditional plastic packaging contributes to pollution and waste accumulation. Biodesign offers biodegradable options like mycelium-based MycoComposite, which breaks down harmlessly, reducing plastic waste and its impact on the environment [2].

Biodesign holds promise for sustainable and environmentally friendly solutions across diverse design disciplines. Designers should be educated to become bio-designers because this field empowers them to create innovative, sustainable, and eco-friendly solutions in fashion, architecture, and packaging. Bio-design equips designers with the knowledge and tools to address pressing environmental and social challenges, transforming industries and contributing to a more responsible and sustainable future.

D. Preparing Future Designers to Become Bio designers

So by standing on this ground, introducing and raising the impact of biodesign in the education sector, with the aim of inspiring, introducing, and educating the next generation of designers to become "biodesigners" for a sustainable and ideal future. The goal is to invest in the far future, towards an idealistic world. Learners can then take the initiative to design for a sustainable future, rather than merely solving problems, by working towards their goals rather than the results.

Sustainable product design education holds significant importance throughout the entire process of design learning, particularly during higher education. This stage provides students with essential training and instruction prior to entering their respective professional design fields across diverse industries. While extensive research has been conducted on sustainable product design education and its implementation using various methodologies, our study specifically focuses on the utilization of the Biodesign challenge as a teaching approach for students enrolled in Korean universities.

II. THE BIODESIGN CHALLENGE

A. BDC's Necessity

Making sustainability, bio-design, and other sustainable practices obligatory and primarily offering effective, accessible tools. However, in the case of South Korea, where bio-design is not well-established, the available information primarily centres on biotechnology, a firmly established field. This merging of life sciences and art/design aims to shape the future across various domains and explore ways to cultivate a more sustainable society by employing techniques of biological craftsmanship, commonly known as bio-design.

Unlike biotechnology, which is firmly rooted in biology and employs biological principles to create innovative products, methods, and organisms to enhance human well-

being and advance societal progress, biodesign remains in a nascent stage. Establishing future-oriented education that aligns with learners' initiatives necessitates educators identifying their needs, providing relevant learning resources, designing a curriculum that empowers learners in problem-solving, and fostering learning that equips them with the essential knowledge, skills, attitudes, and values for shaping a sustainable future.

Education for the next generation of designers in sustainability is essential so that they can make sustainable design choices in their respective fields. Bio design is of utmost importance for the future of humanity. However, the concepts and education of bio design is still unfamiliar in the design education. Therefore, suggesting a new educational approach based on bio design, which aims to cultivate a lasting interest in environmental issues and teach problem-solving skills through community cooperation and a multi-dimensional approach for a sustainable future. The aim is to educate students by integrating bio materials and current technology. Based on this design thinking, proposing the implementation of bio design education with the help of Bio Design Challenge BDC.

B. BDC

Daniel Grushkin, a journalist and co-founder of Genspace, the pioneer of community laboratories globally, envisioned the Biodesign Challenge as a contest involving leading design schools. Its aim was to provide art and design students with a platform to imagine future applications of biology. In order to kick start this project and organization, Dan required guidance, connections, and financial support, all within a tight three-month timeframe [3].

Today the outcome of Biodesign Challenge is that it is an education program and competition that is shaping the first generation of biodesigners. High school and university students are partnered with scientist, artists and designers to envision, create and critique transformational applications in biotech [3].

BDC Biodesign Challenge Bridges art, design, and biotech to develop the first generation of professionals who cross disciplines, anticipate promises and pitfalls, and engages the public in dialogues about the broader implications of emerging biotech. BDC'S are threefold:

1. To create a community of collaboration among artists, designers, and biologists.
2. To seed the first generation of biodesigners.
3. To build meaningful public dialogue about biotech and its uses [4].

C. BDC and Service Design

We present a framework designed to facilitate the convergence of design and biology, employing service design methodologies that emphasize problem definition over immediate problem-solving.

This framework encompasses a suite of service design techniques that guide design teams in the initial phase, helping them articulate the overarching project purpose and pinpoint the specific biological domain where design intervention is warranted.

Service design principles can be integrated into programs like the Biodesign Challenge (BDC) to enhance participant experiences and streamline program processes. This integration involves understanding stakeholder needs, mapping user journeys, defining service blueprints, and conducting co-creation workshops. Methods used include user research, persona development, empathy mapping, service prototyping, and user testing. The potential outcomes of integrating service design into BDC include improved participant experiences, streamlined processes, innovative program components, improved communication, and a commitment to continuous improvement. The extent of integration may vary based on program goals and available resources.

By using these methods, we acquire valuable insights, establish clear value propositions, understand user requirements, plan user journeys, analyze service processes, and design user experiences

D. Project Design Outcome

Five design teams utilized our comprehensive framework and explored innovative solutions to pressing environmental and social challenges, presented detailed concepts for their Biodesign projects.

Team 1 emphasized the importance of educating people about biodesign and indoor shoes in Asia. They proposed an educational program centered around mycelium-based indoor shoes, aiming to raise awareness and introduce biodesign concepts. The program, called the Mycelium Indoor Shoe Activity program, encourages creative thinking and problem-solving among students.

Team 2 addressed the issue of excessive oil consumption and pet hair disposal in Korea. They developed pet hair scrubbers capable of absorbing oil, demonstrating their effectiveness in experiments. The idea is to collect pet hair and transform it into scrubbers for oil absorption before dishwashing.

Team 3 focused on sustainable fashion and environmental concerns. They created an eco-friendly material using jellyfish collagen and shell calcium carbonate, envisioning its use as an alternative to various plastic products. This material could replace items like glasses frames, hangers, straws, disposable spoons, and nylon, promoting sustainability.

Team 4 tackled the problem of sea desertification caused by sea urchins. They explored the potential of marine collagen extracted from less commercially viable sea urchins. This collagen has medical applications due to its properties like wound healing and burn treatment.

Team 5 focused on waste-related global issues. They proposed the "Jawa Bag," a sustainable bag made from bio-strings and recycled plastic waste. This bag is not only practical but also carries a strong message of waste reduction and the importance of reusability.

These projects showcase innovative solutions addressing diverse challenges, from education and oil absorption to sustainable materials and waste reduction. They emphasize the potential of biodesign to offer sustainable alternatives and promote environmental consciousness.

III. Future Direction of Biodesign Education

Through examining student's design results, we analyse the BDC and propose the future direction of Biodesign education.

A. Strength and Considerations of BDC for Biodesign Education

Examining the BDC program from the perspective of design education, we derived 6 key strengths and 4 contributions.

This idea seems promising and likely to be effective because it offers a completely different method to achieve a sustainable future. Rather than focusing solely on current practices, we are prioritizing the education of the next generation of designers. This bio design education will be experiential, rather than just theoretical, enabling students to enhance their collaborative skills and prepare them for handling unpredictable situations. Ultimately, students will be able to apply what they learn in their bio design education to their respective areas of activity, By providing them with the right knowledge, they will be equipped to drive change in their fields.

Here are some key points to highlight the strengths of this idea:

- *Interdisciplinary Approach:* This proposal recognizes the importance of bridging the gap between art, design, and biotechnology. This interdisciplinary approach is crucial for addressing complex sustainability challenges that require diverse perspectives and expertise.
- *Community Collaboration:* Emphasizing community cooperation is a practical way to instill problem-solving skills and a sense of responsibility among students. This hands-on approach encourages students to work together and apply their knowledge in real-world scenarios.
- *Experiential Learning:* The focus on experiential learning is vital, as it enables students to gain practical skills and a deeper understanding of bio design concepts. This approach goes beyond theoretical knowledge and prepares students to face real challenges in their future careers.
- *Preparation for Change:* This proposal acknowledges the unpredictable nature of future challenges and aims to prepare students to adapt and drive change in their respective fields. This adaptability is essential in the rapidly evolving landscape of sustainability and biotechnology.
- *Integration of Bio Materials and Technology:* Incorporating bio materials and current technology into the education process aligns with the cutting-edge developments in bio design. It ensures that students are exposed to the latest tools and techniques in the field.
- *Alignment with Bio Design Challenge (BDC):* Leveraging BDC as a platform for implementing bio design education adds credibility and provides students with opportunities for showcasing their work and collaborating with professionals in the field.

However, to successfully implement the proposal, we consider the following steps:

- *Curriculum Development:* Developing a comprehensive and adaptable curriculum that covers both foundational bio design concepts and emerging technologies will be essential.
- *Resource Allocation:* Ensuring access to the necessary resources, such as laboratories, materials, and expert guidance, will be crucial for effective bio design education.
- *Assessment and Evaluation:* Implementing a robust assessment and evaluation system to measure students' progress and the impact of the bio design education program will help in refining the approach over time.
- *Engaging Stakeholders:* Collaborating with educational institutions, industry partners, and government agencies to gain support and resources for your initiative can be instrumental in its success.

The proposal for bio-design education to promote sustainability and interdisciplinary collaboration has the potential for significant impact. It focuses on ideation and interdisciplinary collaboration to generate innovative service design solutions for complex biological challenges. Rigorous problem definition and reframing enable solutions that target root causes, not just surface symptoms.

The program prioritizes user-centricity and stakeholder validation, expected to result in high user satisfaction and, consequently, a more significant impact of the service solutions. Ethical considerations are integrated to ensure effectiveness and ethical responsibility, enhancing trustworthiness and societal benefit.

Continuous learning and adaptability are emphasized, ensuring that the service designs remain flexible and can respond to changing circumstances, emerging technologies, and evolving societal needs, ensuring their longevity and relevance over time.

B. Framework for Bridging Design and Biology

In the perspective of a fashion industry Product Designer, it is emphasized that designers have a crucial role in promoting sustainability through their design choices. Designers are responsible for considering both individual and environmental impacts during the product design process, but this requires proper education and awareness. Traditionally, there has been a lack of focus on environmental and social consequences in favor of meeting customer demands. Designing for sustainability is a creative and uncertain task.
To bridge design and biology, a structured framework is used, which encourages interdisciplinary collaboration and creative thinking. The process involves:

- Defining the project's purpose and identifying the specific biological domain for design intervention.
- Forming a multidisciplinary team to deeply understand the biodesign context through research and stakeholder engagement.
- Framing and reframing the problem by challenging assumptions and exploring it from various angles.
- Ideation and concept generation, with low-fidelity prototypes to test feasibility.
- Validation through feedback and insights from stakeholders and users.
- Service blueprinting to visualize the service's functionality.
- Piloting the concept and refining it based on real-world feedback.
- Measuring impact over time to assess the effectiveness of the solution.
- Documenting the process for knowledge sharing.

This approach prioritizes problem definition, collaboration, and stakeholder engagement to create biology-rooted, user-responsive service design solutions. Sharing the program's processes and outcomes fosters innovation and collaborations, contributing to knowledge in design and biology.

Ultimately, the program aims to have a significant impact on sustainability in various biology-related domains, driven by innovative and adaptable service design solutions. Its success depends on execution, interdisciplinary dedication, and adaptability.

IV. Discussion

From the perspective of a Product Designer in the fashion industry, designers have a pivotal role in shaping a more sustainable future through their choices. Their responsibility extends to decisions that directly impact both individuals and the planet during the product design process. However, this necessitates acquiring essential knowledge, which underscores the importance of education and learning. Traditionally, the focus has been on meeting customer demands with insufficient consideration of the environmental and social consequences of design choices. Designing for unknown variables is complex and requires creativity and a willingness to embrace uncertainty. Early guidance on prioritizing sustainability could have expedited the integration of such principles into design. To bridge design and biology through service design methods, a structured framework is employed. This approach encourages interdisciplinary collaboration and innovative thinking, particularly in the initial stages of problem exploration. The first step involves defining the project's overarching purpose and pinpointing the specific domain within biology needing design intervention, be it healthcare, environmental conservation, or food production.

A multidisciplinary teams are then assembled, comprising different design majors, each offering unique perspectives and skills. The teams immerses itself in the biology-related domain through research, laboratory visits, and expert engagement to gain a profound understanding of biodesign's context. Mapping the ecosystem around the identified problem is another crucial step, involving the identification of stakeholders, users, their pain points, and broader societal and environmental implications. Stakeholder engagement is prioritized to gather insights into their needs, desires, and challenges, utilizing methods like interviews and surveys. The subsequent phase entails substantial time framing and reframing the problem, challenging assumptions, and exploring it from various angles to shift focus from predefined solutions to a well-defined problem. The teams engage in ideation and concept generation, encouraging unconstrained brainstorming. Techniques like design thinking, biomimicry, and ideation workshops are used to generate innovative concepts. Low-fidelity prototypes or conceptual models are created to test the feasibility of service design solutions. Validation of problem statements and proposed solutions involves revisiting stakeholders and users to ensure alignment with their needs and expectations. Iteration on problem definition and concept development is based on feedback and insights gained during validation. Service blueprinting outlines processes, touchpoints, and interactions in delivering the proposed solution, aiding visualization of how the service will function. Piloting the service design concept in a controlled environment collects real-world feedback, leading to refinements before full-scale implementation. To measure impact, adherence to biodesign challenge rules and regulations allows assessment of results' effectiveness in solving the defined problem over time. Finally, documenting the process, insights, and outcomes of the project facilitates knowledge sharing with the broader community, contributing to collective knowledge in design and biology. This comprehensive framework prioritizes problem definition, interdisciplinary collaboration, and stakeholder engagement, ensuring service design solutions are deeply rooted in biology and responsive to user and stakeholder needs. Additionally, documenting the program's processes and outcomes promotes knowledge sharing within the community, fostering a culture of continuous improvement in both design and biology. Depending on the chosen domain within biology, the program holds the potential to make a substantial impact on sustainability efforts, whether in healthcare, environmental conservation, or other areas. This comprehensive approach, grounded in interdisciplinary collaboration and problem definition, positions the program effectively to address complex challenges at the intersection of design and biology. Anticipated results encompass innovative, user-centric, ethically responsible, and adaptable service design solutions with the capacity to create positive transformations across various biology-related domains. The ultimate success of the program depends on its execution, the dedication of interdisciplinary teams, and the ability to adapt to changing circumstances and knowledge.

V. Conclusion

In conclusion, this paper propose a program that connects design and biology through a service design framework, with a keen focus on problem definition, offers a powerful and holistic approach to addressing intricate challenges at the intersection of these disciplines. By fostering interdisciplinary collaboration, encouraging innovative problem exploration, and prioritizing user-centricity, this

program holds the promise of yielding transformative results.

Through ideation, rigorous problem framing, and continuous adaptation, the program is poised to deliver innovative solutions that not only address the symptoms of complex issues but also target their underlying root causes effectively. This approach, coupled with ethical considerations and a commitment to knowledge sharing, ensures that the outcomes are not only ground breaking but also responsible and contribute to the broader community.

Furthermore, the potential to enhance user satisfaction and make a meaningful impact on sustainability efforts within various biology-related domains underscores the program's significance. It aligns with the evolving needs of our society, driven by advancements in both design and biology, and holds the potential to inspire further innovations and collaborations.

As we embark on this journey of problem definition and design exploration, the success of the program ultimately depends on the dedication of interdisciplinary teams, effective execution, and the program's ability to adapt to changing circumstances and knowledge. Yet, with its robust framework and unwavering commitment to innovation and ethical responsibility, this program stands as a beacon of hope for creating positive transformations in our world, where design and biology converge to address some of our most pressing challenges.

While this program presents a comprehensive framework for the convergence of design and biology through service design, it is vital to acknowledge its inherent limitations. These constraints include potential variations in program effectiveness contingent on the chosen biology-related domain, the intricate dynamics of interdisciplinary collaboration and teamwork, and the ongoing necessity to evaluate the program's enduring impact and address scalability challenges. Looking ahead, future research endeavors should encompass in-depth case studies across a spectrum of biology-related domains to assess the program's adaptability, the exploration of strategies to bolster interdisciplinary collaboration, the conduct of comprehensive long-term impact assessments, and a commitment to ensuring the program remains attuned to the ever-evolving landscape of the biodesign field, thus ensuring its continued position as a pioneering force at the intersection of design and biology.

ACKNOWLEDGMENT

I want to express my heartfelt gratitude to those who supported and inspired me throughout this endeavor. My deep gratitude goes to my Professor Yong-Ki and Senior Punyotai Thamjamrassri for their invaluable guidance.

I am immensely grateful to the BK21 program for their multifaceted support. Their financial assistance was instrumental in bringing this project to fruition, and their provision of valuable resources and opportunities significantly enhanced its quality and scope.

REFERENCES

[1] W. Myers, Bio Design nature science creativity, 2nd ed., New York: Museum of Modern Art (MoMA), 2018.

[2] Ecovative, "Ecovative," Ecovative, Newyork, 2007.

[3] Kelley, Nancy J, "Biodesign Challenge," *Nancy J Kelley ,* 2020.

[4] bdc Biodesign Challenge, "Biodesign Challenge," Daniel Grushkin, Newyork , 2015.

[5] W. Myers, Bio Art: Altered Realities, 1st ed., London: Thames & Hudson, 2015.

[6] V. M. Ian de Vere, "Socially responsible design: thinking beyond the triple bottom line to socially responsive and sustainable product design," *International Journal of CoCreation in Design and the Arts,* 2011.

[7] P. Subsoontorn, "Hack Biodesign: An Integrative STEAM Education Platform for Biodesign," *IEEE International Conference on Teaching, Assessment, and Learning for Engineering (TALE),* pp. 1016-1021, 2018.

Shaping The Future Entertainment and Media Industries of Bangladesh Through OTT Platforms

1st Nazmus Galib
Student, Dept. of Visual Contents
Dongseo University, Busan, Korea
galibnazmus@gmail.com

2nd Sunghoon Cho
Professor, Visual Contents
Dongseo University, Busan, Korea
mattcho@gdsu.dongseo.ac.kr

***Abstract*— This conference paper aims to explore the impact of OTT platforms on the entertainment and media sectors in Bangladesh and examine the opportunities and challenges they present. By analyzing the current landscape, trends, and consumer behavior, this paper seeks to provide insights into the potential of OTT platforms to drive growth, foster creativity, and enhance the overall ecosystem of the entertainment and media industries in Bangladesh.**

Keywords— OTT, Bangladeshi entertainment and media industries, Content creator, Audience reach, Market diversity.

I. INTRODUCTION

The entertainment and media industries in Bangladesh have undergone a significant transformation in recent years, driven by technological advancements and changing consumer preferences. One of the key catalysts for this transformation is the emergence of Over-The-Top (OTT) platforms, which have revolutionized the way content is created, distributed, and consumed. As the country embraces the digital age, the potential of OTT platforms to shape the future of the entertainment and media industries in Bangladesh cannot be overstated. Radio emerged as the first major mass communication and entertainment medium in 1939 installing a radio station in Dhaka, Bangladesh. Educating people and promoting culture were the main purposes of radio during that time. In 1960, Television was introduced in Bangladesh in the name of Pakistan Television (PTV). At that time Bangladesh was a part of Pakistan. After the independence of Bangladesh in 1971, Television brought a new wave of visual storytelling and entertainment in the name of Bangladesh Television (BTV). After a few years later Bangladesh entered the satellite television era in 1980. Channels like star TV, MTV, BBC world service began reaching Bangladeshi households, introducing global media contents to the local audience. The film industry in Bangladesh, commonly known as Dhallywood, played a significant role in shaping the country's entertainment landscape. The industry experienced ups and downs over the years, but it has produced numerous successful films, celebrated actors, and iconic directors. The emergence of private television channels, and production houses, and the availability of advanced filmmaking technology have further boosted the growth of the entertainment and media industry. OTT (over the top) is a content-providing channel that provides television productions, films, drama, and web series via the Internet. It is considered as a modern alternative to traditional media distribution methods. In recent years, the global media landscape has witnessed the rapid rise of Over-the-Top (OTT) platforms. These platforms, such as Netflix, Amazon Prime Video, and Disney+, provide on-demand streaming services to users over the internet. OTT platforms have gained immense popularity worldwide due to their convenience, accessibility, and wide range of content choices. They have revolutionized the way people consume entertainment and media by giving individuals the freedom to view content at their preferred time and place. The emergence and growth of OTT platforms hold immense significance for the future of the entertainment and media industries in Bangladesh. These platforms offer an opportunity for the industry to overcome challenges faced by traditional media channels. OTT platforms provide a wider distribution reach, allowing content creators to reach a global audience. They also facilitate content diversity by providing a platform for niche and independent content that may not find a place on traditional platforms. Moreover, OTT platforms enable content creators to experiment with different formats and genres, thereby nurturing creativity and supporting the growth of the local entertainment industry.

II. CURRENT STATE OF ENTERTAINMENT AND MEDIA INDUSTRIES IN BANGLADESH

Television, with its vast reach and popularity, has been a primary source of entertainment for the masses. As per BTRC (Bangladesh Telecommunication Regulatory Commission), from 1997 after the journey of private channels to till now there are 45 private TV channels in Bangladesh. Drama, News, music, documentaries, and sports are the major contents of these channels to entertain the people around the clock. Similarly, radio plays a significant role in reaching out to a wider audience, especially in rural areas. At present time, there are 29 Radio stations including one state-owned station namely 'Bangladesh Betar'. The Bangladesh NGOs Network for Radio and Communication (BNNRC), in special consultative status with the United Nations economic and social councils, considers community radio a special area for intervention. With their help, 19 more radio stations are going on across the country as a part of promoting advocacy to the government in relation to community radio with other organizations. The cinema industry, although faced with challenges, produces a considerable number of films each year, contributing to the cultural and economic growth of the nation. BFDC (Bangladesh Film Development Corporation), widely known as Dhallywood, still plays the most vital role in entertaining the mass people still. Most importantly, Bongo is the first OTT platform launched in 2013 in Bangladesh. Since then, Bangladesh has a variety of OTT video streaming

platforms. Among them, some are international (Netflix, Hulu), and some are national (Bongo, Chorki, bioscope). As per a report in Statista, the number of users of OTT channels in Bangladesh has been increasing significantly. By the year 2026, it will reach 11.3 million.

A. Challenges faced by the industry

Traditional media platforms often struggle with limited content diversity, mainly due to commercial considerations and limitations on available airtime. Moreover, the distribution channels for traditional media, such as terrestrial television and cinema halls, have certain limitations in terms of reach and accessibility. Additionally, traditional media formats may not cater to the evolving demands of audiences who seek content on-demand and personalized viewing experiences.

B. Emergence and growth of OTT platforms in Bangladesh

In recent years, Bangladesh has seen the emergence and growth of several OTT platforms to address these challenges. There are three types of OTT channels in Bangladesh according to their characteristics and origins.

Fig. 1. International OTT platforms

- Foreign platforms - The group of video streaming platforms available in Bangladesh is operated by international service providers like Netflix, Prime Video, and Hulu who do not have an office in the country.
- Semi-foreign platforms - Hoichoi and Binge, a Bangladeshi company operating from Bangladesh but originating from abroad.
- Local platforms - The platforms that are truly Bangladeshi service providers originating and operating from Bangladesh. Chorki, Bioscope, and Bongo are from this category.

Fig. 2. Bangladeshi local OTT platforms

Local platforms like Grameenphone Bioscope, Hoichoi, Bongo BD, and Chorki have gained popularity by offering a variety of content, including movies, web series, and TV shows. These platforms have capitalized on the increasing

Fig. 2. Bangladeshi local OTT platforms

internet penetration and smartphone usage in the country. With their affordable subscriptions and user-friendly interfaces, they have gained a sizable user base, particularly among the youth. The success of these platforms has opened new avenues for content creators and industry stakeholders, fostering the growth of the entertainment and media sectors in Bangladesh.

III. Role and Impact of OTT Platforms on the Entertainment and Media Industries

OTT platforms have played a significant role in revolutionizing the entertainment and media landscape in Bangladesh, leading to the growth and diversification of the industry. Furthermore, OTT platforms have provided a platform for local content creators to showcase their work and have empowered independent filmmakers to reach a global audience.

A. Expansion of audience reach and market opportunities

OTT platforms have played a crucial role in expanding the audience reach and market opportunities for the entertainment and media industries in Bangladesh. With the increasing penetration of smartphones and affordable internet access, OTT platforms have become popular among the younger generation and those residing in remote areas. These platforms have provided access to a wide range of local and international content to a previously untapped audience. As a result, the entertainment and media industries have experienced significant growth, attracted investment, and created employment opportunities.

B. Enabling content diversity and catering to niche markets

Unlike traditional media outlets, OTT platforms are not constrained by time or space limitations. This allows for a wider variety of content, including films, TV shows, documentaries, and web series, to be made available to viewers. Furthermore, OTT platforms have facilitated the production and distribution of content targeting specific demographics, preferences, and interests. This has resulted in the promotion of local culture and the emergence of new genres and storytelling formats.

C. *Transforming consumption patterns and preferences*

Viewers are no longer bound by fixed television schedules and are adopting a more personalized approach to content consumption. The availability of on-demand content has led to a shift towards binge-watching and the desire for instant gratification. Moreover, the data-driven algorithms used by OTT platforms to recommend content have influenced viewers' preferences and choices. Consequently, the entertainment and media industries have had to adapt their production, distribution, and marketing strategies to cater to the evolving demands of viewers.

D. *Empowering local content creators and independent filmmakers*

OTT platforms have provided a platform for local content creators and independent filmmakers in Bangladesh to showcase their work and reach a global audience. Previously, these creators faced numerous obstacles in getting their content distributed and monetized. However, with the advent of OTT platforms, they have been able to bypass traditional gatekeepers and gain direct access to viewers. This has led to an increase in the production of original and innovative content, fostering creativity and entrepreneurship within the industry. Additionally, OTT platforms have made it easier for content creators to monetize their work through revenue-sharing models and direct audience engagement.

IV. Strategies for Shaping the Future of Entertainment and Media Industries through OTT Platforms in Bangladesh

The rise of OTT platforms has disrupted the traditional entertainment and media industries, offering users a wide range of content at their fingertips. In Bangladesh, the potential of these platforms is vast, as they can bridge the gap between content creators and users, provide access to a diverse range of content, and create opportunities for economic growth.

A. *Collaboration between OTT platforms and local content creators*

Collaboration between OTT platforms and local content creators is crucial for creating culturally relevant and diverse content. By partnering with local talents, OTT platforms can ensure the production of high-quality and appealing content for Bangladeshi audiences. They can support through funding, training, and promotional campaigns. This collaboration can also provide a platform for emerging content creators, giving them the opportunity to showcase their work and contribute to the growth of the industry.

B. *Investment in infrastructure and digital connectivity*

According to the Bangladesh Telecommunication Regulatory Commission (BTRC), the number of internet users in Bangladesh reached around 126 million in 2023, indicating a growing potential market for OTT platforms. In this stage, investing in broadband expansion and 4G/5G technology will enhance the user experience and enable smooth streaming of content. Moreover, the development of content delivery networks in rural areas will expand the reach of OTT platforms, making content accessible to a broader audience.

C. *Content regulation and quality control*

To maintain the integrity and credibility of OTT platforms, content regulation, and quality control are essential. Implementing guidelines and standards for content creators will ensure the production of ethical and socially responsible content. Additionally, user feedback mechanisms and content rating systems can help improve the overall quality and relevance of the content.

D. *Building audience engagement and monetization models*

OTT platforms can boost audience engagement by leveraging data analytics and personalization techniques. By understanding user preferences, platforms can provide tailored recommendations, enhancing user satisfaction and retention. According to Statista, the revenue in the Video-on-Demand (VoD) segment in Bangladesh is projected to reach US $16 million in 2021, with an annual growth rate of 6.5%. From this perspective, developing effective monetization models, such as subscription plans and targeted advertising, will enable sustainable revenue generation for both the platforms and content creators.

E. *Addressing legal and copyright issues*

Copyright protection and intellectual property rights play a crucial role in fostering creativity and innovation within the entertainment industry. Implementing strong legal frameworks and enforcing copyright laws will protect the rights of content creators and provide a secure environment for investment. Collaboration between OTT platforms, content creators, and legal authorities is necessary to address these issues effectively.

V. Opportunities and Challenges for the Future of Entertainment and Media Industries in Bangladesh

The future of the entertainment and media industries in Bangladesh holds significant opportunities for growth and expansion. However, addressing infrastructure and technological gaps, competing with international OTT platforms, and finding the right balance between cultural sensitivity and global appeal are essential challenges that need to be overcome. With strategic investments, innovative approaches, and collaborative efforts, Bangladesh can harness its potential to become a vibrant hub for entertainment and media in the region.

A). Growth Prospects and Market Potential

Bangladesh has a population of over 160 million, with a significant portion being young and increasingly connected to digital media platforms. This demographic dividend provides a vast market potential for the entertainment and media industries. According to a report by PwC, the entertainment and media industry in Bangladesh is expected to reach a value of BDT 328.7 billion ($3.9 billion) by 2024.

B). Addressing Infrastructure and Technological Gaps

While the market potential is immense, Bangladesh still faces challenges in terms of infrastructure and technological gaps. The penetration of high-speed internet, particularly in rural areas, remains limited, hindering the growth of digital media consumption. The government and industry stakeholders need to invest in expanding internet connectivity and improving the overall digital infrastructure to unlock the industry's full potential.

C). Competition from International OTT Platforms

Bangladesh's entertainment industry is witnessing increasing competition from international over-the-top (OTT) platforms such as Netflix and Amazon Prime Video. These platforms offer a wide range of content that appeals to global audiences, attracting a significant portion of the country's urban population. As a result, local content creators and broadcasters face the challenge of creating compelling and locally relevant content to retain audiences.

D). Balancing Cultural Sensitivity and Global Appeal

Bangladesh is a country with a rich cultural heritage and traditions. Content creators and broadcasters need to strike a balance between preserving the country's cultural sensitivity and creating content with global appeal. This requires careful consideration of local values, customs, and storytelling techniques while maintaining quality and relevance for global audiences. Collaborations with international production houses and leveraging global distribution networks can facilitate this balancing act.

VI. Conclusion

The entertainment and media industries in Bangladesh have a rich history but face challenges like limited content diversity and distribution channels. The rise of global Over-the-Top (OTT) platforms has brought about significant changes, expanding audience reach and content diversity. The impact of OTT platforms has transformed consumption patterns, empowered local content creators, and offered market opportunities. With strategic collaboration and investment, Bangladesh's entertainment and media landscape can become a hub for innovative content. A robust infrastructure will ensure that even remote areas have access to high-quality OTT services, opening new markets. By striking a balance between local cultural sensitivity and global appeal, Bangladesh can create content that resonates both locally and internationally. In this vision, the country's entertainment and media industry will thrive, serving as a testament to the power of OTT platforms in shaping the future.

Acknowledgment

The authors wish to thank members of the Dongseo University Research Lab and anonymous referees for their helpful comments on earlier drafts of this paper.

References

[1] Ahmed, T. (2020). Impact of OTT platforms on the entertainment industry in Bangladesh. Journal of Media Research, 10(2), 45-62.

[2] Khan, R. (2018). Transforming the media landscape: The rise of OTT platforms. In Proceedings of the International Conference on Media Studies (pp. 123-136). Dhaka: University Press.

[3] Rahman, S. (2019). Empowering local content creators: A case study of OTT platforms in Bangladesh. Journal of Film and Television Studies, 15(4), 78-91.

[4] K. Elissa, “Title of paper if known,” unpublished.

[5] Hossain, S. M. I. (2022). A Study on Over-the-Top (OTT) Video Streaming Platforms in Bangladesh. Advances in Journalism and Communication, 10, 355-376.

[6] PwC (2020). Global Entertainment & Media Outlook 2020-2024. Retrieved from https://www.pwc.com/gx/en/industries/tmt/entertainment-and-media/outlook.html

[7] ITU (2020). Measuring the Information Society Report 2020. Retrieved from https://www.itu.int/en/ITU-D/Statistics/Pages/publications/misr2020.aspx.

[8] McDonald, J. (2020). Bangladesh Broadcasting: From State Control to Market Liberalization. In The International Encyclopedia of Media Studies (pp. 1-19). John Wiley & Sons, Ltd.

[9] Pearce, L. (2021). Mapping media policy change in Bangladesh: the case of film. Media, Culture & Society, 43(1), 105-122.

[10] Bangladesh Telecommunication Regulatory Commission (BTRC) - https://www.btrc.gov.bd/Young, The Technical Writer’s Handbook. Mill Valley, CA: University Science, 1989.

[11] Statista - Video-on-Demand (VoD) segment revenue in Bangladesh https://www.statista.com/outlook/211/100/video-on-demand/bangladesh#market-arpu

A Study on Storytelling AI Utilizing Minhwa Elements

Sang-Nam Kim
Department of Artificial Intelligence
Dongguk University,
Seoul, Rep. of Korea
naminamiya@naver.com

*Young-Suk Lee
Institute of Image and Cultural Contents
Dongguk University,
Seoul, Rep. of Korea
tonacoco@dongguk.edu

***Abstract*—This paper is a study to identify objects appearing in Minhwa and use them to generate sentences to create a Minhwa story. Words for sentences are generated through Minhwa's object extraction, and sentences that may finally be the basis of the story are generated through the positional relationship of the objects. To generate a Korean-style story from an object in Minhwa, we set an inter-object state label for a language generation model. Through this, we will study the Minhwa story generation model in the future.**

Keywords—Minhwa, Minhwa image Objects, story generation, sentence automation, cultural archetype

I. INTRODUCTION

Minhwa is in itself original and atypical. In the late Joseon Dynasty, Minhwa, painted by the nameless public, is not just an expression of still life, but contains various stories. Minwa, which sometimes contains origins and sometimes superstitious ideas, can know the daily lives and interests of people at the time. Therefore, the object in Minhwa is not just an object, but a semantic existence containing symbols and people's aspirations. In addition, Minhwa has originality that is not limited to artistic techniques because it was painted by people who did not go through formal education. Due to these characteristics, Minhwa can be used as a base story to develop other content. However, due to the atypical characteristics of Minhwa, systematic classification has been difficult. However, with the development of AI and image processing technology, object extraction has become easier with the classification of Minhwa. Therefore, in this paper, in order to generate a Minhwa story, we want to identify objects that appear in Minhwa and generate Minhwa object words. I want to use him to create a sentence that is the basis of the story.

• Minhwa: Minhwa is a painting style that appeared in the middle of the Joseon Dynasty and was painted by ordinary people who did not learn how to paint. It was mainly painted on animals and plants that appear in everyday life or stories[1]

II. RELATED STUDIES

A. The prototype value and classification of Minhwa

Minhwa is a work of art created by the people in the late Joseon Dynasty and is a life painting symbolically expressing the life that the people want. Minhwa, the art of the people, is freely expressed under the theme of Korean folk beliefs and customs. One of the characteristics of Minhwa is the independence of the object. The object in the picture contains symbols and stories in itself. This means that each object contained in it, as well as Minhwa, contains a cultural prototype value. In other words, it can be used as a material for the story itself.

Resources for the cultural prototype are classified into cultural and related areas [2]. Among Korea's cultural heritage classification systems, the standard relics system used by the National Museum of Modern and Contemporary Studies is the most authoritative classification system [3]. In accordance with these criteria, research is being conducted to establish a DB by extracting objects appearing in Minhwa, Korea [4][5]. This study intends to utilize the ongoing Minhwa object classification system.

B. Generative AI and visual story creation

Generative AI is currently the hottest issue. Recently, ChatGPT3, which has been upgraded from the previous version, and Dall-E3, an image-generating AI, have been released, providing detailed and accurate results to users' needs. Amazon's Alexa and Google's Bard also show advanced results to users' needs by adding large-scale language model (LLM)-based Generative AI capabilities from simple features.

The visual story generation technology that generates a story using an image among Generative AI generates a story using a series of images [6]. In image learning, expressive learning is achieved through hostile training or reinforcement. The analysis model is given common sense reasoning and semantic association capabilities to contain overall information [7][8][9].

In this study, we intend to study generative AI that deduces information on objects appearing in Minhwa through such image learning and generates new stories through its elements. In addition, we want to design a Korean-style story-generating model by using traditional folktales for learning to generate the story.

III. Creating Sentences Using MINHWA Objects

In order to generate a sentence that is the basis of the story, word generation, which is a component of the sentence, must be preceded. In order to generate words, the object in the image of Minhwa was extracted and a dataset for the information of the object was constructed.

First, the dataset required to classify the Minhwa object follows the classification system created in the previous study [10]. After that, the object of Minhwa was extracted using YOLOv5. The extracted information of the object is classified into the type of the object and the location of the object. This data is defined as a story dataset through inter-object state labeling for the language generation model. The classification of labels to be applied at this time is shown in Table 1.

TABLE Ⅰ. Inter-object state labels for language generation models

Number	1
Type	Tiger, magpie, etc
Location	Top, middle, bottom

The type and location data, which are the result values extracted through the state label, are generated as sentences through the language generation model. KoGPT-2, a Korean natural language processing model, is used as a model to extract story elements. Words and location data, which are information extracted through image processing, are generated as sentences through KoGPT-2. The entire process of the research model is shown in Figure 1.

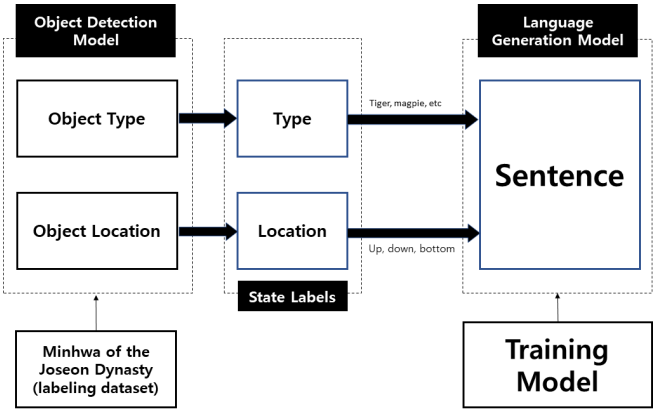


Fig 1. Minhwa Object Sentence Automatic Generation Study Model

Words for sentences are generated through Minhwa's object extraction, and sentences that may finally be the basis of the story are generated through the positional relationship of the objects. The results of learning sample 1 among Minhwa through the Minhwa object sentence automatic generation research model are as follows.

Captions

- This is a Minhwa and it is the Korean folk paintings.
- This painting is a Hojakdo which is one of the Korean folk paintings.
- This is a Hojakdo which is the Painting of Tiger and Magpie and Pine Tree.
- In Hojakdo there are mammals and birds of animals and trees of plants and rocks of inanimate objects.
- The keyword of Hojakdo is a tiger and a magpie and a pine tree and an oddly shaped stone.

Fig 2. Sentence results generated by applying the model of sample 1

Sample 1 is Minhwa featuring tigers and rabbits. Through image object extraction, word data such as rabbits, tigers, Minhwa, flower trees, pine trees, and rocks were extracted. After that, a complete sentence was generated by connecting words and words using the location data of each object. In addition, by combining the extracted words and location information, a number of sentences were generated rather than one sentence.

Through the Minhwa object automatic sentence generation model, it was confirmed that various sentences could be generated using objects in the Minhwa genus. The information of the visualized object is worded through the inter-object state label. The words extracted from the object are automatically generated as basic sentences that may compose a story again through the language generation model. Through this sentence generation model, it is possible to extend to the Minhwa story generation model in the future.

IV. Conclusions

In this paper, we set up an inter-object state label for a language generation model to generate a Korean-style story from Minhwa's object. Through the study of the Minhwa object sentence automatic generation model, visual information could be converted into text information for story composition. However, the amount of data Minhwa can learn is limited and there is a limit to the wordization of visual information. It will find alternatives through future research. In addition, based on this study, we will study a story generation model using Minhwa in the future. Through this, if the Minhwa story can be created, it is thought that the original value of Minhwa can be utilized in various ways. Through this study, it is expected to contribute to the production of various contents using traditional culture.

Acknowledgement. This work was supported by Institute of Information & communications Technology Planning & Evaluation (IITP) under the Artificial Intelligence Convergence Innovation Human Resources Development (IITP-2023-RS-2023-00254592) grant funded by the Korea government(MSIT)

This research was supported by Basic Science Research Program through the National Research Foundation of Korea(NRF) funded by the Ministry of Education, Science and Technology(NRF-2021R1F1A1050452)

REFERENCES

[1] Sunghoon Hong, "A study on the expressivity and formative consciousness of Korean folk painting." Korea National University of Education, 2009..

[2] UNESCO Institute for Statistics, The 2009 UNESCO Framework for Cultural Statistics(FSC) , p.24, 2009.

[3] National Museum of Korea, Standardization of Relics Classification for Computerization of Museums, 1995

[4] Eunjin Choi and Young-Suk Lee, "Designing a Classification System for Minhwa DB," Journal of Korea Multimedia Society, vol. 25, no. 1, pp. 135-143, 2022.

[5] Soorim Yoon and Young-Suk Lee, "A Study on the Classification Model of Minhwa Genre Based on Deep Learning," Journal of Korea Multimedia Society, vol. 25, no. 10, pp. 1524-1534, 2022.

[6] Qiuyuan Huang, Zhe Gan, Asli Celikyilmaz, Dapeng Wu, Jianfeng Wang, and Xiaodong He. "Hierarchically structured reinforcement learning for topically coherent visual story generation", arXiv preprint arXiv:1805.08191, 2018.

[7] Taehyeong Kim, Min-Oh Heo, Seonil Son, Kyoung-Wha Park, and Byoung-Tak Zhang. "Glacnet: Glocal attention cascading networks for multi-imagecued story generation", arXiv preprint arXiv:1805.10973, 2018.

[8] Kyunghyun Cho, Bart van Merrienboer, C¸ aglar G¨ulc¸ehre, Dzmitry Bahdanau, Fethi Bougares, Holger Schwenk, and Yoshua Bengio. "Learning phrase representations using RNN encoder-decoder for statistical machine translation", In EMNLP, pp. 1724–1734, 2014.

[9] P Yang, F Luo, P Chen, L Li, Z Yin, X He, X Sun, "Knowledgeable Storyteller: A Commonsense-Driven Generative Model for Visual Storytelling", IJCAI, 2019

[10] Seokhwan Yang and Young-Suk Lee, "A Study of an AI-Based Content Source Data Generation Model using Folk Paintings and Genre Paintings," Journal of Korea Multimedia Society, vol. 24, no. 5, pp. 736-743, 2021.

A Study on Cultural Memory and Identity of Chinese Myth Animation*

Shao - Yuwen
Design Department
Dongseo University, Busan, South Korea
2565711329@qq.com

Kim, Hae Yoon
Design Department
Dongseo University, Busan, South Korea
haeyoonkim@empal.com

***Abstract*—This study compares the color features of Chinese School of animation with modern myth animation and reveals the similarities and differences of Chinese animation color features in different periods. It illustrates the change in Chinese society and aesthetic sense by the change of characteristics of colors. In order to evoke cultural memory and promote identity from Chinese Modern animation, the author propose the integration for Che differences in the use of color between the Chinese School and modern myth animation. The cultural memory is evoked, and the cultural identity is promoted through the color integration. It also maintains the appeal and character of modern Chinese animation.**

Keywords—Myth animation, Chinese School of animation, culture memory, culture identity

I. INTRODUCTION

Myth is an important part of culture. It concludes the value, history and tradition of a society.(1) It can convey these culture memory by integrating myth story into animation and enhance the culture identity of a society.(2) People can feel proud and close to their culture, when they see myth stories symbols appearing in the animation. This helps not only strengthen social cohesion and identity but also preserve and inherit the cultural heritage.

In the history of Chinese animation, the 1950s and 1980s were the peak periods of Chinese animation, and the animation produced in this period is called "Chinese School" animation. It could be the important part of Chinese culture and these works focus on traditional Chinese value, philosophical ideas and historical stories and pursue unique aesthetics. Using traditional Chinese myth, traditional painting and art techniques, and classical music or sound design convey a profound cultural heritage experience. However, (3) with the influx of foreign animation and sino-foreign joint production, the ethnicity in Chinese animation was downplay. It was a turning point of Chinese animation when an animated movie adapted from a mythological theme in 2015.

As an important part of Chinese cultural memory, Chinese School of Animation not only carries important memories of culture, but also has unique characteristics of color. Therefore, this paper aims to have a deep insight into the characteristics of color used in Chinese School of Animation in order to have a deeper understanding of the evolution and trends of color use in animation works of these two periods.

"Memory" is an important source of information for the continuation and enrichment of Chinese culture. Through the preservation and transmission of cultural memory, Chinese culture can continuously exist and develop, and provide valuable information about traditional values, historical events, artistic expressions and literary traditions to future generations. These "memories" are the foundation of culture to help people understand and appreciate the diversity of Chinese culture, as well as promote a sense of cultural identity and empathy. By analyzing the color characteristics in these animation works, we can provide valuable suggestions for contemporary animation production to convey and promote the cultural values represented by the Chinese School of Animation, and to ensure its traditional colors to continue to inherit in today's animation field.

II. THEORETICAL BACKGROUND

A. Color System of Chinese School of animation

The creative production of Chinese School of Animation are borrowing from the forms of national art. Chinese School of Animation's has strong connection with traditional Chinese art. Chinese animation absorbs and borrows the traditional Chinese painting art, which leads to the formation of a unique Chinese School of animation in the world. Chinese School of Animation integrates the traditional Chinese color system (folk color system and literati color system) in the use of color. The traditional Chinese color system presents the aesthetic ideals of Chinese nation and reflects the inner world of Chinese nation's culture, as well as remain the independence and stability in the development of thousands of years. The nationalized color system of Chinese School of Animation mainly shows the following characteristics

- Firstly, the most of Chinese School of Animation presents simple, bright and concise color with strong decorative meanings.
- Secondly, Chinese School of Animation borrows plenty of folk-art forms in its creation, for example, Peking Opera, Paper-cutting, shadow puppets, folk New Year painting, etc. The colors of these folk-art forms adopt the Chinese five-color system. Five elements and Five colors are often used as the basis for the colors used in Chinese folk-art creations. It indirectly shows that the Chinese School of Animation also borrows and uses the color system based on the "five-color concept". It becomes a specific cultural and psychological cognition.

B. Color system of Chinese modern animation

In the early nineties, foreign animation occupied the main market in China, the animation processing factory and Shanghai Film Studio's lagging management system have

suggestions for color application in modern animation. It helps realize the reproduction and transmission of cultural This research was supported by BK21 Four Service Design driven Social Innovation Educational Research Team in Dongseo University.

caused a drastic loss of domestic animation talents, which makes the lack of talent become more serious problem. Since then, the Chinese animation has entered a low period. With increasing modern animation companies, causing difference of color without specific color matching system. Since the Chinese animation borrows from foreign animation is so frequently, it is impossible to use the original traditional Chinese colors reasonably. There is a difference between the traditional Chinese color system and the Western's. The Western color system is Scientific Color, which is spectral color, i.e., with the action of light, it discusses darkness and brightness. Tricolor, blue, red, yellow in modern color are similar with Chinese "Five-color" system. Modern color theory, through scientific research, the eye needs the sum of tricolor to achieve balance. The color generally is bright. On the contrary, traditional Chinese color is built up based on the spontaneity of the original color, focusing on the inherent color of the object, and will not appear bright-colored color. But, modern animation borrows much from Western colors, it causes that bright color are showed in modern animation.(4)

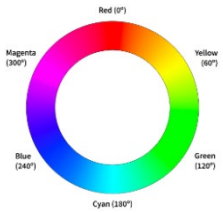
C. *Concept of HSB(V)*

The HSB color model is applied for research and analysis. HSB mode corresponds to the human eye. It describes the basic characteristics of color in terms of hue, saturation and brightness/value, i.e., the three elements of color are quantified, with saturation and brightness presented with a percentage value (0%-100%), and hue expressed with angle (0°-360°).(5) Therefore, HSB is more intuitive than other color models to calculate the value of three elements of color.(6)

Although RGB model is important in computer graphics and displays, the HSB is more applicable to art creation, design, and color selection.(7) Because it provides a more intuitive and easier way to describe and control colors. (8) This is the reason that this study selects this model.

D. *HSB Distinctive Range*

Hue represents the basic property or kind of color, which is the fundamental attribute of colors, such as red, green or blue etc. In the HSB model, hue is represented by an angular value from 0° to 360° , and this range can be imagined as a circle in which each angle corresponds to a different color. 0° and 360° means red in hue circle, 120° indicates green and 240° indicates blue, etc. By changing the value of the hue, different colors can be selected.(9)



Saturation and brightness of HSB color model can be divided into three degree. The low degree is 0-33%; medium is 34-66% and the high degree is 67-100%(10).

Distinction between saturation and lightness		
saturation	Low saturation	0 - 33%
	Middle saturation	34 - 66%
	High saturation	67 - 100%
lightness	Low lightness	0 - 33%
	Middle lightness	34 - 66%
	High lightness	67 - 100%

III. Comparison of Color Characteristics of Chinese School of Animation and Modern Myth Animation

A. *Case selection*

Taking 2015, a turning point of China's animation industry, as a starting point, the author selects the modern animation for analysis. Based on the mainland box office rankings of Chinese animated films from 2015 to 2022, the author select all animated films with mythological theme in the top 100.(11) A total of eight modern Chinese animation were selected.

The Chinese School of Animation has a long history, so the original movie is stored on film, which is affected by the damage during the process of use and other factors. There are quality problems in different degrees, including low clarity and smoothness, color distortion and flaw.(12) For this reason, 100 original animations restored with 4K technology are selected. And eight movies are chosen among them as the object of this study.

TABLE I. Selected Animation Titles

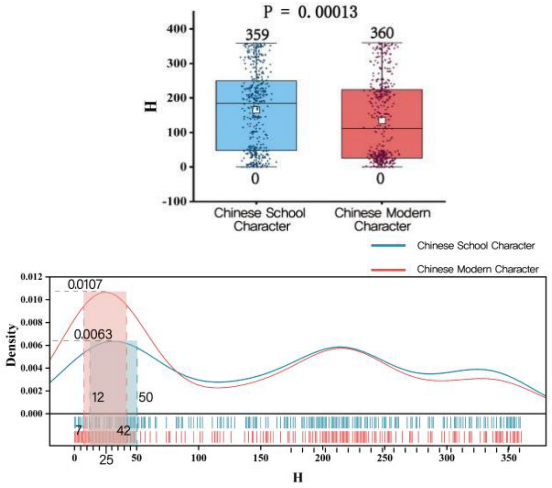
Chinese School of Animation	Year	Chinese Modern animation	Year
fisherman	1959	Monkey King: Hero Is Back	2015
The Monkey King: Uproar in Heaven	1961,1964	Big Fish & Begonia	2016
Nezha Conquers the Dragon King	1979	White Snake	2019
The Deer King	1987	Nezha: Birth of the Demon Child	2019
The Legend of Sealed Book	1983	Jiang Ziya：Legend of Deification	2020
Monkey King Conquers the Demon	1985	The Tribulation of the Green Snake	2021
Journey to the West Five Magic Weapons	1989	New Gods: Nezha Reborn	2022
Lotus Lantern	1999	New Gods: Yang Jian	2022

B. *Analytical method*

- Selecting 131 characters and 223 colors of background of Chinese School of Animation and Chinese Modern animation, 2139 RGB color data is selected by KSCA.
- The study uses HSB color mode to compare and analyze the color characteristics. Therefore, Python and conversion code converts 2139 RGB color values into 2139 HSB color values.
- Box-plot is completed by Origin 2021. Box-plot, called Box-whiskerPlot, which shows the distribution of the data. It is mainly used to reflect the characteristics of the distribution of the original data, and can also be used to compare the distribution characteristics of multiple groups of data. (13)
- In order to analyze whether there is a difference in the color of the animation in the two periods, HSB color value is compared by Student's t-test. When P is less than 0.05, it indicates that the difference reaches a significant level.

C. *Comparison of color characteristic of character in Chinese School and Modern animation*

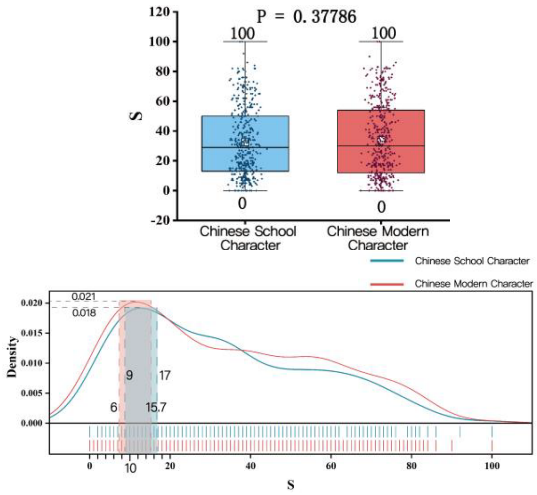
- Comparison of hue of character



By comparing the hue characteristic of characters of Chinese School and Chinese Modern animation, the box-plot shows that the range is 0-359, 0-360, which illustrates that the both use a wild range of hues. P-value is 0.00013, which indicates that the two data has significant difference.

Additionally, the two data sets have a relatively even distribution of hues, so the values indicating the highest range of use are representative of the characteristics of this data. In terms of common use of hue range, the Chinese School concentrates between 12-50, while modern Chinese mythological animation concentrates between 7-42, both favoring warm tones.

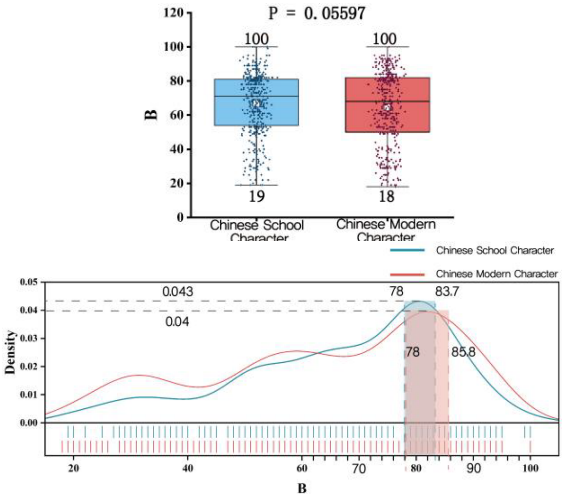
- Comparison of character saturation



The results of analyzing and comparing the characteristics of character saturation in Chinese School and Chinese Modern myth animation from box-plot show that the distribution ranges are 0-100, and both two sets of data involved, high, medium and low saturation ranges. P-value is 0.37786 and there is no significant difference between the two sets of data.

From the box-plot can be seen that the two sets of data are equally distributed. The range of the highest usage rates can characterize this data, with the Chinese School focusing mainly on values ranging from 9-17 and the Chinese Modern focusing mainly on values ranging from 6-15.7. Both belong to low saturation.

- Comparison of character brightness

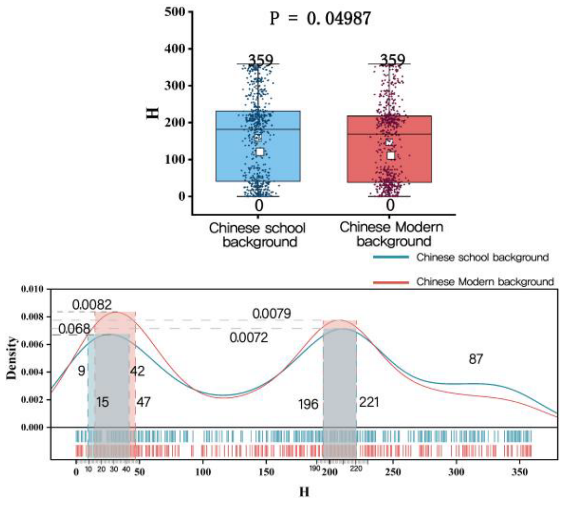


The study compares the characteristics of character brightness in Chinese School and Chinese Modern myth animation. The results show that the distribution ranges are 19-100 and 18-100 as can be seen from the box-plot, and both sets of data involve, high, medium and low ranges of brightness. The significant difference p-value is 0.05, although there is no significant difference, but there is a smaller difference.

From box-plot, it can be seen that the hue of two sets of data have an even color distribution. Thus, values with high frequency of occurrence can relatively objectively characterize the data. The value range of Chinese School is 78-83.7, while Chinese Modern's is 78-85.8. The both belong to high brightness.

D. *Comparison of background characteristics between Chinese School and Modern myth animation.*

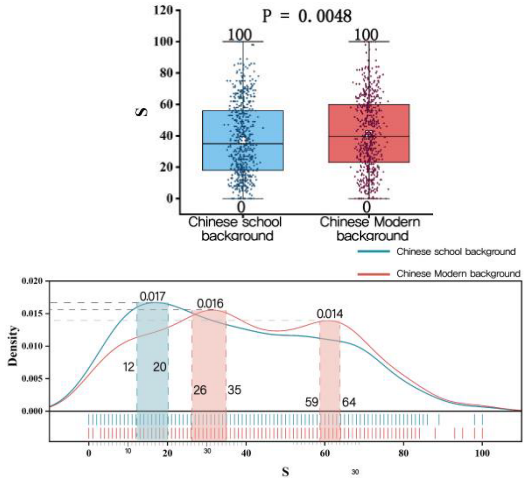
- Comparison of the hue of background



After analyzing and comparing the hue situation of the backgrounds in the Chinese School and modern Chinese myth animation, the results show that the two sets of data have the same range of hue, which includes 0-359 degrees, indicating that they widely use different colors. Statistically analyzing the p-value of 0.04987, there is a significant difference between the two sets of data.

It can be seen from the box-plot that the contribution of data is even. So, the most frequently used hue ranges can represent the characteristics of this set of data. The highest range of Chinese School is 196-221, which is cool tone, while the Chinese Modern is 15-47, which is warm tone.

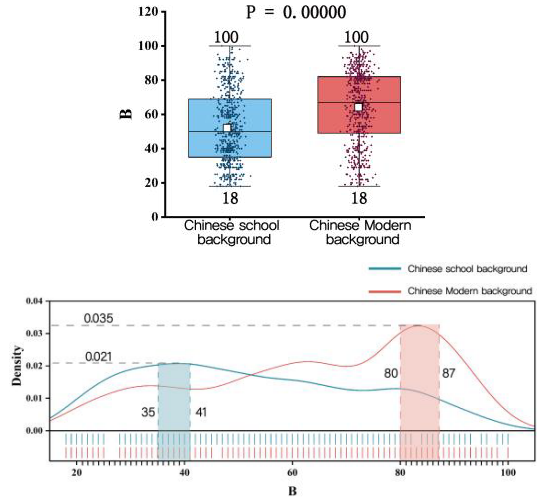
- Comparison of saturation of background



After comparing the saturation in Chinese School and Chinese Modern myth animation. The results shows that the saturation ranges are similar in both sets of data and include high, medium, low saturation. The p-value is 0.0048, which means that the two sets of data are significantly different.

In addition, the data is evenly distributed in box-plot. The range of highest usage can represent the characteristics of this data set. The highest range of Chinese School is 12-20, which is in the low saturation range, while the Chinese Modern animation is 26-35, which is in the low-medium saturation range.

- Background brightness comparison



Comparative analysis of the brightness of the backgrounds in the Chinese School and modern Chinese myth animation shows that the two sets of data have the same range of brightness, which is 18-100, covering high, medium and low brightness. P-value is 0/00000, which means that the two sets of data are significantly different.

From the box plot, an even distribution of the data can be seen. The range with the highest usage can represent the characteristics of this data. The brightness of Chinese School is mainly concentrated in 35-41, which is medium brightness, while the brightness of Modern Chinese is mainly in 80-87 which is high brightness.

E. Comparison of color characteristics between Chinese School and modern myth animation

TABLE II. COMPARISON OF COLOR CHARACTERISTIC OF CHARACTER IN CHINESE SCHOOL AND MODERN ANIMATION

	Chinese School Character	**Chinese Modern Character**	**Difference**
H	12-50 warm tone	7-42 warm tone	Significant difference
S	9-17 Low saturation	6-15.7 Low saturation	no significant difference
B	78-83.7 High birhgtness	78-85.8 1High birhgtness	no significant difference

It is seen that there is not much difference in the characteristics of color usage in the characters in the animation of the two periods. From the viewpoint of the three elements of color, the hue mainly uses warm colors, the saturation is in the low range and the brightness in the high range.

In traditional Chinese myth culture, many stories and god characters have fixed color symbols and specific aesthetic concept and emotional expression. In traditional Chinese art works and Chinese philosophical concepts, color also has a specific symbol, and the use of colors in line with traditional aesthetic concepts and emotional expression, especially the emotion and recognition of traditional Chinese culture.

Indirectly, it shows that in previous animations, colors may be simple and less used due to technology. Modern myth animation, due to technological advances, although more complex and brilliant color effects can be achieved, the colors of the characters are chosen to retain the traditional color style as much as possible. This belongs to China's unique cultural memory.

TABLE III. COMPARISON OF BACKGROUND CHARACTERISTICS BETWEEN CHINESE SCHOOL AND MODERN MYTH ANIMATION.

	Chinese School background	**Chinese Modern background**	**Difference**
H	196-221 Cool tone	15-47 Warm tone	significant difference
S	12-20 Low saturation	26-35 Low-medium saturation	significant difference
B	35-41 Medium saturation	80-87 High brightness	significant difference

As can be seen from the table, there are significant differences in the characteristics of color use in the backgrounds in the two periods of animation, both in terms of the overall data and the characteristics of the data. However, the characteristics of color use in the two periods are different in terms of the three elements of color.

Chinese School myth animation is on the cool side, while modern myth animation is on the warm side. In terms of the saturation, the school myth animation belongs to the low range, and the brightness belongs to the middle range. The saturation of modern mythological animation is in the low-middle range, and the brightness is in the high range. Modern

mythological animation has more vibrant and brighter colors in the background compared to Chinese School of Animation. And the backgrounds of school animations show gray and cold tones.

The cold tones in the background of Chinese School myth animation reflect the aesthetic concept of traditional Chinese culture, which focuses on introspection, elegance and calmness. This color selection can be traced back to ancient Chinese culture, which emphasizes on the humility and thoughtfulness in traditional social values. The application of these colors in animation carries Chinese cultural memories, makes viewers feel close to traditional values, and evokes a sense of cultural identity.

In modern mythological animation, the background colors are warm tones. The reason is that China has experienced profound social, economic and cultural changes in the process of modernization, leading to changes in people's aesthetic concepts. Modern society places more emphasis on individuality, vitality and openness, and these qualities are more easily expressed through warm colors. In addition, warm tones are more popular among the color expressions commonly used internationally, so modern mythological animation may be more inclined to use these warm tones, which can be better integrated into the international market.

IV. CONCLUSION AND SUGGESTION

As a medium for transmitting the memory of culture, it can help animation to transmit the values and aesthetics of traditional Chinese mythological culture by preserving the traditional style and evoke Chinese cultural memory, thus, promoting cultural identity.

The color usage of Chinese School and modern mythology animation shows the evolution of Chinese culture, society and the identity with traditional and modern values. The animation work delivers the cultural memory and identity through color and also reflects that China has been constantly changing and adapting on the global cultural stage.

- Because the characters are depicted in warm tones, the background conversely is in cool tones. Such visual contrast highlight the characters and story.
- In terms of modern animation production, the advanced technology and artistic method are used for the color of background in order to present more colorful background details. This gradation can be working in different parts and scenes to highlight key plot or at the time of character appearing. Moreover, such gradation can balance the mystery of cool tones and the modernity of warm tones creating a more layered background. Also, it can evoke the identity of the audiences with traditional culture and contribute to enhance the memory and identity with traditional Chinese culture.
- Integrating the background color tones of background into modern myth animation and the symbols and patterns of traditional Chinese culture into the background can strengthen cultural memory and identity. These symbols are presented in warm colors but blend with the cool tone background to create a unique visual effect. This integration saves the characteristics of Chinese School animated myth and combines with modern color, which manifests national culture and adapts to international trends, achieving a balance between culture identity and international integration

This paper discusses animation as medium for transmitting cultural memory, especially the inheritance and expression of traditional Chinese mythological culture. Also, the paper mentions that how the animation evokes the cultural memory in traditional way and how it shows the evolution of Chinese culture and the culture identity of traditional and modern values through the use of color in modern myth animation.

- The paper points out that the use of color and background design in modern myth animation not only reflects the evolution of Chinese culture, but also contributes to the development in global stage.
- It promotes exchange and integration between Chinese and international cultures and the understanding and cooperation of cross-culture.
- The research method concludes the analysis of color use and background design, which demonstrates how the academic approach explains the importance of cultural expression and memory.

In conclusion, the significance and value of this paper lies in the in-depth discussion of how animation, as a carrier of cultural memory, expresses cultural memory through color and design to promote national identity. It contributes to have a better understanding of cultural inheritance and the importance of culture identity, and also provides valuable insights for cross-cultural studies and the animation industry.

[1] 高有鹏, 2023.05.16, 《中国传统神话的文化价值与当代表达》, 人民论坛, To.8, [N][OL],http://paper.peoplepeople.com.cn/rmlt/html/2023-05/15/content_25999957.htm

[2] Hao Kong, Yibo Qi, Yuqian Qi, 2022, "A Study of Communication Competence of the Chinese School in the Context of Cultural Confidence," Advances in Social Sciences, vol.11, No.3, [J], pp.687-695

[3] Shen Hao, 2019.01, "Aesthetic features and symbolism of the Traditional Color Concepts in animation of Chinese", Journal of Communication Design, No.66, [J], p191.

[4] Shao Yuwen, Kim HaeYoon, 2022, "Research on the Difference of Colors of Animation Between Chinese Schools and Modern Animations in China", Journal of Communication Design, No.79, [J], p125.

[5] 钱品辉主编；刘涛副主编, 2013.09,《色彩构成》, 人民美术出版社, [M], p112

[6] 陶晓欣, 2004, "中文版 Photoshop 7 图形图像处理基础与应用", 海洋出版社, [M], pp34-35

[7] Erik D. Kennedy, 2020.05.01 "The HSB Color System: A Practitioner's Primer", https://www.learnui.design/blog/the-hsb-color-system-practicioners-primer.html, [OL], 2023.07.11

[8] 康博创作室编著, 1999.03, "Photoshop 5.0 中文版实用教程",人民邮电出版社, [M], p109

[9] HSB 컬러 시스템, https://brunch.co.kr/@blackindigo-red/26, [OL], 2023.07.11

[10] HSB 简介, https://blog.csdn.net/Flamus/article/details/78583856, [OL], 2023.07.23

[11] 中国国产动画电影内地票房排行榜, https://zh.moegirl.org.cn/, [OL], 2023.09.25

[12] 百部中国动画完成 4K 修复

https://baijiahao.baidu.com/s?id=1749272770958112951&wfr=spider&for=pc 2023.09.21

[13] 箱形图(Box-plot), https://baike.baidu.com/item/箱型图/10671164, [OL], 2023.07.21

Mimetic and poetic visualization of war in film as a means of raising awareness based on the analysis of visual style of "Come and See"

1st Olena Lastivka
Student, Dept. of Visual Contents
Dongseo University
Busan, Korea
olenalastivka93@gmail.com

2nd Sunghoon Cho
Professor, Visual Contents
Dongseo University
Busan, Korea
mattcho@gdsu.dongseo.ac.kr

3rd Yong-Ki Lee
Professor, Design
Dongseo University
Busan, Korea
ykl177@gdsu.dongseo.ac.kr

***Abstract*—This paper explores the potential of film as a tool for raising global awareness about wars that continue to occur in the world. It also discusses mimetic and poetic styles of visualization and highlights the significant role of visual style in how people perceive and understand war. By selecting the appropriate visual style, films can have a unique potential to challenge war stereotypes and assumptions, encourage empathy, and raise awareness of the causes and aftermath of wars due to films' widespread availability and the storytelling techniques.**

Keywords— visualization, film, empathy, mimetic and poetic, awareness

I. Introduction

It is important to first define what is meant by the war visualization. In the analysis on visualization of war through photography Dursun says that war is a collective, direct, open, personal, intentional, organized, institutional, instrumental, imposed and sometimes ritualized form of violence caused by the conflict of interests of states, nations, or cultures [1]. Although death, violence, injury, refugees, economic problems are the usual and obvious topics of war visualization, other aftereffects such as the huge social and psychological changes created by wars are highly underestimated [2]. That is why in this paper by visualization of war in film we mean not only visualization of shocking side of war like battles, arms, death and destruction but also showing stories about displacement, disabilities, economic problems, government problems, educational problems, ecological problems, famine, poverty, phycological individual trauma, collective trauma and other aftereffects caused by war, which in some cases may happen long after the war is over. And by visualization or representation itself we mean that the storyline and visuals of the film (both documentary and feature) includes one of the mentioned above topics. In other words, visualization of war in this paper does not necessarily mean films of war genre but visualization of the events related to causes, process, or consequences of wars in films of any genre.

While visualization of war in film is often political and overly simplified, film is a powerful tool for raising global awareness about war if filmmakers use visual style and storytelling possibilities properly. These days film is widely available to the mass audience. Storytelling and visuals of film make it easier for people to feel empathy towards those affected by wars, and, moreover, various visual styles can make it possible to show such complex topics as war from different perspectives. By carefully selecting and employing appropriate visual styles, films can challenge usual war stereotypes and assumptions, encouraging empathy and deepening awareness regarding the causes and consequences of conflicts. This way, films can be effective in promoting peace and minimizing the catastrophic impact of wars on civilians and society.

Thus, in our research we focus on the following question: how should we use the visualization of war in film to raise global awareness about war and in its turn mobilize people for more active support? The characteristic features of mimetic and poetic styles in war representation are also discussed under the angles of ethics and efficiency for raising global awareness. The scope of the research is limited to analysis of previous publications and visual style analysis of one of the most well-known anti-war films, "Come and See". Ultimately, this research is aimed to underscore the potential of film as a catalyst for global awareness and in its turn for significant actions towards war aftermath relief.

II. Literature Review

First, our analysis of previous research works on the specifics and role of visualizing war in films for raising global awareness showed that film is a powerful tool for conflicts understanding and transformation because film can be a stimulus for and/or mean of communication on the social level, which is crucial for conflict transformations [3]. Kellett and Matyók explore the role of communication in promoting conflict transformation and peacebuilding at the local, regional, and global levels. Through their research they highlight the crucial role that effective communication can play in promoting dialogue, mediation, and reconciliation in conflict situations. Furthermore, the storytelling possibilities of film could be used to challenge war stereotypes and assumptions, encourage empathy [4]. The key idea is using creative arts (art, music, and storytelling) in promoting reconciliation and social transformation in the aftermath of conflict and trauma. That is why creative arts can be effectively used to facilitate empathy, understanding, and dialogue between individuals and communities. Another reason why film is a powerful tool for raising awareness is the fact that film is one of the forms of art. That actively utilizes storytelling, and it can be used as an effective tool for peacebuilding [5]. Manojlovic concludes with the statement

that education is the most powerful tool for peacebuilding, and how innovative approaches to education, like film, can help to transform conflict and build sustainable peace.

The second insight was that war, as any other historical event involving many people and sometimes many nations, is extremely complex and should not be simplified for the purposes of visualization in film. First, there should be a critical approach to the representation of historical events. Previous research shows how media can shape our understanding and memory of historical events [6]. And thus, a more complex critical approach to the representation of such events is needed. Moreover, it is important for filmmakers to engage with the complexity of the conflict and to resist simplifying the narrative into easily digestible tropes of good versus evil. On the example of the Syrian Revolution and War Chaudhuri shows how conflict has been represented in the media and in films primarily from the perspective of the distant spectator, who is often outside of the conflict zone and viewing the events from a safe distance [7]. A more complex approach would contribute to a deeper understanding and shaping of the public perceptions of conflicts, so it is important to consider the perspective of the spectator in the visualization and narratives.

Furthermore, since war is such a complex thing to portray, the third insight was that the visualization of war in any kind of visual contents, and film in particular, has its ethical issues which should be taken into consideration. Digital media has become an increasingly important tool for representing and shaping our understanding of war. Robinson and Schulzke discuss the potential benefits and limitations of using digital media to represent war, such as the ability to engage audiences in new ways, but also the risk of perpetuating certain stereotypes or misrepresentations [8]. Besides the ethical issues of misrepresentations, there is also an ethical issue of imagery depicting human bodies in pain, particularly in the context of war and conflict. Dauphinée argues that the use of the imagery of bodies in pain can be deeply problematic, as it can reduce the suffering of individuals to a spectacle or a tool of political manipulation [9].

Following the two previous insights, the final insight was that the choice of visual style is crucial for the ethical and comprehensive visualization of war. For example, mimetic and poetic style each has an important but different role in representing painful pasts in films, and thus a careful balance between the two can help to create a more nuanced and complex understanding of the human experience of war. Pötzsch demonstrates on the examples of different films how these styles can be used to create different types of meaning and emotional impact [10]. Furthermore, it is important that after the First World War the cinema started to be used as a tool for propaganda and this led to the development of a new language of film [11]. Through these previous studies we can see that the history and evolution of visual style of war visualization is very complex, and that film plays an important role in shaping public opinion and political discourse.

III. Poetic realism of war in "Come and See"

"Come and See" directed by E. Klimov tells the story of the occupied and burnt Belarussian villages by Nazi Germans during the World War II through the eyes of a young boy Flyora. The film is based on the historical events when allegedly 628 Belarussian villages were burnt together with all their inhabitants. This information is given in the title at the end of the film, and it gives an even stronger sense of realism to everything that happened on the screen.

"Come and See" starts with Flyora joining local partisans, full of excitement and inspiration to fight Nazis. However, among more experienced partisans he is treated as a total beginner and when they move further, he is left behind. In the woods he meets a beautiful girl Glasha. Together they hide from the bomb shelling and decide to go back to Flyora's village.

Even though the film mainly employs an objective point of view where we see the events from the perspective of an observer, sometimes we can see or hear things subjectively from the point of view of the main protagonist Flyora. For example, when bomb shelling starts in the woods, we can hear whistle as if the bomb explosion stunned the viewer too. After the shelling for some time during the film we hear all the sounds muffled. This effect gives a strong sense of realism, which is one of the main characteristics of a mimetic style of visualization.

When Flyora and Glasha come to Flyora's house, they notice that there are no Flyora's mother and younger sisters. While they eat this still warm food prepared by Flyora's mother we can hear the sound of flies, which is getting more and more. Even Flyora who subconsciously doesn't want to admit that things are off in his house, eventually cannot stand the sound of flies and the atmosphere that they create and goes out to search for his family and other villagers. The flies over the Flyora's sisters' toys symbolize death in this scene. It is a very poetic means of conveying the message since it does not show directly what is wrong. It is possible to say that this is one of the exemplary scenes of poetic visualization of war. At this point in the film the viewer has only the frame of the events, what exactly and how it happened is up to the active re-imagining of the spectator.

On the other hand, in the next scene, when Flyora decides to run and find the people of his village who, as he convinced himself, ran away and are hiding somewhere on the island, Glasha looks back and sees the pile of dead bodies behind the house. Ther are shown on the screen only for a couple of seconds, but it gives an explicit image of the death. At this point there is no need to imagine, the viewer gets answers, which is one of the most prominent characteristics of mimetic style of visualization.

Another important detail in "Come and See" is that in the first half of the film audience does not see Germans at all. The only things that symbolize them are the planes that are constantly flying over the villages and the shelling in the woods. We only hear about Germans from people whom Flyora meets. So, the viewer does not get any certainty and actively imagines. These are the poetic elements of visualization of war. Such elements make the audience more curious and engaged and make a film more objective with the conclusions.

However, in the second part of the film we can see the Nazi Germans, who occupied one of the Belarus villages, the way they treat local people and the way they eventually burn them all together in a barn. These explicit visual images of what happened give direct answers to the viewer. Now, the viewer does not need to imagine, it is clear that the Nazi Germans were inhumane demons whose goal was to destroy and wipe out innocent Belarus people with extreme cruelty. Such scenes

are characteristic examples of mimetic style. After watching such shocking scenes, viewer will undoubtedly be left with this image in the head. Especially because it is based on true historical events.

In the end of the film there is a scene when Soviet partisans come to the village and kill in anger a group of Germans and their collaborators who were left behind by their people. These explicit images show the righteous anger of Soviets and could have been considered as a quite unethical call for revenge, if not for the last scene, where Flyora sees the portrait of Hilter in a puddle and starts to fire at it from his rifle. Klimov adds a montage of real videos played backwards of Hitler's speeches, young, inspired Nazi followers greeting Hitler until it stops at the photo of Hitler as a child sitting on the laps of his mother. Then Flyora stops shooting too and starts crying. This scene is a perfect combination of mimetic and poetic styles. It shows the real documentary videos of Hitler and destructions of Nazi Germans, which gives the clear answers. However, on the other hand when we see how Flyora stops shooting at the photo of young Hitler, the viewer undoubtedly will start to think if Hitler was born evil or became evil, if the revenge is moral or there are other ways. This last scene is just a poetic way of creating a frame to make the audience actively re-imagine and reflect on everything they have just seen on the screen to fill this frame with their own answers.

It is also worth mentioning that the face of the main character Flyora changes all the time throughout the events of the film. And if in the beginning audience see him as a young boy who laughs a lot, in the end the horrors of the war turn him into a suffering old man in the body of a boy. This change in his face, which could be seen in Figure 1, is a poetic symbol of the trauma that people suffered in the occupied territories.

IV. DISCUSSION

War has always been quite a popular topic for visualization and storytelling. Since it usually has action, drama and brings out very strong emotions, it is a widely used theme for both classical art and new media art like computer games. In their work on visualizing war [8] N. Robinson and M. Schulzke say that even though in the modern world most of the people have never experienced real war, they experienced it and have the general idea of it through news, TV, computer games, films, and pop culture.

After analyzing previous research papers there can be seen a clear division into mimetic and poetic styles of visualization. As Pötzsch describes in their paper [10], mimetic style is basically the "imaging of the past". It means that past events are shown directly the way they happened but in the form of a film. And through this direct representation of the war the past sufferings and pains of the people involved are shown directly too. The purpose of such visualization is to give the illusion of documenting reality on the screen. This style of visualization positions the audience of the film as passive spectator. It usually gives all the answers without giving the opportunity for the audience to imagine what happened on their own.

On the contrary, poetic visualization is called the "re-imagining of the past" in the paper. It means that the past pains are not shown as they happened but through symbols and poetic meanings of some other events. This style just gives the frame of the historical past without showing anything too certain. The audience is given the chance to actively think and imagine past events as well as their true meaning without getting the precise answers.

(a) (b)

Figure 1. Screenshots from "Come and See" when (a) Flyora is happy to join Soviet partisans in the first act of the film and (b) Flyora witnessed the burning of the people in a barn by Nazi Germans in the last act of the film.

Since the war is inextricably linked to death, destruction, injuries, and traumas, it is crucial to keep in mind possible ethical issues while visualizing war in film. One of the biggest ethical issues is the depiction of sufferings of others. Even though it may seem not a big problem, it has several ethically problematic consequences when it comes to visualizing sufferings of real historically existing people. First, when people see the exact depiction of the horrors that happened to someone else in films too much, it makes them lose sensitivity to such sufferings. People start to think that such pain is a normal thing, which they have experienced already multiple times (even though it was only through the screen) and thus they start to depreciate the sufferings of those people who went through the horrors of war. Such visualization reduces the significance of the victims to mere objects and makes the viewer just a passive consumer and potential voyeur.

Moreover, depiction of past pains of war is problematic in a way that it makes victims of the war feel "double betrayal", which means the use of their past pain for political purposes. Victims of the war are forced to re-live their past pains again and usually it is not even done in commemoration of their pain but for political or entertainment purposes. This aspect is not ethical and should be taken into consideration when visualizing the sufferings of war in film.

Apart from the issue of visualizing past pains, there is also another important ethical issue – political propaganda. Since war brings out a lot of strong emotions related to national or ideological questions, in a lot of cases it is easy to use those emotions for political purposes. As it was mentioned previously, using past pain of victims of war for such purposes is not ethical. It is especially dangerous if a film visualizing war is promoted as being educational but is also political.

Mimetic representation is a good choice if we want to show the events and sufferings very close to or precisely as they happened in the past. Such realistic war visualization usually has a great impact because viewers get shocked from seeing the painful images (for example, the pile of dead people or burning people alive in "Come and See") and knowing this is based on true historical events. However, the impact is usually short-term as it does not make people engaged. It just lets people be passive consumers of the past pains and it also may result in being ethically incorrect visualization. Such visualizations potentially can turn the viewer into a voyeur who just finds it entertaining to watch the sufferings of the victims of the war without critically thinking about the reasons, consequences, and responsibility. Also, in a lot of cases mimetic representation is done for political propaganda using the pain and sufferings as emotional manipulation of people's political opinions.

The poetic style on the contrary has less ethical issues since it does not depict the past pains of victims of the war precisely but just gives the frame of the past events. The

person who watches a film with such war visualization is given the chance to think and imagine what happened, what were the reasons, who was right and who was wrong and what can be done to avoid such conflicts in the future. Such style of war visualization gives more space for critical thinking and self-reflection. Moreover, it achieves much longer and deeper engagement of the viewers because even though the stories are not fully true to the history, people can feel much greater sympathy towards the characters. Usually, such poetic war visualization is much more impactful in terms of forming the socio-political ideas of people in general. It is an effective way to raise awareness if used properly.

As we have seen during the analysis of "Come and See", mimetic and poetic styles of visualizing war are not mutually exclusive, they can be combined, and, this way, create an even more powerful effect on the audience. The mimetic elements of "Come and See" show events of the film precise to the history, adding real footage of Hitler in the end and ending with the title that say about 628 Belarusian villages burnt with its inhabitants. All this creates the impression of just documenting the crimes of Nazi Germans in Belarus at that time, which could lead to viewers just consuming this information without re-thinking what they have just seen.

However, the masterful combination of documenting with multiple poetic symbols and metaphors in this film creates a much bigger frame of the past that should be filled with audience's own imagination. This gives the opportunity to the viewer not only to consume past pains but also to reflect on it, which is a much more ethical approach based on our research of previous studies.

V. CONCLUSION

This research concludes with the statement that film, like any other form of art, can play an important role in not only portraying the wars but also in shaping our perception of the past and ongoing conflicts. As a result, it can be an immensely powerful drive for forming our present and future. That is why it can be and should be used as a way of education and raising awareness about ongoing and past conflicts. It is especially effective if it involves engaging storytelling and gives the opportunity for the audience to re-imagine and re-think war.

However, it is important to remember that war visualization is the visualization of other people's pain. So, such aspects as political agenda or reviving other people's traumas on the screen might become a big ethical issue. That is why it should be kept in mind and artists should comply with such ethical aspects if they want to use visualization of war in film as an effective but ethical way of raising awareness about globally important topics such as war and other historical events related to society development and human rights.

Since this paper is planned to be used as a foundation for the future research of the influence of visualization of war in film on the global awareness about war and peace, we could also say that combination of mentioned above features of mimetic and poetic styles of visualization will help making the anti-war film messages more powerful and this way contribute to the increasing of the awareness in general.

REFERENCES

[1] O. Dursun, F. Yildiz, and S. Bulut, "Dichotomy between war and visualization of war: an analysis of the war photos awarded by the WPP", Moment Dergi, 6(2), Article 2, 2019.

[2] A.A. Stein, and B.M. Russett, "Evaluating war: outcomes and consequences", 1980.

[3] P.M. Kellett, and T.G. Matyók, "Communication and conflict transformation through local, regional, and global engagement", Lexington Books, 2016.

[4] C. Cohen, "Creative approaches to reconciliation", 2006.

[5] B. Manojlovic, "Peacebuilding through education: innovative ways of dealing with conflict", 2018.

[6] H. Pötzsch, snd V. Šisler, "Playing cultural memory: framing history in Call of Duty: Black Ops and Czechoslovakia 38-89: Assassination", Games and Culture, 14(1), pp. 3–25, 2019.

[7] S. Chaudhuri, "The alterity of the image: the distant spectator and films about the syrian revolution and war", Transnational Cinemas, 9(1), pp. 31–46, 2018.

[8] N. Robinson, and M. Schulzke, "Visualizing war? Towards a visual analysis of videogames and social media. Perspectives on politics", 14(4), pp. 995–1010, 2016.

[9] E. Dauphinée, "The politics of the body in pain: reading the ethics of imagery", Security Dialogue, 38(2), pp. 139–155, 2017.

[10] H. Pötzsch, "Imag(in)ing painful pasts: mimetic and poetic style in war films", "Ethics and Images of Pain", Routledge, 2012.

[11] C. Stojanova, "The Great War, cinema propaganda: the emancipation of film language", Acta Universitatis Sapientiae, Film and Media Studies, 14, pp. 131–156. 2017.

Compressed Character Arcs in Short Film: Exploring Character Development within Time Constraints

1st Yang Wanye
Student, Dept. of Visual Contents
Dongseo University
Busan, Korea
wanyeyang33@gmail.com

2nd Sunghoon Cho
Professor, Visual Contents
Dongseo University
Busan, Korea
mattcho@gdsu.dongseo.ac.kr

3rd Yong-Ki Lee
Professor, Design
Dongseo University
Busan, Korea
ykl177@gdsu.dongseo.ac.kr

***Abstract*—Character arc is a commonly used technique in film creation, which can be simply explained as the trajectory of a character's development and change from the beginning to the end of a story. Since the early stages of film development, character arcs have played a vital role in making characters more appealing. In traditional two-hour films, there is ample time for character development. However, with the emergence and popularity of short films, the challenge has been how to effectively create character arcs within limited time frames, avoiding one-dimensional characters. This study explores methods to enhance character arcs within time constraints through strengthened character development and narrative structures.**

Keywords—Character arcs; Short film; Compressing; Time Constraints

I. INTRODUCTION

Since the word "character arcs" comes from English, the keyword of most studies on movie characters in China is "characterization". The author searched for the keyword "characterization" in CNKI Knowledge Resource Library, and found that there are more than 10 articles related to this subject under the category of art. One of the main research objects of the character arc is Liu Ziyan of Northeast Normal University, "On the Movie Scriptwriting Technique - "Character Arc"", which focuses on the deep analysis of the character arc from the perspective of movie scriptwriting. The keyword of this paper is compressed character arcs, aiming to explore how to make character arcs complete under the time constraints of short films. When I searched for the keywords "short film" and "character arcs", there was only one related master's thesis <Research on the creative method of character arcs in micro films -- taking the short film "gain> is very detailed from the character setting, narrative techniques and expression methods to explore the creative method of character arcs applicable to short films. This essay provides a very great help for this paper.

Character arcs refer to the "development trajectory or change in a character's nature, whether it's for the better or worse."In simple terms, it can be understood as the development and change of a character's personality. Characterization is more varied and realistic because of the character arc. Character arcs are highly regarded by Hollywood screenwriters, but it's hard to find out exactly who first came up with them. But there is no doubt that character arcs have been used for a long time in film and television.

Character arcs are of great importance in movie narratives. A complete character arc means that the character's personality or goal is fully transformed in the movie, which can reflect the character's authenticity and three-dimensionality.

Whether it is Don in Green Book, who is arrogant and self-contained at the beginning but accepts himself more in the end and shows more vulnerability and truth; or the son Will in Big Fish, who pursues rationality and truthfulness at the beginning but accepts his father's fantastic story and understands more about the wonders of life in the end, we can see obvious character arcs in them, and therefore the characters are more Realistic. This allows the audience to better empathize with the characters and better convey the core theme of the film.

Unlike regular two-hour movies, short films are generally defined as movies that are no longer than 40 minutes in length. Due to the short duration of the movie, the advantage of low cost makes making short films a must for new filmmakers entering the industry. However, the short duration also brings with it some limitations.

How can character arcs be compressed within the constraints of time so that both character arcs and narrative remain intact at the same time? The answer to this question is of great importance to filmmakers, audiences, and researchers. This paper will use case studies to examine techniques, strategies, and creative approaches to compressing character arcs and provide insights into effectively portraying character development within limited screen time.

II. BODY

To make the character arcs as well as the narrative as complete as possible within the constraints of time, we can separately strengthen the two major elements. For instance, while ensuring the narrative remains intact, enhance the characterization of the individuals; or conversely, while ensuring the depth of the characters, strengthen the narrative structure.

A. Using narrative techniques: highlighting the climax

Firstly, when we focus on narrative techniques, we can shape character arcs by enhancing the climax of the story and emphasizing key moments. By downplaying the beginning and ending, and trimming some non-essential or less crucial moments, we can reduce the time required for character arc development, conveying a concise yet impactful arc.

The climax emphasizes the drama and impact of the conflict, which can reach a pinnacle of emotion, form and content. In short films, the arrival of the climax means the highlight of the character's arc. Therefore, weakening the beginning and the end and strengthening the climax can highlight the character's arc in the shortest time and break through the time limit of the short film.

In the short 18-minute movie FEELING THROUGH, the climax is due to the bus driver's perfunctory attitude towards the old man, the boy has been pent up mood suddenly broke out, he for the deaf old man's eyes full of tears yelling at the driver, this action reaches the climax of the film, but also the boy's character arc of the highlights of the moment, so far the boy's character arc is complete, the next boy's character will be completely to the good, which is from the end of the boy will be stolen money into the beggar's bowl can be seen, the boy was eventually cured, the human nature of the trajectory of the comprehensive and profound display.

B. Using narrative techniques: Emphasizing Key Moments

Instead of depicting every nuance of a character's journey, the short film could have chosen key moments that represent the essence of their character arc. These moments become the focal point and represent the essence of character development. These key moments specifically include key decisions, turning points, and emotional climaxes. By emphasizing these key moments, creators are able to portray character arcs deeply and vividly in a limited amount of time. This focused approach helps create memorable and engaging character development while ensuring that the story maintains a clear and strong narrative over the length of the short film.

Through the short film research methodology, it was found that many short films focused on showing key moments in the character's life, with a particular emphasis on turning points. Choosing to show major turning points in a character's arc can be effective in demonstrating shifts in character development. This could be an unexpected encounter, a blow, or a spiritual realization, and through these pivotal moments, characters are able to experience significant development in a short period of time.

For example, in the short film ON ANOTHER NOTE, the main character Nathan, on his way home to work for his parents after deciding to quit the job he is currently working at, is stranded on the side of the road because his car has a problem, he accidentally meets the music teacher who used to teach him, and the two of them have a conversation. This conversation is what the story is all about, one scene, two people, just the conversation. The conversation itself is to make Nathan solve his own knot, so the whole film is emphasizing this accidental encounter, and in the end of the film, Nathan understands the teacher's good intentions for himself, so the film through the use of such a pure dialogue and the form of a single scene to emphasize the turning point of this character. In the end, Nathan returns the car the way it came, implying that he will continue to pursue his dream, that he has changed his mind and has grown. The character arc is realized in the shortest possible time.

C. Character Setting Tips: Enhancing Characters

In addition to compressing the length of the narrative to make up for the character arcs, the character arcs can also be compressed directly through the enhancement of the character settings. Generally, conventional movies have a large number of characters and complex relationships, and the conflicts between characters are also divided into many complex narrative lines to tell, so that the character arc can be shown completely. However, the time constraints of a short movie, where the characters are usually one or two, limit the narrative. In order to ensure the integrity of the character arc as well as the narrative, the characterization needs to be strengthened in two ways.

One of them is the reinforcement of character gaps, that is, to set obvious and very imperfect flaws in the character's personality or background information, such as character or physical defects, life adversities. The setting of character gaps should be as prominent as possible to provide the characters with a precise orientation, so that the characters are oriented in the process of creating confrontation with the consequent transformation of their character, values, and attitudes.

In FEELING THROUGH, the hero comes out with a message telling the audience that he is a homeless youth in dire straits. He seeks metro cards from passersby to take public transportation, seeks to borrow a room from a friend but is rejected, plays games with friends on the street, squats on the side of the road to eat cheap food, and has nowhere else to go. The character's adversity of being homeless is set up to reinforce the character's gaping hole. And immediately after he refused a beggar to find him for money, as well as later stole the blind man's money can reflect the character of indifference, sinister and "bad", here the bad is not really bad, but the weakness of human nature. Because the back from his patience to help the blind can see the goodness of his character, the character at the beginning of the character defects and the life background of the defects of the reinforcement, as a foothold in the development of the character, so that the main character with the development of the story slowly change, for the character arc to show the light of the good enough to prepare.

The second is to strengthen the explosive power of the characters. Because the length of the microfilm is too short, the character's space is too small, if the character can't complete the dramatic action in time with a strong explosive force will not be able to move towards the end of the story under the dual role of internal and external conflicts, the character can't be transformed.

In the short film movie HEAR ME, Kennedy, a girl with selective mutism who is unable to speak due to the grief of losing her mother in childhood, has the ability to speak. She starts practicing her speech in order to be able to attend an interview at the college of her choice. In a silent performance, the teacher reads from her diary and she writes what she wants to say on stage. Along with the memories of Kennedy's outbursts of tears and accelerated writing speed when she writes in her diary, although it is silent, it enables the audience to fully feel Kennedy's inner struggle and outbursts. It is after such outbursts that the character is able to reconcile with her past self towards the end where she finally manages to speak, and this is where the character arc shines the brightest.

III. Conclusion

Through the study, we can conclude that under the limitation of time duration, strengthening the climax part of the plot and the key moments of the characters in the narrative, and strengthening the gaps as well as the explosive power of the characters in the character setting can compress the time needed for the character arcs, so as to overcome the problem of insufficient character arc shaping brought about by the time limitation of the short film.

Whether it is the enhancement of character gap and explosiveness or the enhancement of the climax plot, the character arc can be shaped in a short period of time. However, it should be noted that not all short films necessarily need excellent character arcs. In romantic or horror films, the audience cares more about the atmosphere and feeling expressed in the movie than the characters, and at this time, character arcs are not so important. However, in growth and inspirational films, the character's growth trajectory cannot be separated from the character arc, even though it is a compressed character arc, it does not affect our ability to create one memorable three-dimensional character after another.

References

[1] Johnson D. “Fantagonism, franchising, and industry management of fan privilege”, [M]. The Routledge companion to media fandom. Routledge, 2017: 395-405.

[2] Geoff King. “American independent cinema”, [M]. London: IB Tauris, 2005.

[3] Kotler P, Keller K L, Ang S H, et al.“Marketing management: an Asian perspective”, [M]. Harlow: Pearson, 2018.

[4] Liu Ziyuan. (2013). On the Creative Techniques of Film Screenwriting: "Character Arcs" (Master's thesis, Northeast Normal University). Retrieved from https://kns.cnki.net/KCMS/detail/detail.aspx?dbname=CMFD201401&filename=1013363739.nh

[5] Zou Jialu. (2020). Research on the Creative Methods of Character Arcs in Micro Films (Master's thesis, Shanghai Normal University). Retrieved from https://kns.cnki.net/KCMS/detail/detail.aspx?dbname=CMFD202002&filename=1020807952.nh

[6] Xiao Renfeng. (2021). The Composition and Setting of Character Flattening in Short Film Stories. Journal of Nanjing University of Aeronautics and Astronautics (Social Sciences) (02), 98-102+114. doi:10.16297/j.nuaass.202102017.

Interdisciplinary Development Trends and Research Hotspots in Digital Humanities and Film: a Bibliometric Analysis from 2000 to 2023

Lin, Fang Ru
Design Department
Dongseo University
Busan, South Korea
lfr19910929@gmail.com

Kim, Hae Yoon
Design Department
Dongseo University
Busan, South Korea
haeyoonkim@empal.com

***Abstract*—In order to understand the latest research trajectories of digital humanities and film studies, this article uses bibliometric methods to analyze and visualize the literature of Web of Science (WoS) and Scopus from 2000 to 2023. Through quantitative literature analysis, the study outlines key hot spots, traces evolutionary paths, and identifies popular trends in digital humanities and film studies. This study provides a comprehensive overview of the main research priorities and outlines the dynamic evolution of the field. (1) The development of digital technology is leading film analysis from traditional methods to data-centric methods. (2) The integration of different disciplines such as computer science and psychology enriches and expands insights into film. (3) Nonetheless, the transformative impact of digital technologies extends beyond the realm of film, creating new opportunities and challenges for the broader humanities and social sciences.**

Keywords—Digital humanities, film, bibliometrics, visual analysis

I. INTRODUCTION

With the rapid advancement of digital technology, the digital humanities have emerged as a significant area of interest within the humanities and social sciences. Merging knowledge from computer science, information technology, and the humanities, digital humanities seek to probe the influence of the digital era on culture, history, literature, and other disciplines. Concurrently, film, an essential medium of cultural expression, has been deeply transformed by digital technology. Innovations in digital filmmaking, storage, distribution, and film analysis present both unparalleled opportunities and challenges for film scholars. As a consequence, the nexus between digital humanities and film studies has emerged as a cross-disciplinary area of immense interest.

However, despite the notable strides in both digital humanities and film studies, there has been a paucity of intersecting research between the two. Prevailing literature predominantly centers on the deployment of digital technologies in film creation and distribution, as well as the digital evaluation of film as a cultural artifact. Yet, exhaustive research delving into how digital humanities methodologies can enhance film studies, and how films can offer fresh research avenues for digital humanities, is notably sparse. This underscores the imperative for rigorous exploration to unearth the latent ties binding digital humanities and film studies and to delve into how bibliometric methodsologies might deepen comprehension of this confluence.

The objectives of this study, geared toward elucidating the interconnections and trajectories between digital humanities and film studies via the bibliometric methods , are as follows: (1) Systematic Analysis: Assess the prevailing state of research in the realms of digital humanities and film studies by amassing, categorizing, and scrutinizing pertinent literature. This encompasses gauging shifts in literature volume, the metamorphosis of research themes, and the prevalence of pivotal concepts; (2) Identification of Research Hotspots: Leveraging the bibliometric methods, pinpoint the research epicenters and avant-garde themes within the domain of Digital Humanities and Film Studies, thereby illuminating scholarly predilections and concerns; (3) Exploring Cross-Influences: Dissect the symbiosis between digital humanities and film studies to ascertain how the former's methodologies can augment film studies and how film can offer novel research paradigms for digital humanities.

II. RESEARCH METHODS AND DATA SOURCES

A. Research Methods

Bibliometrics is a measurement method used to describe and analyze the dynamics and evolution of a specific discipline or research field. With the aid of modern computer technology, the results of bibliometric analysis can be represented visually using a well-defined knowledge map. This method frequently creates knowledge maps based on parameters such as "countries, institutions, authors, and keywords," visually representing the network structure to quantitatively detail the characteristics of data relationships.[1] Several tools, including CiteSpace, Vosviewer, and Ucinet, are widely used for knowledge mapping and bibliometrics. Among these, CiteSpace, a literature visualization and analysis software, was developed by Chaomei Chen's team at Drexel University in the United States. It is adept at showcasing the structure, evolution, and collaboration within research fields. Moreover, it evaluates the current research status in a field and forecasts emerging trends using visualization techniques. For this study, CiteSpace 6.2.R4 is employed to visualize and analyze international literature pertaining to "digital humanity" and "film."

B. Data Sources

This research was supported by BK21 Four Service Design driven Social Innovation Educational Research Team in Dongseo University.

This study utilized the Web of Science database for literature collection and searching, specifically selecting the SSCI and A&HCI databases. Additionally, the Scopus database was employed, which is recognized as an authoritative indexing tool for humanities, social sciences, and arts literature, making it an ideal source for bibliometric data. The search parameters were set as TS = ("Digital Humanities" OR "Digital" OR "Quantitative" OR "Quantify") and TS = ("Film" OR "video" OR "film style"), covering the time interval from January 1, 2000, to December 31, 2023. This search yielded 7,076 relevant documents. From these, we filtered out results to include only the "Article" literature type, reducing the count to 6,453. Within WoS categories, there were 360 articles categorized under "Film Radio Television," 174 articles under "Computer Science Interdisciplinary Applications," and 148 articles under "Humanities Multidisciplinary," amounting to a subtotal of 590 articles. Scholars involved in the study reviewed the preliminary data, excluding articles unrelated to "Digital Humanities" and "Film." Articles pertinent to the theme but with incomplete titles were also identified. From this process, a total of 308 irrelevant and duplicated articles were removed, resulting in 282 valid documents. Using the same search and refinement criteria on Scopus yielded an additional 261 documents. Combining the data from both sources, we obtained a comprehensive dataset of 543 documents, which served as the foundation for the bibliometric analysis.

The 543 papers used in this study were written by 637 authors from 316 organizations in 7 countries, published in 190 journals, and cited 2796 references from 2557 journals.

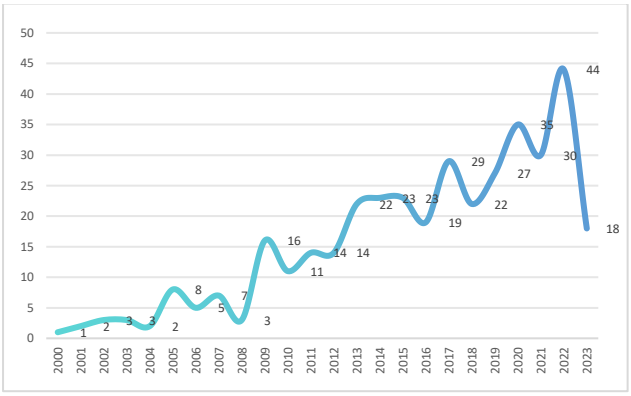


Fig. 1. Trends in the number of publications (2000-2023)

As illustrated in Figure 1, research related to "digital humanities" and "film" demonstrates a stepwise growth across distinct phases: 2000-2008, 2009-2012, 2013-2016, and 2017 to the present. In the 2000-2008 phase, "digital humanities" research was still in its infancy, with scholars beginning to explore how digital technology could bolster humanities research, focusing on areas like film archives, electronic libraries, and other R&D projects. The annual publication rate during this period averaged fewer than 4 articles. Between 2009 and 2016, the domain witnessed rapid growth, with the average annual article count tripling to 12. This surge underscored a growing recognition of digital technology's significance in humanities research, leading to increased attention to diverse theories and methodologies. Notably, text analysis and computer linguistics emerged as primary focal points. Between 2009 and 2012 alone, the annual publication count nearly doubled to 22. This period also marked an uptick in academic and societal appreciation for "digital humanities," prompting several universities to launch related academic programs. From 2017 onward, the field has seen consistent growth, approximating a linear trend, with an annual average of 37 articles. The realm of digital humanities continues its expansion, transitioning from foundational explorations to more encompassing roles in teaching, research, and societal applications.

III. DESCRIPTIVE STATISTICS

A. Bibliometric Analysis of Journals

The distribution of literature across journals and their respective publication volumes reflects shifts in the knowledge structure within a field. This not only equips scholars with efficient references to scientific research outcomes but also holds substantial importance in identifying research topics intrinsic to the field. Figure 2 enumerates the top 15 journals based on their literature contributions. Among these, seven journals boast an impact factor greater than 1, signifying that the core journals within this set uphold rigorous academic standards and offer high scholarly value.

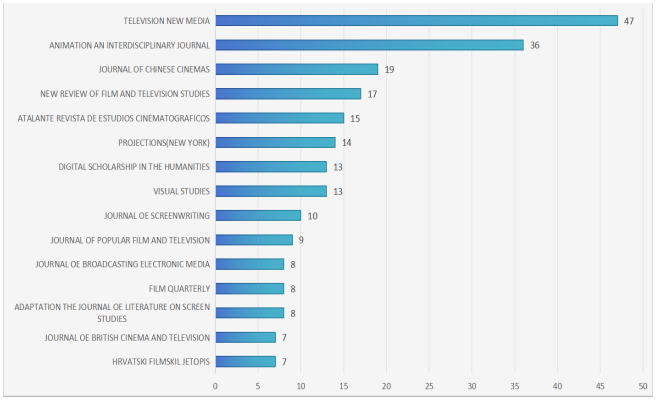


Fig. 2 Distribution diagram of the top 15 journals in the field of “digital humanities and film” research

The journal landscape for "Digital Humanities and Film" is diverse with a total of 190 publications. Of these, 29 journals have published more than two articles, eight have more than three, and six journals have contributed over six articles each. The TELEVISION NEW MEDIA journal stands out with 47 articles. From Figure 2, the distribution of these journals can be categorized into three core areas: research focusing on film and media, digital humanities, and cognitive science and psychology. (1) Film and Media-Centric Research: Journals such as New Review of Film and Television Studies and Television New Media primarily delve into film and media. However, their specific themes and areas of interest vary. For instance, New Review of Film and Television Studies often explores television and new media from diverse angles like historical, theoretical, and critical perspectives, including audience reception and the cultural implications of film and television series; (2) Digital Humanities Research: Notable journals in this category include Digital Humanities Quarterly and Digital Scholarship in the Humanities. Many of these are interdisciplinary, like the Digital Humanities Quarterly, which tackles intersections of humanities and computing. This journal, in particular, emphasizes computational

analysis of humanities data, media studies, and digital preservation; (3) Cognitive Science and Psychology Research: This segment is represented by journals such as Journal of Vision and Cognitive Science. These journals incorporate psychological and computational models to understand topics like visual neuroscience, perception, and visual cognition. Advanced tools and techniques in these areas are propelling in-depth analyses in the arts and humanities, introducing innovative solutions to longstanding challenges.

B. Bibliometric Analysis of Countries

Using CiteSpace to organize country information from the literature, we analyzed the contributions from various foreign countries. Our findings indicate that the thrust of digital humanities-related research originates from 60 countries or regions. Figure 3 illustrates 58 nodes and 56 links representing collaborating countries. Two primary timelines are discernible with "digital humanities" as the central keyword. In Timeline 1, initial research began in the United States and gradually branched out to several European and Asian countries, including the Netherlands, Italy, Belgium, China, France, and South Korea. The second timeline, predominantly led by the United Kingdom, demonstrates significant collaboration among European nations, encompassing countries such as Germany, the Netherlands, Estonia, and even Australia. Table 1 showcases the top 5 countries based on their publication counts, detailing the total number of publications and the inaugural year of research for each country.

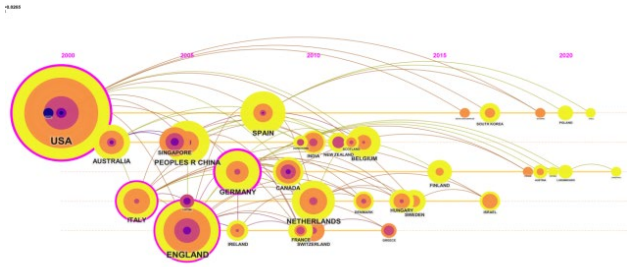


Fig. 3 Timeline of national or regional publishing networks

TABLE I. TOP 5 COUNTRIES IN THE EDUCATION BIG DATA RESEARCH FIELD

Rank	Country	Publications	Average Citation	Start Year
1	USA	95	0.27	2000
2	ENGLAND	44	0.16	2005
3	PEOPLE R CHINA	22	0.08	2005
4	SPAIN	22	0.10	2008
5	GERMANY	21	0.28	2007

The United States is the number one country for total publications in the digital humanities. In 1964, the University of Virginia established the Center for Anthropology and Digital Computing (Humanities and Social Sciences Online (HSSO)) to promote digital research in the humanities, which has led to an increasing number of scholars in the local area to pay attention to and participate in research related to the digital humanities. humanities-related research]. In 1993 the Humanities and Computing (Humanist) email list was established and became the main platform for digital humanities scholars to exchange ideas and projects. After that, many universities and institutions also participated in related research and construction, opening in the United States, Canada and Europe, such as the University of Virginia, the University of California, Stanford University, Harvard University, etc. opened a department of digital humanities, and carried out a lot of cross-disciplinary digital projects in a variety of academic fields such as history, archaeology, art history, and so on. The United Kingdom, Spain, Germany, and other countries have made significant contributions to the field of digital humanities. Digital humanities research topics cover a wide range, including text digitization, digital libraries and archives, digital cultural heritage, virtual reality, computational humanities, etc. As a branch, or a cultural carrier, digital humanities and imaging will continue to make new progress and breakthroughs globally with the continuous development of technology and the integration of disciplines.[2]

B. Bibliometric Analysis of Authors

Prolific authors are the core strength of the work in the research field, while highly cited authors represent significant influence in the field. To determine the core strength of digital humanity research field. Table 2 lists the volume, and starting year of the publication contribution of the top ten scholars in the foreign publication contribution rankings. Based on the Locarte theory in bibliometrics, half of the papers in the same topic are authored by a group of highly productive authors, a set of authors that is quantitatively approximately equal to the square root of the total number of all authors.[3]

$$\sum_{m+1}^{I} n(x) = \sqrt{N}$$

m=2 I=7 m+1=3 3-7 n（3-7） 30

where n(x) denotes the number of authors who have written x papers, I = is the number of papers by the most productive authors in the field, N is the total number of authors, and m is the minimum number of publications by core authors. , the minimum number of publications by core authors N = 0.749 × ɳ Max½ (ɳ Max i.e., the number of publications by the most prolific authors), and ɳ Max = 21, which yields N = 4. Thus, authors with 4 or more publications are positioned as core authors in the field, totaling 10 core authors, which account for a total of Approximately 6.45%, less than one-tenth of the core number, totaling 83 publications, accounting for 21.8% of the total number of publications. As shown by bibliometrics, the total number of authors who have published literature on the topic of digital humanity in the last 20 years is 155, of which the scholar with the most publications is Cutting, JE, with a total of 21, whose research interests are mainly in the areas of visual perception, cognitive psychology, and the intersection between psychology and art. He has a wide range of research interests in the field of visual perception and psychology, with a particular focus on time perception, dynamic vision, film and art, The rest of the top ten scholars were Salt, B, Tsivian, Y, Benini, S, Redfern, N, Savardi, M, Signoroni, A, Svanera M, Candan, A, Pearlman, K.

TABLE II. TOP 10 MOST PROLIFIC AUTHORS OF INTERNATIONAL DIGITAL HUMANITY RESEARCH, 2000-2023

Rank	Name	Publications
1	Cutting, JE	21
2	Salt, B	12
3	Tsivian, Y	11
4	Benini, S	9
5	Redfern, N	9
6	Savardi, M	5
7	Signoroni, A	4
8	Svanera, M	4
9	Candan, A	4
10	Pearlman, K	4

Cutting JE is a Professor of Psychology at Cornell University, where his research focuses on exploring the interactions between film editing techniques and human psychological and emotional responses in response to the psychological and emotional responses of film editing, and has conducted numerous studies in collaboration with his student DeLong, J. E.. They have contributed extensively to the field of psychology through their intensive research efforts.[4]

IV. Visual Analysis and Discussion

A. Cluster Analysis on Keywords

Keywords summarize the research theme, key contents and technical methods of the article, which is a high degree of refinement of the author's research focus. Knowledge network analysis of the keywords can be used to access the specific research content and flow of research in the field of "digital humanity". The international digital humanity research clustering network map consists of 349 nodes and 558 links, with a network density of 0.0092. Among them, the modularity value of the clustering view Q (Modularity) and the average silhouette value S (Mean Silhouette) as an important index for evaluating the effect of the view drawing, $Q > 0.3$ indicates that the structure of the clusters is significant, $S > 0.5$ indicates that the structure of the clusters is significant, and $S > 0.5$ indicates that the structure of the clusters is significant. $Q > 0.3$ indicates that the clustering has a significant structure, $S > 0.5$ that the clustering is reasonable and effective, and $S > 0.75$ that the clustering has a high degree of confidence. The study clusters the keywords of the literature with LLR algorithm, and finally arrives at the keyword clustering map Figure 7, the specific clustering information and the keyword information within its clusters are shown in Table 5. Its $Q=0.7845 > 0.5$, the clustering structure is significant, $S=0.9231 > 0.75$, the clustering results are highly credible. The clustering map is divided into 10 clustering taxa. They are #0 film style, #1 gender, #2 file editing, #3 narrative, #4 statistical analysis, #5 computer vision, #6 arts computing, #7 digital humanity, #8 music, #9 social network analysis.

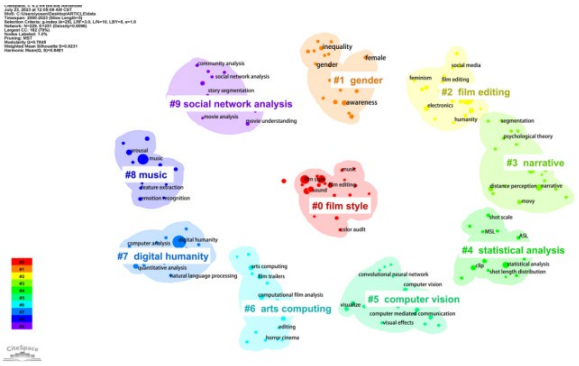


FIG. 4 Keyword clustering mapping for digital humanity research

TABLE III. KEYWORD CLUSTERING CONTENT TABLE FOR INTERNATIONAL DIGITAL HUMANITY RESEARCH 2000-2023

Rank	Cluster Name	Cluster Content	Cluster Size
0	film style	film style ;film editing ; sound ; music ; color adult	27
1	gender	gender ; female ; inequality ; electroencephalography ; awareness	20
2	film editing	film editing ; social media ; electronics ; humanity ; feminism	19
3	narrative	narrative ; segmentation ; psychological theory ; movy ; distance perception	19
4	statistical analysis	statistical analysis ; shot length distribution ; ASL ; MSL ; shot scale ; clip	18
5	computer vision	computer vision ; visualize ; convolutional neural network ; computer mediated communication ; visual effects	18
6	arts computing	arts computing ; film trailers ; computational film analysis ; editing; horror cinema	17
7	digital humanity	digital humanity ; quantitative analysis ; natural language processing ; computer analysis ; social media	16
8	music	music ; arousal ; feature extraction ; emotion recognition ; time perception	14
9	social network analysis	social network analysis ; movie analysis ; movie understanding ; story segmentation ; community analysis	14

By analyzing the keywords with high frequency and high centrality in Figure 4 and the keyword clusters and information in Table 3, the three main hot research topics in digital humanities are as follows:

(1) Digital technology-oriented research . The methods and tools of digital humanities provide a new analytical approach to film studies, In recent years, many new digital methods and tools have emerged in digital humanities research. The keywords clustering cluster #4 statistical analysis,#5 computer vision, #6 arts computing,#7 digital humanity are all keywords related to digital technology. The high frequency and high centrality keywords of digital technology are as follows: deep learning, statistical analysis, content analysis, information retrieval, convolutional neural network, factor analysis, etc. analysis and so on. Therefore, how to effectively recognize and semantically analyze the text, music, and video contents in movies, and how to

continuously improve the multidimensionality of the research object and the efficiency and precision of the research have become an important part of digital humanities research. In addition, some scholars are committed to applying these methods to build digital heritage and digital programs for better inheritance and preservation. Some scholars are also interested in how to use data visualization and virtual reality to display complex data and information of humanities research for better understanding and communication.[5]

(2) Interdisciplinary Research on Digital and Film. Digital technology has brought new perspectives and approaches to the study of humanities, utilizing new media technologies, mining textual data for explicit and implicit problems, and then proposing new solutions and research paradigms. The keyword clusters #0 film style, #3 computer vision, #8 music are the keywords related to the content of humanities. High frequency and high centrality keywords for the humanities are as follows: film, movie, article, television, cinema, rhythm and so on. Researchers work on converting printed text, handwritten text, etc. into computable digital form and use techniques such as natural language processing and semantics to extract semantic information from them in order to better understand and interpret the content of the text. Thus, computations related to the arts, such as computational analysis of films, understanding of movies, analysis and segmentation of stories. Feature extraction for art such as music.[6]

(3) Digital society-oriented research . With the popularity of social media, digital humanities also focuses on the analysis and research of social media data. By mining social media data, researchers explore social networks, user behavior and information dissemination in digital society . Keyword clustering cluster #0 gender, #9 social network analysis are all keywords related to the content of digital social crowd. The high frequency and high centrality keywords about digital society people are as follows: motion pictures, emotion, male, social media and so on. As a result, the research and analysis are related to the attributes of emotions, preferences and behaviors of people in social media. For example, the application of natural language models to analyze the behavior, expressions and emotions of people. We also focus on the role of digital technologies and social media in shaping modern society and human interaction. Related research includes social network analysis, user behavior studies, social media, digital inequality, social interaction and virtual space.

B. Time Zone Analysis on Keywords

The visualization of time zones can clearly show the evolution of knowledge over time. Figure 5 shows the high-frequency keywords in digital humanity studies in different time zones, demonstrating the focus of scholars' research, which varies from one stage to another. In each time zone, the font size indicates the frequency of keyword occurrence. That is, the larger the font size, the more frequently the keyword appears. The connecting lines indicate the inheritance between research hotspots. Combined with the most emergent terms in each time zone, the evolution and development process of each phase will be identified through Figure 6, based on the above division of research phases.

Fig. 5 Keyword time zone map of digital humanity research

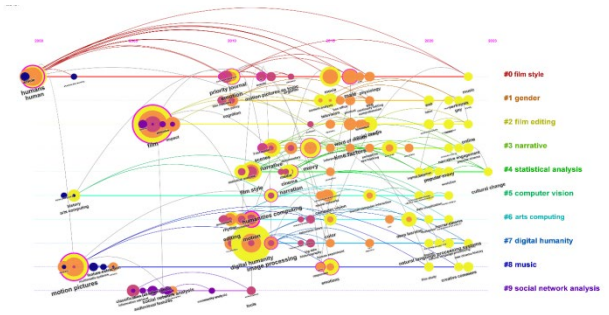


Fig. 6 Keyword timeline map of digital humanity research

In 2000-2008: although scholars in some areas have carried out more sustained study and practice, research in digital humanities as a whole is still in the early stage of exploration and development. The keywords with the highest intensity in this stage are film, information retrieval; the most cited keywords are human, motion picture, classification of information , audiovisual analysis. in the field of digital humanities, film studies is a very important and colorful field. At this stage, relevant research is gradually being established. As a cultural phenomenon and a form of entertainment, movie has far-reaching influence in society, history and culture. Digital humanities researchers utilize the intersection of computer science and the humanities in film studies, exploring aspects of film works, the industry, and audience response through large-scale data processing, text analysis, information retrieval, and other technical means.

In 2009-2012:In this phase, Digital Humanity gradually attracted the attention of more scholars. The research field was extended to data mining, text analysis, visualization and virtual reality. The keywords with the highest intensity in this phase are film style, motion; the most cited keywords are cognition, humanities computing, image processing, motion. scholars began to apply computer technology and data science methods to mine meaningful information from large-scale text and data. Scholars are beginning to apply computer technology and data science methods to mine meaningful information from large-scale text and data. Their specific research includes the application of text analysis to analyze texts such as movie scripts, commentaries, and reviews to reveal the themes, plot development, and character relationships in the movies, so as to gain a deeper understanding of the connotations and composition of the movie works; it can also be based on the large amount of data on social media, such as Facebook and Twitter, to study

the reactions of moviegoers. By analyzing users' discussions and interactions on social media, it is possible to understand the popularity and influence of a movie as well as the emotional tendencies of the audience.

2013-2016: digital humanities gained wider academic recognition and became an important area of research at the intersection of humanities and computer science. The keywords with the highest keyword intensity in this phase are male, female, neural networks; the most cited keywords are word of mouth, time factors, eye tracking ,human computer interaction. in the subject matter of cinema studies, the contradiction between men's and women's power Issues have a long history in humanities research, and the field of digital humanities allows for a more comprehensive understanding of the representation and impact of film as a form of cultural expression on gender and social issues. As a medium of wide dissemination, film plays an important role in shaping and changing social perceptions and values. Through the exploration of gender studies, feminist and patriarchal issues in movies, we can better understand and think about the topics of gender equality and social justice. Its specific research such as the analysis of gender roles, the exploration of women's rights and feminist theme films , the patriarchal system and family relations theme films, the director-producer and the gender inequality data behind. In terms of data collection and analysis of moviegoers, in order to more effectively obtain the emotional data of moviegoers, the quantitative analysis of the research gradually introduces the measurement of physiological indexes, shifting from the quantification of human words and language to the quantification of human physiological data, such as eye-tracking technology, EEG technology and so on.

Along with digital humanities continues to evolve to this day, becoming an important part of the humanities. Research has gradually expanded to include online analytics, digital socialization, and cultural interaction. The keywords with the highest keyword intensity in this phase are natural language processing, news, communication; the most cited keywords are narrative engagement, online, cultural change and so on. With the maturity and application of artificial intelligence, natural language modeling and other technologies, the openness and public participation based on digital humanities have been emphasized more, scholars and the public continue to share research results and data, more and more multi-dimensional information began to be synthesized and analyzed and integrated, and more and more in-depth content began to be mined. There is a move away from traditional linear narratives towards narratives that are multidimensionally fused, including the fusion of multimedia, metadata, and static environments. Examples include the " The Valley of the Shadow" project at the University of Virginia, and the " Vectors Journal of Culture and Technology in a Dynamic Vernacular" project. More, such as cross-cultural communication and cultural change brought about by online socialization, and online sharing of creativity have also been increasingly emphasized by scholars.

V. Conclusion and Limitation

A. Conclusion

This article uses CiteSpace 6.2.R4 to analyze relevant literature on the interdisciplinary subject of digital humanities and film studies in the literature database wos and scopus from 2000 to 2023, conducts a basic analysis of the number of publications, Journals, Countries, and Authors, and conducts a keyword analysis Visual display - keyword co-occurrence and keyword time zone analysis, the research hot spots and development trends in this field are as follows:

The significant stepwise growth trend of published articles from 2000 to 2023 highlights the growing academic attention to "digital humanities and film studies", especially in fields as diverse as film, media, digital humanities, and cognitive sciences Making key contributions to the intersection, this growth is particularly evident in regions such as the United States and the United Kingdom, which have championed the establishment of digital humanities platforms through the establishment of dedicated research centers. The academic contributions of outstanding individuals such as Cutting, JE, Salt, B, and Tsivian, Y, have become pioneers in interdisciplinary research.

Current research hotspots illuminated by keyword co-occurrence cover areas such as data mining, text analysis, visualization, deep learning, artificial intelligence, and the use of digital tools to protect cultural heritage. This broad scope is reflected in the integration of disciplines, emphasizing the digital transformation of the humanities and leveraging technology to enhance public engagement and openness. The development of digital humanities from 2000 to 2008 was marked by fundamental research, primarily focused on film studies through digital tools. From 2009 to 2016, the field expanded its scope to incorporate computational techniques for data analysis, highlight social issues such as gender roles, and integrate physiological audience data for deeper insights. Post-2016 is characterized by the integration of advanced technologies such as natural language processing with film narratives, providing a more comprehensive perspective for analyzing film narratives and their social impact.

The future of digital humanities and film studies promises to be rich in technological innovation, interdisciplinary collaboration, and public engagement. This evolution will not only change the way we study and understand film, it will also reshape cultural narratives and the way the public engages with film.

B. Limitation

This chapter employs the Citespace visualization software for bibliometric analysis of literature in the digital humanities field. Despite its comprehensive approach, this study has limitations. It relies on keyword-based retrieval, potentially omitting some gray literature. Publications were sourced exclusively from WoS and Scopus databases; while reputable, they may not encompass all relevant articles. Additionally, while CiteSpace standardizes terminology and author names, its processes might exclude certain small samples. Nonetheless, these limitations are unlikely to significantly affect the overarching findings.

Acknowledgment

This research was supported by BK21 Four Service Design driven Social Innovation Educational Research.

REFERENCES

[1] Merigó J M, Gil-Lafuente A M, Yager R R. An overview of fuzzy research with bibliometric indicators. Applied Soft Computing, 2015, pp.420-433.
[2] Fricke-Steyer H. Burdick, Anne/Drucker, Johanna/Lunenfeld, Peter/Presner, Todd/Schnapp, Jeffrey (eds.)(2016): Digital. DHd 2019, 278.
[3] Friedman A. The power of Lotka's law through the eyes of R[J]. Romanian Statistical Review, 2015, 63: pp.69-77.
[4] Cutting J E, DeLong J E, Nothelfer C E. Attention and the evolution of Hollywood film. Psychological science, 2010, 23: pp.432-439.
[5] Li L, Goh T T, Jin D. How textual quality of online reviews affect classification performance: a case of deep learning sentiment analysis. Neural Computing and Applications, 2020, 32: pp.4387-4415.
[6] Manovich L. Computer vision, human senses, and language of art. AI & SOCIETY, 2021, 36: pp.1145-1152.

Machine Learning-Based Detection of Hazardous Road Objects

Tee Yi Loong
Department of Computing and Information System
Sunway University
Bandar Sunway, Malaysia
tyloong57@hotmail.com

Athirah Mohd Ramly
Department of Computing and Information System
Sunway University
Bandar Sunway, Malaysia
athirahr@sunway.edu.my
0000-0001-8282-297X

***Abstract*—This research aims to enhance pedestrian safety in low-light environments by developing a machine learning-based approach to detect dangerous road objects. The study utilizes image enhancement techniques and machine learning algorithms to improve the accuracy of object detection in low-light conditions. The motivation for this research stems from the high number of pedestrian-vehicle collisions, particularly during nighttime, and the limited research on utilizing machine learning for pedestrian safety in low-light environments. The proposed methodology involves training a machine learning model using the exclusively dark, ExDark dataset and you only look once version 5, YOLOv5 object detection framework. Prior to training, the dataset undergoes a transformation process using the attention-guided low-light image enhancement network, AGLLNet image enhancement technique to enhance image quality and visibility. The transformed dataset is then preprocessed and converted into the YOLOv5 PyTorch format using Roboflow. The results of the study demonstrate the model's decent performance in predicting dangerous objects, with precision, recall, and mean average precision scores varying for different classes such as bicycles, buses, cars, motorbikes, and people. However, the model exhibits limitations in distinguishing between similar or overlapped objects, suggesting the need for further improvements such as data augmentation and architectural refinements. Future work in this area could explore advanced deep learning architectures, incorporate additional data sources, conduct experiments on larger datasets, and collaborate with transportation authorities to integrate the developed model into existing road safety systems. Ultimately, this research contributes to the development of a machine learning-based approach for detecting dangerous road objects in low-light environments, with the potential to enhance pedestrian safety and reduce the number of pedestrian-vehicle collisions.**

***Keywords*—Machine learning, Object Detection, ExDark, YOLO, AGLLNet**

I. INTRODUCTION

Pedestrian-vehicle collisions have been causing significant injuries and fatality starting from years back then. According to statistics reported, 436,400 people were injured and 39,600 people were killed in pedestrian-vehicle accidents worldwide in the year 1997 alone [3]. It was the second largest sources of death and injury related to traffic [4]. Not only that, the number of vehicles on the roads is increasing every year as the world population is increasing. According to L. David Roper, there are approximately 1.446 billion cars in the year 2022 where about 17.7% of the people in the world have a car, and it is not slowing down anytime soon [5]. Accident rates might increase at night as the visibility of the road is reduced significantly. Other factors may also lead to pedestrian collision such as pedestrians with dark clothing at night, dazzling light from cars coming from opposite, bad weather conditions, poor condition of bulbs and headlamps, lack of traffic lights, etc. [8]. According to data from National highway Traffic Safety Administration, between the year 2010 and year 2015, majority of the fatal pedestrian accidents occurs during 6pm to 12am and when light conditions are dark which take up to 72% of the total pedestrian fatalities in those years [2].

The topic object detection has been popular in recent years, it plays a vital role in computer vision, which is a sub-field of artificial intelligence that derives and manipulates images and digital videos [1]. Computer vision and object detection has been widely utilized in many areas such as autonomous vehicles, face detection, intelligence video surgery, and other more to provide automated service, improving security, as well as providing ease to humans. In the pedestrian safety domain, object detection can be utilized in low light environments to detect dangerous objects such as moving vehicles which can pose a danger to pedestrians on darker roads or vice versa. Hence, this project is intended to utilize image enhancement technology and machine learning algorithms to help to detect dangerous objects like vehicles during nighttime or in low light environments in order to improve the safety of road users and reduce accidents rates during the night. Next, this research also tends to contribute to the growth of low-light domain research by expanding into the pedestrian safety domain.

There is limited research made on improving pedestrian safety under low-light environment using machine learning. Challenge persists when image with noise, low quality, and low illumination are affecting the accuracy of object detection algorithms. Many of the studies proposed methodology regarding pedestrian-collision in recent years are only applicable on autonomous vehicles, mainly cars. Not only that, but the previously proposed methods are also generally complex and require high costs to be implemented. Hence, it is not suitable to be widely implemented.

A. *Objectives*

- Address the limited attention given to road safety in previous research on low-light object detection by extending the focus beyond autonomous vehicles to include various vehicle types, such as bicycles, motorcycles, and others, in the context of pedestrian collisions.
- Propose and design a combination of existing image enhancement and object detection techniques, utilizing the convolutional neural network, CNN algorithm for easy implementation, to provide a cost-efficient and easy-to-understand low-light object detection method for pedestrian safety.
- Evaluate the accuracy of the proposed model upon implementation to assess its effectiveness in detecting dangerous objects, including different vehicle types, in low illumination environments.
- Contribute to reducing the number of pedestrian collisions by improving the capability of low-light object detection and enhancing road safety across various vehicle categories.

II. METHODOLOGY

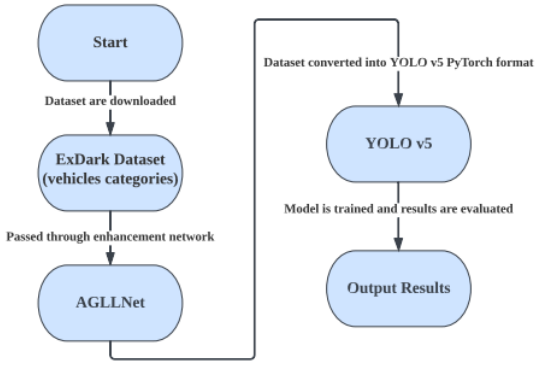


Fig. 1. Implementation steps of this study

This study begins by acquiring the dataset. The dataset that will be used in this study is a publicly available exclusively dark, ExDark dataset published by [6]. The dataset consists of 12 object classes, namely table, people, motorcycle, dog, cup, chair, cat, car, bus, bottle, boat, and bicycle, summing up to a total of 7363 images. The low-light images were taken from a variety of sources such as websites and search engines, namely Google Search, Gettyimages.com, Deviantart.com, Imgur.com, Photobucket.com, and Flickr.com. some images from public datasets like Microsoft common objects in context, COCO, ImageNet, Pascal visual object classes, VOC, and other datasets were also sub-sampled. Not only that, different models of digital cameras and smart phones were also used to manually capture low-light scenes from movies collection. Next up, class annotation and bounding box annotation of the images were then labelled using Piotr's Computer Vision MATLAB toolbox. Some of the images contain single, multiple, or different instances. The dataset also identified 10 different types of low-light conditions, namely twilight, shadow, window, screen, strong, weak, single, object, ambient, and low. To ensure relevancy, a selection process is applied to filter and choose suitable vehicle categories and images from the dataset.

Subsequently, the selected dataset undergoes image enhancement using the Attention-Guided Low-Light Image Enhancement Network, AGLLNet method. It is an end-to-end attention-guided method based on multi-branch convolutional neural network to learn the mapping from low-light images to normal light ones. It is an improved version of multi-branch low-light enhancement network, MBLLEN, with new design techniques such as two attention guidance, reinforce-net, as well as a new and more complete low-light data synthesis pipeline [9]. One key difference between AGLLNet and MBLLEN is the attention-guided enhancement. AGLLNet incorporates attention maps to guide the enhancement process in a region-adaptive manner. It learns two attention maps: the underexposed, UE attention map and the noise map. The UE-attention map distinguishes underexposed regions from well-lit regions, allowing the method to focus more on enhancing the underexposed areas. On the other hand, MBLLEN does not utilize attention maps for guiding the enhancement and denoising tasks. AGLLNet also introduces a reinforcement-net in the final step to further enhance the color and contrast of the output image. This additional network improves the overall quality of the enhanced image in terms of color reproduction and contrast enhancement. In contrast, MBLLEN does not incorporate a specific module for color and contrast enhancement. To address the lack of large-scale paired low-light image datasets, AGLLNet proposes a low-light image simulation pipeline. This pipeline is used to synthesize realistic low-light images with well-exposed ground truth images. The synthetic dataset generated by AGLLNet is larger and more diverse than the dataset used in MBLLEN. Having a more extensive data set aids in training a more effective network. AGLLNet demonstrates superior performance compared to MBLLEN both quantitatively and visually, as shown by extensive experiments on multiple datasets. The proposed attention-guided enhancement, combined with the reinforcement-net and the large-scale dataset, contributes to a significant margin of improvement over the state-of-the-art methods [7]. By applying this technique, the quality of the dataset images is improved, resulting in enhanced object detection performance.

To facilitate model training, the dataset is converted into PyTorch format using Roboflow, a platform that streamlines the conversion process. The dataset can also be processed for tasks such as auto-orientation and checking dataset health. Next, the converted dataset is then utilized in a training environment like Google Collab, which comes pre-installed with Torch and computer unified device architecture, CUDA, providing the necessary tools for efficient training. One advantage of using Google Collab is its availability of graphical processing unit, GPU and tensor processing unit,

TPU hardware acceleration, which speeds up the training process.

Furthermore, you only look once version 5, YOLOv5 repository, along with its dependencies, is cloned and installed. YOLO is a real-time and unified object detection. YOLO can obtain boundary box and class probability directly from image pixels. This is very different from R-CNN which uses the combination of proposal and classification for detecting objects. Small model size and fast calculation speed is the core of YOLO object detection algorithm. This is because the straightforward structure of YOLO allows it to produce detection results directly from pictures input into the network. The fast detection speed of YOLO makes it suitable for real-time video detection. It is employed to train the model on the enhanced dataset. Customized parameter values, including image size, batch size, number of training epochs, dataset configuration path, model configuration, weights, and result naming, are set to optimize training performance.

During the training process, parameter tuning is conducted to fine-tune the model's performance. Various parameters are adjusted to achieve the best possible results. Lastly, performance metrics such as generalized intersection over union, GIoU, mean average precision, mAP, and others are evaluated to assess the effectiveness of the trained model.

III. RESUTLS AND DISCUSSION

Model summary: 157 layers, 7023610 parameters, 0 gradients, 15.8 GFLOPs

Class	Images	Instances	P	R	mAP50	mAP50-95
all	525	1803	0.865	0.692	0.777	0.497
Bicycle	525	211	0.863	0.72	0.789	0.489
Bus	525	111	0.92	0.856	0.885	0.703
Car	525	414	0.828	0.681	0.787	0.515
Motorbike	525	208	0.872	0.62	0.725	0.414
People	525	859	0.841	0.583	0.7	0.362

Fig. 2. Model summary for the trained model

Precision, P measures the accuracy of the model's positive predictions. In this case, the overall precision of 0.865 suggests that when the model predicts an object, it is correct around 86.5% of the time. This indicates that the model has a relatively low false positive rate, meaning it doesn't often incorrectly classify background regions as objects. Recall, R captures the model's ability to correctly identify positive instances. The recall value of 0.692 indicates that the model detects only 69.2% of the actual positive instances present in the dataset. In other words, there are instances that the model fails to detect, resulting in false negatives. This suggests that the model may miss some objects or struggle with detecting smaller or more challenging instances.

Mean average precision at 50% IoU (mAP50) provides a measure of the model's performance in localizing objects with a moderate level of overlap with the ground truth bounding boxes. The mAP50 value of 0.777 indicates that, on average, the model achieves a precision of 77.7% when the overlap threshold is set to 0.5. This means that the model can reasonably localize objects that have some degree of overlap with the ground truth, but there is room for improvement. Mean Average Precision at 50%-95% IoU (mAP50-95) offers a broader evaluation of the model's performance across a range of IoU thresholds. The mAP50-95 value of 0.497 suggests that the model struggles with objects that have high levels of overlap with the ground truth bounding boxes. This means that when objects are densely packed or occluded, the model may have difficulty accurately localizing and classifying them.

Analyzing the results for individual classes, we can see variations in performance. The model performs relatively well in detecting buses, with a precision of 0.92, recall of 0.856, and an mAP50-95 of 0.703. This indicates that the model can successfully identify and localize buses with high precision. On the other hand, the model struggles with motorbikes, as indicated by a lower precision of 0.872, recall of 0.62, and mAP50-95 of 0.414. This suggests that the model may have difficulty distinguishing motorbikes from similar-looking objects such as bicycles or may struggle with their smaller size. The model's performance for people is relatively lower, with a precision of 0.841, recall of 0.583, and mAP50-95 of 0.362. This indicates that the model faces challenges in accurately detecting and localizing people instances in the given dataset. People are a complex class with significant variations in appearance, poses, and occlusions, which can pose difficulties for the model's object detection capabilities.

An interesting observation is that while the AGLLNet image enhancement technique improves the precision of object detection by enhancing image illumination, some of the enhanced pictures may exhibit blurriness or pixelation. This could be attributed to the relative immaturity of the AGLLNet technique, as it was recently published. Alternatively, it might be an inherent characteristic of image enhancement techniques in general.

Image enhancement techniques aim to enhance the visual quality of low-light images by adjusting brightness, contrast, and other relevant visual features. These techniques are effective in improving object visibility, making objects more distinguishable and facilitating easier detection by the model. Consequently, the model trained on the image-enhanced dataset achieves higher precision, reducing false positives and making more accurate positive predictions.

However, there is often a trade-off in terms of recall. While image enhancement improves visibility, it can also introduce artifacts or alter the original characteristics of objects. These alterations can affect the model's ability to accurately detect objects, increasing the likelihood of false negatives or missed detections. Consequently, the recall value may decrease, as the model may struggle to detect certain objects that were present in the normal dataset but are more challenging to identify in the image-enhanced dataset.

IV. CONCLUSION AND FUTURE WORKS

In conclusion, this research successfully trained a machine learning model to detect dangerous road objects in low-light environments using an image enhancement technique, with a decent level of accuracy. The ExDark dataset was utilized for YOLOv5 object detection, following a transformation process using the AGLLNet image enhancement technique to improve image quality and visibility. The transformed dataset was then preprocessed and converted into the YOLOv5 PyTorch format using Roboflow. The results of the study demonstrated that the model achieved decent performance in predicting dangerous objects.

The model statistics showed that the overall precision (P) was 0.865, recall (R) was 0.692, mean average precision (mAP) at 50% intersection over union (IoU) was 0.777, and mAP from 50% to 95% IoU was 0.497. The model performed well in detecting specific classes such as bicycles, buses, cars, motorbikes, and people, with varying precision, recall, and mAP scores for each class. However, there are still areas that require improvement, particularly in scenarios where objects are similar or overlapped. The model showed limitations in distinguishing between similar objects, indicating a need for techniques such as data augmentation and architectural refinements to enhance its discriminative capabilities. Additionally, addressing the model's performance in detecting and localizing overlapping objects can be achieved through the application of post-processing techniques and the inclusion of augmented data with overlapping instances.

Overall, this research contributes to the development of a machine learning-based approach for detecting dangerous road objects in low-light conditions. The findings highlight the model's strengths and limitations, providing valuable insights for future enhancements. By addressing the identified limitations and incorporating suggested improvements, the accuracy and effectiveness of the model can be further enhanced, ultimately contributing to improved road safety in low-light environments.

Future work in this area of research holds significant potential for further advancements. To improve the model's performance, future studies could explore the integration of advanced deep learning architectures, such as object detection models with attention mechanisms or transformer-based models. These architectures have shown promising results in various computer vision tasks and may help address the challenges of distinguishing between similar and overlapped objects. Additionally, incorporating additional sources of data, such as thermal imaging or LiDAR, could enhance the model's ability to detect objects in low-light environments where visual information alone may be limited. Furthermore, conducting experiments on larger and more diverse datasets would provide a better understanding of the model's generalizability and enable its deployment in real-world scenarios. Finally, collaboration with transportation and urban planning authorities could facilitate the integration of the developed model into existing road safety systems, ultimately contributing to the reduction of pedestrian-vehicle collisions and enhancing overall road safety.

ACKNOWLEDGMENT

This work was supported by the Malaysian Ministry of Education Grant FRGS/1/2022/ICT09/SYUC/03/1. The authors would like to thank members of the INTERNAL GRANT SCHEME (IGS) 2023 GRTIN-IGS-DCIS-07-2023 for their contribution and collaboration.

REFERENCES

[1] Brownlee, J. (2019). *Deep learning for computer vision: image classification, object detection, and face recognition in python*. Machine Learning Mastery.

[2] Coleman, H., & Mizenko, K. (2018). Pedestrian and bicyclist data analysis (Research Note. Report No. DOT HS 812 502). Washington, DC: National Highway Traffic Safety Administration.

[3] G. Sala and D. Neunzig. Protector: Accident Analysis. 2000. Deliverable No. D02.1.

[4] Grubb, G., Zelinsky, A., Nilsson, L., & Rilbe, M. (2004) 3D vision sensing for improved pedestrian safety. In IEEE Intelligent Vehicles Symposium, 2004 (pp. 19-24). IEEE.

[5] Kemp, A. (2022). How many cars are there in the world? PD Insurance. https://www.pd.com.au/blogs/how-many-cars-in-the-world/.

[6] Loh, Y. P., & Chan, C. S. (2019). Getting to know low-light images with the exclusively dark dataset. *Computer Vision and Image Understanding*, *178*, 30-42. https://doi.org/10.1016/j.cviu.2018.10.010

[7] Lv, F., Li, Y., & Lu, F. (2021). Attention guided low-light image enhancement with a large scale low-light simulation dataset. International Journal of Computer Vision, 129(7), 2175-2193. https://doi.org/10.1007/s11263-021-01466-8

[8] Nowosielski, A., Małecki, K., Forczmański, P., Smoliński, A., & Krzywicki, K. (2020). Embedded night-vision system for pedestrian detection. IEEE Sensors Journal, 20(16), 9293-9304.

[9] Yu-Li. (2021). agllnet: Attention guided low-light image enhancement with a large scale low-light simulation dataset, IJCV 2021. GitHub. https://github.com/yu-li/AGLLNet

Unlocking Anomaly Dynamics: A Masked Autoregressive Model and GCN Fusion for Openset Video Anomaly Detection

Trisha Prasad
Centre for Digital Innovation
CHRIST University,
Bengaluru-560029, Karnataka, India

Harsha B M
Centre for Digital Innovation
CHRIST (Deemed to be University),
Bengaluru, India

Dr. Guydeuk Yeon*
Centre for Digital Innovation,
CHRIST (Deemed to be University),
Bengaluru, India
*Corresponding author:
yeon@christuniversity.in

***Abstract*— Video anomaly detection is pivotal in computer vision, especially for surveillance and security. This paper, "Towards Openset Video Anomaly Detection," advances existing literature with a novel model integrating Graph Convolutional Networks (GCN) and a refined Normalizing Flow for realistic anomaly detection. The model adapts to unforeseen scenarios, leveraging XD-Violence and UCF Crimes datasets. Enhancements to the Normalizing Flow architecture, transitioning from Inverse Autoregressive to Masked Autoregressive, improve the model's ability to capture complex temporal dependencies, enhancing anomaly detection. Emphasizing broader applications, beyond security to self-driven automobiles, business, IT, and application performance, XD-Violence and UCF Crimes datasets are chosen for realistic benchmarks in criminal activities. The roles of GCN and Normalizing Flow in Openset video anomaly detection challenges are discussed. The methodology draws from a pre-trained action recognition backbone, employing a C3D-like 3D ConvNet based on ImageNet-pretrained Inception-V1. Two similarity graphs capturing temporal relationships and feature similarity, along with triplet learning constraints, enhance discrimination between anomalous and normal classes. Integrating GCN refines anomaly detection by modelling complex relationships within video frames. Noteworthy enhancements to the Normalizing Flow model, particularly the Masked Autoregressive shift, are presented with a focus on computational implications. Aligning with intended usage, this shift optimises efficiency for openset anomaly detection.**

In conclusion, this study offers a comprehensive openset video anomaly detection approach, leveraging a pre-trained action recognition backbone and GCN for improved spatial-temporal understanding. The refined Normalizing Flow underscores commitment to enhancing anomaly detection frameworks, addressing computational trade-offs, and optimizing efficiency for realistic applications. Experimental results validate the proposed approach's efficacy in overcoming openset video anomaly detection challenges.

Keywords*— *Openset Video Anomaly Detection, Graph Convolutional Networks (GCN), Normalizing Flow, Masked Autoregressive Model, Spatial and Temporal Dependencies

I. INTRODUCTION

In recent years, video anomaly detection has emerged as a critical task in computer vision, finding applications in various domains such as surveillance and security. The paper titled "Towards Openset Video Anomaly Detection" addresses the challenge of realistic anomaly detection by training a deep neural network on specific anomaly types, mitigating the inadvertent omission of potential anomalies. The primary focus lies in datasets renowned for their portrayal of anomalous events, such as the XD-Violence Dataset and the UCF Crimes dataset.

The existing literature has made significant strides in anomaly detection, yet the inherent limitation of capturing all conceivable anomalies persists. This paper extends the current paradigm by proposing a novel model that combines Graph Convolutional Networks (GCN) with Normalizing Flow to adapt to unforeseen scenarios. The utilisation of Normalizing Flow enables the learning of intricate parameters associated with previously unseen anomalous activities.

Specifically, this work builds upon the original paper by introducing refinements to the Normalizing Flow model. While the initial model employed an Inverse Autoregressive architecture, our proposed architecture employs a Masked Autoregressive model. This modification enhances the model's capability to capture the complex temporal dependencies inherent in video data, thus improving its effectiveness in anomaly detection tasks.

In the following sections, we delve into the methodology employed, detailing the integration of GCN and the enhanced Normalizing Flow model. Subsequently, we present experimental results that validate the efficacy of our proposed approach in addressing the challenges posed by openset video anomaly detection.

II. LITERATURE REVIEW

These different approaches in anomaly detection highlight varying trade-offs in terms of data requirements, accuracy, and adaptability to diverse scenarios. The scenarios in question however are unable to quantify the real-world scenario of a

plausible blindfold to instances and classes of anomalies that have not been witnessed or recorded .

Fully-supervised methods ([Borisyak, M. et al, Ryzhikov, A. et al]): These techniques demand extensive labelled data for training, which can be resource-intensive and time-consuming to collect. Moreover, when dealing with high-dimensional video data, these methods might encounter challenges in scalability and computational efficiency due to the sheer volume of information. Weakly-supervised methods: ([Charpentier B. et al, Chen, Z. et al., Cho, M. et. Al.]): Operating with less labelling, these methods leverage video-level labels or partial annotations to train anomaly detectors. However, their reliance on weaker supervision might lead to lower accuracy in detecting anomalies or necessitate additional post-processing steps to enhance results, potentially impacting efficiency. Multiple instance learning (MIL) methods: ([Dietterich, T.G et al., Dinh, L. et al., Doshi, K.et al.]) MIL approaches treat video frames as collections of instances (bags) and learn to classify entire videos based on the presence of anomalous instances. While this method handles variations within videos, it may struggle to distinguish between diverse types of anomalies or adapt to new, unseen anomalies during testing phases, posing limitations in its generalizability.

The proposed model for open set video anomaly detection is developed as a multiple instance learning (MIL) framework, which consists of two graph convolutional networks (GCN) followed by an evidential deep learning (EDL) head. The instances are structured as graph data to address the temporal consistency and feature similarity of video clips. The classification confidence and uncertainty are used to select the most informative instances for training. The model is trained by weakly supervised MIL loss and triplet loss, and it can handle semantically shifted testing data that contains unseen anomaly events. The effectiveness of the proposed architecture is demonstrated on multiple real-world video datasets, outperforming existing methods by a large margin [G.Papamakarios et al.]. Masked Autoregressive Flow achieves state-of-the-art performance in a range of general-purpose density estimation tasks. However, the text does not provide a direct comparison to other types of normalizing flows [7]. One key advantage of using Masked Autoregressive Flow (MAF) over Inverse Autoregressive Flow (IAF) is that MAF enables transforming from data x to random numbers u and thus calculating p(x) in one forward pass through the flow, eliminating the need for sequential recursion as in IAF. Additionally, the architecture of MAF directly computes μ and α from previous data variables x, whereas in IAF μ and α are directly computed from previous random numbers u [S. S. Chen et al].

III. Methodology

The methodology employed in this study draws inspiration from a pre-trained action recognition backbone architecture, specifically a C3D-like 3D ConvNet using ImageNet-pre-trained Inception-V1, prominently known as I3D features. The I3D features are of smaller parameter count over C3D and have been trained on ImageNet dataset. The process involves constructing two key similarity graphs - one capturing temporal relationships and the other focusing on feature similarity. Triplet learning (Eq. 1) constraints are introduced to enlarge discrepancies between anomalous and normal classes, enhancing the model's discriminatory power. The triplet loss function is commonly used in tasks like metric learning, where it aims to learn embeddings in such a way that the distance between an anchor instance and a positive instance (from the same class as the anchor) in the embedding space is smaller than the distance between the anchor instance and a negative instance (from a different class). This helps in creating distinct clusters for different classes while minimizing the intra-class variations.

$$Ltriplet = [dap - dan + m]+ \quad (1)$$

A. Authors and Affiliations

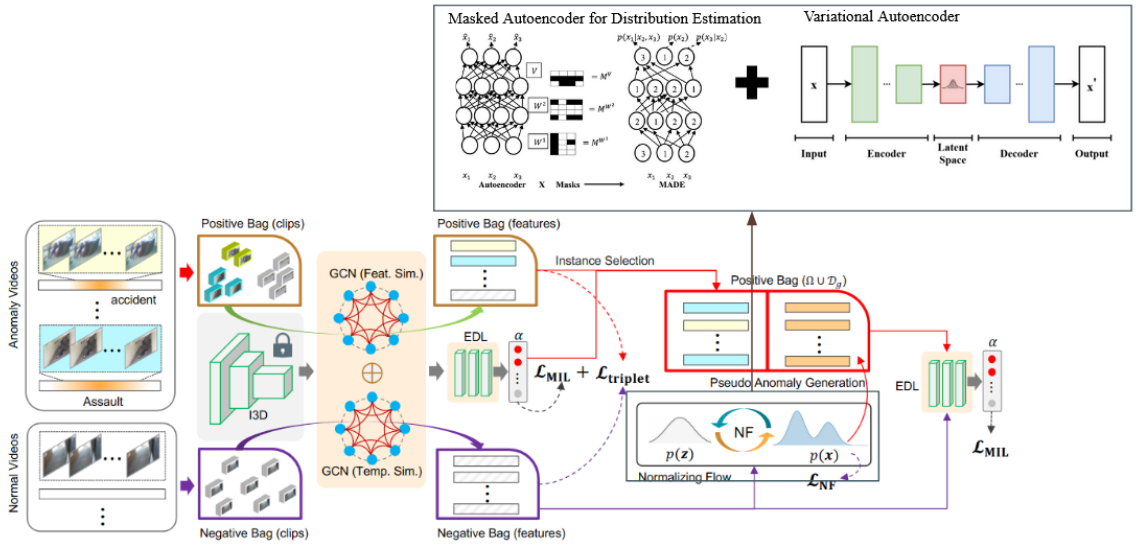
Essentially, the features are used directly, however, is the light of computational limitation, the paper adapts to the features by reducing the feature size from 1024 parameters to a workable size by employing a simple convolutional block.

To further refine the anomaly detection framework over the use of I3D features which account for RGB and flow features, Graph Convolutional Networks (GCN) are incorporated. GCN plays a pivotal role in modelling complex relationships within video frames, providing a nuanced understanding of anomalous patterns. Leveraging the power of GCN contributes to the effective integration of spatial and temporal dependencies in the anomaly detection process. The EDL head is used trained by weakly-supervised MIL loss and triplet loss, and its predicted evidence is used to help select anomaly instances with high cleanness for robust MIL training. This is constituted by a simple neural network of linear layer combined with the ReLU activation function.

The enhancements made to the Normalizing Flow model are noteworthy. The transition from an Inverse Autoregressive model to a Masked Autoregressive model signifies a substantial improvement. While both models utilise MADE (Masked Autoencoder for Distribution Estimation) as their component layer, the architectural distinction lies in the computation of functions {fμi, fαi}. In the Masked Autoregressive Flow (MAF) model, these functions are directly computed from previous data variables, whereas in the Inverse Autoregressive Flow (IAF) model, they are computed from previous random numbers.

This transition bears computational implications. MAF excels in calculating the density p(x) of any data point in one pass through the model but requires D sequential passes for sampling, where D is the dimensionality of x. On the other hand, IAF can generate samples and calculate their density with a single pass, making it more efficient for certain applications. The choice between MAF and IAF depends on the intended usage, and the transition to a Masked Autoregressive model aligns with the objectives of this study.

Based on the framework of the initial paper of reference – Towards Openset Video Anomaly Detection, the framework follows its distinctive architecture. The aforementioned improvement is highlighted in the box in Fig. 1, which uses a masked autoregressive flow model rather than the initial approach of the inverse autoregressive flow model with a Variational Autoencoder. The variational autoencoder is responsible for an optimal distance between the classes.



IV. Results

The model's performance across diverse datasets showcases its robustness in anomaly detection. Across three distinct datasets—XD Violence, UCF Crimes, and Shanghai Tech—the model consistently demonstrated strong discrimination between normal and anomalous instances. In XD Violence and UCF Crimes datasets, with 4 and 9 anomaly types respectively, the model achieved notably high AUC scores of 0.9224 and 0.9204, indicating its ability to effectively discern anomalies. The corresponding loss values of 0.5885 and 0.6489 suggest well-fitted models for these datasets. Even in the more complex Shanghai Tech dataset with 25 anomaly videos, the model sustained commendable performance with an AUC score of 0.8986, implying reliable anomaly detection. Despite a slightly higher loss value of 0.69, the model's ability to distinguish anomalies in a dataset with diverse anomaly types substantiates its versatility and efficacy across varied scenarios.

TABLE I. Comparison Table Results of Towards OpenVAD

Dataset	OpenVAD (IAF-VAE)		OpenVAD (MAF-VAE)	
	AUC Score	Loss	AUC Score	Loss
XD Violence (4 Seen Anomaly Types)	0.69	NA	0.9224	0.5885
UCF Crimes (9 Seen Anomaly Types)	0.85	NA	0.9204	0.6489
Shanghai Tech (25 Seen Anomaly Videos)	0.93	NA	0.8986	0.69

In the literature survey results provided in Table 2, the results achieved by OpenVAD (MAF-VAE) are compared to other works in the field. For XD Violence, OpenVAD (MAF-VAE) achieved an AUC score of 92.24, which was the highest among all the works compared. For UCF Crimes Dataset, OpenVAD (MAF-VAE) achieved an AUC score of 92.04, which was also the highest among all the works compared. Overall, the results achieved by OpenVAD (MAF-VAE) were better than the other works compared in the literature survey.

TABLE II. Comparison Table with Other Works (UCF Crimes)

Paper Titles	XD Violence	UCF Crimes
OCSVM (unsupervised) [12]	27.25	NA
Conv-AE (unsupervised) [10]	30.77	50.60
WU et. al. (online) [16]	67.77	79.11
Wu et. al. (offline) [16]	64.29	79.96
RFTM [14]	63.65	79.55
GODS (unsupervised) [15]	NA	70.46
OpenVAD (IAF-VAE)[17]	69.61	80.14
OpenVAD (MAF-VAE)	***92.24***	***92.04***

V. Conclusion

In conclusion, "Unlocking Anomaly Dynamics: A Masked Autoregressive Model and GCN Fusion for Openset Video Anomaly Detection" presents a novel approach to anomaly detection in video surveillance. The proposed model integrates Graph Convolutional Networks and a refined Normalizing Flow to capture complex temporal dependencies and learn intricate parameters associated with previously unseen anomalous activities. The paper compares the proposed architecture with existing methods and shows that it outperforms them by a large margin. The model was tested on multiple real-world video datasets, and the results demonstrate the effectiveness of the proposed architecture in handling semantically shifted testing data that contains unseen anomaly events. The paper notes a significant performance boost in the problem of Openset Video Anomaly Detection, evident in the above tabulation and quantified at a 0.92 AUC score over two datasets.

Overall, the proposed architecture presents a promising approach to anomaly detection in video surveillance. The integration of Graph Convolutional Networks and Normalizing Flow allows for the learning of intricate parameters associated with previously unseen anomalous activities. The use of Masked Autoregressive Flow adds a computational advantage and makes the model more efficient and effective in density estimation tasks. The proposed architecture has the potential to be applied to other fields beyond security and surveillance, and future research can explore its applications in other domains.

REFERENCES

[1] Borisyak, M., Ryzhikov, A., Ustyuzhanin, A., Derkach, D., Ratnikov, F., Mineeva,O.: (1 + epsilon)- class classification: An anomaly detection method for highly imbalanced or incomplete data sets. arXiv preprint arXiv:1906.06096 (2019)

[2] Charpentier B., Zügner, D., Günnemann, S.: Posterior network: Uncertainty estimation without ood samples via density-based pseudo-counts. (2020)

[3] Chen, Z., Tian, Y., Zeng, W., Huang, T.: Detecting abnormal behaviors in surveil-lance videos based on fuzzy clustering and multiple auto-encoders. In: 2015 IEEE International Conference on Multimedia and Expo (ICME). pp. 1–6. IEEE (2015)

[4] Cho, M., Kim, T., Kim, W.J., Cho, S., Lee, S.: Unsupervised video anomaly detection via normalizing flows with implicit latent features. Pattern Recognition 129,108703 (2022

[5] Dietterich, T.G., Lathrop, R.H., Lozano-Pérez, T.: Solving the multiple instance problem with axis-parallel rectangles. Artificial intelligence 89(1-2), 31–71 (1997)

[6] Dinh, L., Sohl-Dickstein, J., Bengio, S.: Density estimation using real nvp. In: ICLR (2017)

[7] Individual household electric power consumption data set. Accessed on 15 May 2017.

[8] Doshi, K., Yilmaz, Y.: Rethinking video anomaly detection-a continual learning approach. In: Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision. pp. 3961–3970 (2022)

[9] G. Papamakarios, T. Pavlakou, and I. Murray, "Masked Autoregressive Flow forDensity Estimation," Jun. 2014.

[10] Hasan, M., Choi, J., Neumann, J., Roy-Chowdhury, A.K., Davis, L.S.: Learning temporal regularity in video sequences. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 733–742 (2016)

[11] Ryzhikov, A., Borisyak, M., Ustyuzhanin, A., Derkach, D.: Normalizing flows fordeep anomaly detection. (2019)

[12] Smeureanu, S., Ionescu, R.T., Popescu, M., Alexe, B.: Deep appearance features for abnormal behavior detection in video. In: International Conference on Image Analysis and Processing. pp. 779–789. Springer (2017)

[13] S. S. Chen and R. A. Gopinath. Gaussianization. Advances in Neural Information Processing Systems 13, pages 423–429, 2001.

[14] Tian, Y., Pang, G., Chen, Y., Singh, R., Verjans, J.W., Carneiro, G.: Weakly supervised video anomaly detection with robust temporal feature magnitude learning. In: Proceedings of the IEEE/CVF international conference on computer vision (2021)

[15] Wang, J., Cherian, A.: Gods: Generalized one-class discriminative subspaces for anomaly detection. In: Proc. of ICCV. pp. 8201–8211 (2019)

[16] Wu, P., Liu, J., Shi, Y., Sun, Y., Shao, F., Wu,Z., Yang, Z.: Not only look, but also listen: Learning multimodal violence detection under weak supervision. In: European Conference on Computer Vision. pp. 322–339. Springer (2020)

[17] Y. Zhu, W. Bao, and Q. Yu, "Towards Open Set Video Anomaly Detection," Aug. 2022.https://www.ecva.net/papers/eccv_2022/papers_ECCV/papers/136940387.pdf[2] G. Papamakarios, T. Pavlakou, and I. Murray, "Masked Autoregressive Flow for Density Estimation," Jun. 2014.

Intrusion Detection and Prevention System for IoT Networks

Ajay Kumar
SCSE, Galgotias University, Noida, India
ajay@galgotiasuniversity.edu.in
Kumar Abhishek
National Institute of Technology, Patna, India
kumar.abhishek@nitp.ac.in
Abhay Kumar
National Institute of Technology, Patna, India
kumar.abhishek@nitp.ac.in

Rutvij H. Jhaveri*
Pandit Deendayal Energy University, Gandhinagar,India
Rutvij.Jhaveri@sot.pdpu.ac.in
Loknath Sai Ambati
Meinders School of Business Oklahoma City University, USA
lambati@okcu.edu

Abstract— There was a need for instant, seamless and secure access to information for Internet of Things devices based networks. Additionally, security issues had reinforced the need for collaboration among the IoT devices and clound networks. Firewalls were utilized as a security check point to IoT devices but as cloud based network was capable of working on both terrestrial as well as satellite modes of communication, security issues continued to develop. Keeping in mind the end goal to shield the IoT network from illegal access, the idea of implementation of IPS and IDS in IoT network is proposed in the current paper. An IDS observes and examines the IoT in real-time for recognizing the indication of intrusions occurrences. These occurrences may either be infringement or the dangers that are going to happen disregarding the cloud network security or standard security arrangements. The current project also aims to introduce IPS that is utilized to [1] distinguish indications of interruptions into systems or frameworks and make the right move. Together the IDS and IPS creates cautions potentially blocking intrusions to IoT Networks. The attempts could be breaches or potential threats that undermine the DCN's security or standard security measures. The project also aims to introduce IPS, which detects intrusion signs in systems or networks and takes appropriate actions [1]. By combining IDS and IPS, alerts are generated to potentially block intrusions into the IoT network or Cloud services.

Keywords - Tactical Network, Network-based Intrusion Detection System (NIDS), IPS,Server, Firewall, Defence Communication Network, Internet of Things.

I. INTRODUCTION

Each country and organisation is prioritizing the develop-ment of a robust and seamless communication network for their IoT based devices. The tactical satellite communication network and Cloud services for the IoT based devices have necessitates a secure and uninterrupted wireless network with multiple hops to enable safe communication among the different branches of the nation's military as well as any organisation.

In the present "network-centric battlefield," self-forming networks play a crucial role in facilitating the operations of mobile forces that are rapidly deployed to meet tactical requirements. However, tactical network environments pose significant challenges due to unreliable connectivity, remote unit/regiment locations, limited bandwidth, and latency issues. Extensive research has focused on various aspects including data link layers, routing and transport protocols, and cross-layer optimization [2]. Many armies possess a large number of narrowband tactical radios operating in very high frequency and ultra-high frequency ranges. The performance of these legacy radios can be enhanced to support digitization on the battlefield. This article proposes several guiding principles and important considerations for the design and implementation of legacy tactical radio networks for the tactical forces.

The defense wireless network faces a significant challenge with cyber attacks aiming to steal confidential data from the armed forces. Attackers actively search for vulnerable points in the network to breach and gain access to classified information related to tactical armed forces.

A hacker is an individual who exploits weaknesses in computer systems or network connections. Hackers are motivated by different factors, like gain of finance, notoriety, personal challenge, or enjoyment [3], [4]. Through the breach of security measures on a victim's network, they are able to gain control or infiltrate the system, utilizing software vulnerabilities, engaging in identity theft, or launching large-scale denial-of-service (DoS) attacks. Security breaches and data compromises result in substantial financial losses, with companies reportedly losing billions of dollars annually to hackers. In fact, it is projected that these losses may reach one trillion dollars by 2020, highlighting the importance of studying effective mechanisms to prevent network attacks and safeguard our work environments [5], [6].

A computer network refers to a collection of interconnected devices, including computers and other hardware components, facilitating resource sharing and data communication through communication channels. The devices connected within the network are referred to as nodes [7]. In our paper, we focus on monitoring Local Area Networks (LANs) to detect any instances of intrusion [8]. LANs are typically deployed within organizations, educational institutions, etc.,

Corresponding Author : Rutvij.Jhaveri@sot.pdpu.ac.in

to facilitate resource sharing and efficient communication among the nodes present in the network. Despite employing a wide range of security measures [9], [10], LANs are still vulnerable to highly sophisticated hacking attacks.

The network firewall serves as the initial layer of security measures implemented. However, as the techniques for detecting disruptions evolve, so do the number of attacks. Although window defenders, firewalls, and antivirus software provide a certain level of security for the network nodes, sometimes these security measures themselves become po-tential entry points for attackers attempting to compromise the system [8], [11].

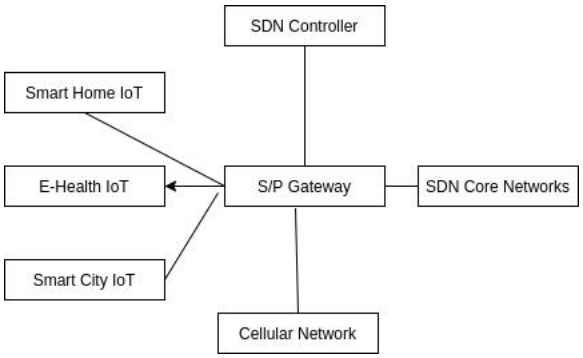


Fig. 1. Internal architecture of SDN system

”Intrusion prevention systems (IPS), also known as intrusion detection and prevention systems (IDPS), are network security systems that monitor both network infrastructure and system activities for malicious behavior” [12]. The primary objective of the Tactical Network-Based Intrusion Detection System (NBIDS) and Network-Based IPS is to continuously monitor the traffic of the tactical network server, detect any suspicious activities, and report them to the NBIPS for filtering the tactical communication traffic.

The paper primarily focuses on addressing network se-curity issues, specifically data privacy and security during network communication. Through a comprehensive analysis and comparison of research contributions in each area of network security, the study aims to identify current research challenges and highlight promising future research direc-tions.

The essay’s structure is as follows: Section II lists various network attack methods, Section III discusses IDS concepts, Sections IV to VI cover IDS system categorization and the proposed system, and Section VII provides an overview of the conclusion and suggested future work.

II. LITERATURE REVIEW

An Intrusion Prevention System has different preferences to organize based IPS are: System based IPS/IDS don’t devour framework assets. Uninvolved executions of IPS/IDS don’t interfere with organize activity stream. A few frameworks based security devices should even now be utilized alongside IPS which moreover may identify the threats [13]. Device Incorporate Hostile to Anti-infection and against spyware. You require this at any rate to stop the shoddy stuff, and these may likewise stoppage Well-suited assaults. Likewise Firewalls. Parcel sifting at the framework level may even now be a smart thought, especially if outbound associations are likewise constrained to those administrations that are genuinely required.

A. Regular Identification Methodologies

IPS uses special techniques to detect security incidents. The creators of IPS have found that no one technique is adequate to identify and stop most types of events; rather, they have committed to various understandable approaches to achieve this [14].

Stateful convention analysis

IPS can identify occurrences by watching singular system associations, for example, also, settling on cautioning or blocking choices in light of what’s viewed as typical for different kinds of exercises. For instance, an IPS may take in the grouping of occasions when the client of a web application sign in, and subsequent to signing in issues orders to the application to perform work[15]. The IPS may consider a client issuing charges without signing in to be an occasion that ought to be blocked, on the grounds that this might be an indication of a gatecrasher who is endeavoring to perform unapproved exchanges [16].

False Positives

With regards to interruption counteractive action, a false positive is an IPS pronouncing great activity as terrible, bringing about either a false caution (if the IPS is in latent recognition mode) or administration disturbance (if the IPS is in online avoidance mode). A false positive is normally caused by an insufficient IPS control or mark. A false positive ought not be mistaken for a ”genuine” assault that is incapable against the working framework or application it is focusing on. For instance, if Conficker assaults a Linux have, and an interruption occasion is activated, it is in fact not a ”false positive” but rather to a greater extent a ”not appropriate” since Conficker just influences Windows working frameworks. Original IDSs were unbelievable for making enormous amounts of cautions, overpowering executives who invested hours endeavoring to block out the commotion. Gaining from those agonizing circumstances, IPS sellers have improved their frameworks much through canny learn modes, less demanding organization, and exceptionally tuned administer sets [17].

False Negatives

The contrary issue is that of a false negative, where an IPS neglects to perceive an interruption or other security occasion. This can happen if the IPS doesn’t have exceptional principles, or if the IPS merchant hasn’t discharged a run for another kind of assault or powerlessness. At the point when an IPS is set in online blocking mode, false negatives are for the most part significantly more harming to an association than a false positive. A false negative allows

terrible activity to enter the system, conceivably prompting traded off frameworks and perhaps stolen or lost information. A false positive squares great activity from entering the system, conceivably prompting lost business or efficiency.

B. IPS Scope DOMAIN

IPS are intended to square a wide range of sorts of assaults. It is less demanding to comprehend IPS on the off chance that we better comprehend the kinds of things they're intended to identify and forestall. [?] The accompanying are different sorts of assault modules utilized by programmers to access a framework or influence the best possible working of any system.

Buffer overflows

A cradle flood is a particular kind of assault against a framework, where the assault is composed to confound the framework into executing the assailant's instructions.A support flood assault works this way. An assaulting program sets up an interchanges session with a particular segment on the objective framework, and sends an extraordinarily made message to the target framework. The message intentionally sends excessively information into the objective framework's input support. In a program that is defenseless against a support flood assault, the abundance information will overwrite program guidelines in the defenseless program, and in the long run the program will execute those directions (feeling that it is executing its unique guidelines). [1] Worms, Trojans, infections, and different kinds of malware regularly utilize buffer floods as a method for picking up a decent footing in another casualty framework. Buffer floods represent a huge part of the assaults against frameworks on the Web.

Zero Day Attacks

"A zero day assault is a fresh out of the box new assault on a formerly obscure weakness, or a new sort of an assault on a current weakness The term multi day originates from the quantity of long periods of caution between the time when the weakness is reported and when it is misused. At the end of the day, these are vulnerabilities for which no patches are accessible"[15]. Zero-day assaults are huge in light of the fact that mark based (misuse based) IPS gadgets are for the most part vulnerable against them. Be that as it may, IPS that likewise utilize abnormality based location and use helplessness based standards (instead of endeavor based marks) can secure adequately against multi day assaults.

III. CONCEPTS

Now days wireless based and satellite Communication based in the tactical network plays a pivotal role for the any country armed forces for operation which fulfill the tactical need. However, tactical network environments also present significant challenges due to unreliable connectivity [2], different type of cyber attack on the military server, Remote location of the units/regiments should be analysed and taken on first response towards the any malicious activity to theft the confidential information of the tactical forces movement to conquer the war towards the adversary country. To avoid this type of infiltration or attack in the tactical network we proposed the first regular analysis of flow of data in and out through the server/network and then first line prevention action should be taken automatically there itself.

IV. PROPOSED SOLUTION

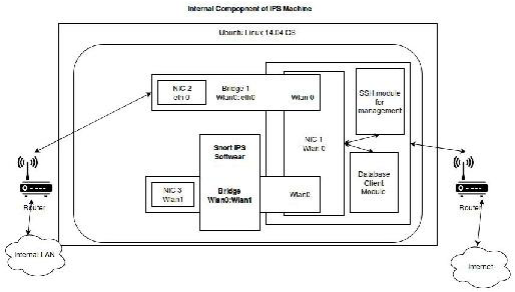


Fig. 2. Architecture of Network Based IPS system

An Intrusion Prevention System is any risk in the system or channel avoidance innovative that examines IoT environment based network protocol which identify and restrict abuse through helplessness. Impotence abuse largely comes as a malicious contribution to a targeted the IoT based network or Environment, which attackers use to disrupt and take control of an autonomous system or IoT enlivenment. After a successful attack, the attacker can disable the targeted machine (causing a denial of service condition) or he can gain access to all Scan permissions and authorization that the traded machine has access to the user.

Prevention

"The Tactical NBIPS placed behind the firewall and allow to access the adversary only incoming after the clearance given by NBIPS system. In other hand the NBIDS will carry on the analysis of the all traffic passes through the firewall as well as NBIDS and report to admin and forward the log to the NBIPS for further filtering of block or allowing of the traffic

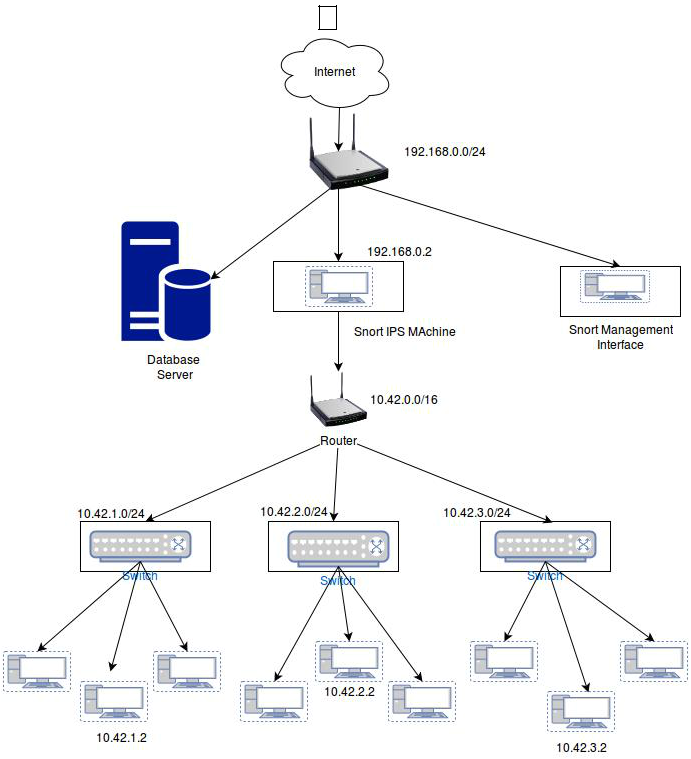


Fig. 3. NBIPS system External architecture

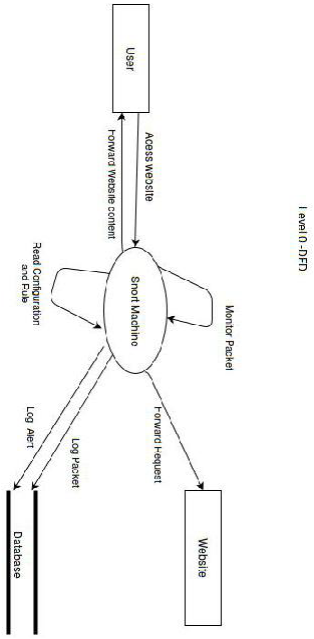


Fig. 4. Data Flow Diagram 1

VLAN port Scan: This is a common technique for hackers which is user for finding the open door or weak point in the network or IoT Environment. Using the NIDS system we can prevent the hackers to enter in our network/channel which one also backings an aloof IDS setup.

Data Flow Graph

DFD represent flow of information through a data framework (Graphical) that demonstrates its procedural perspectives. A DFD is often used as an introductory project to provide a framework review that can be discussed later. DFDs can also be used to represent information processing. A DFD specifies what type of data contributes to and emerges from the framework, where the information came from and where it is going, and where the information is placed. It does not show data on process planning or data on whether processes are grouped or working in parallel[2]. The accompanying figure indicate Data Flow graph of actualized framework. Using DFD we also verified the NIDS and IDS in IoT envirnment.

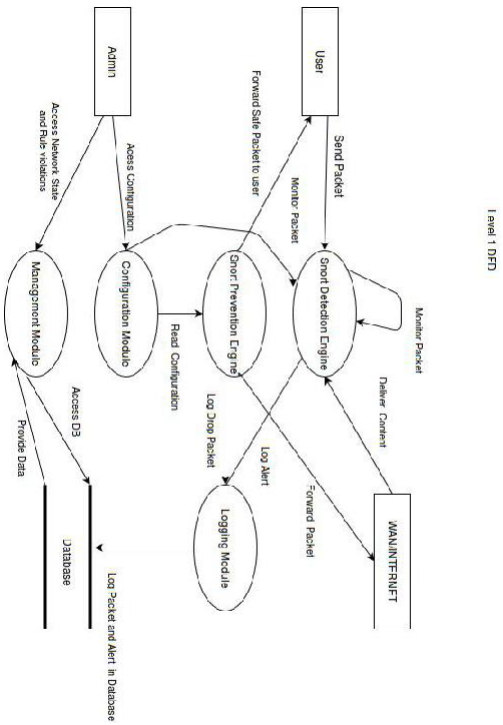


Fig. 5. Data Flow Diagram 2

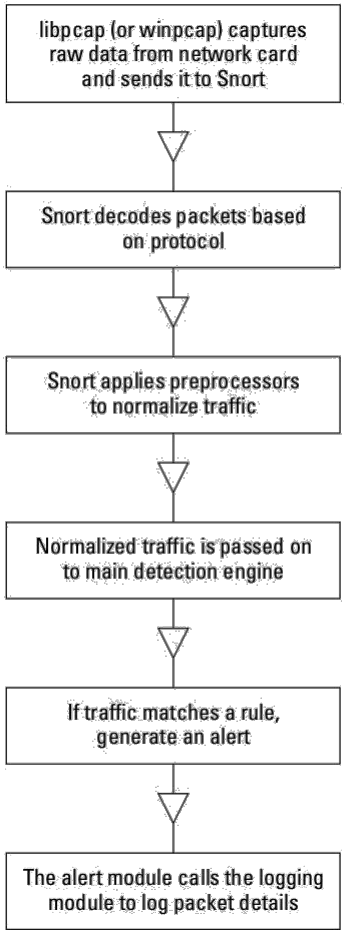


Fig. 6. Snort Internal Packet processing diagram

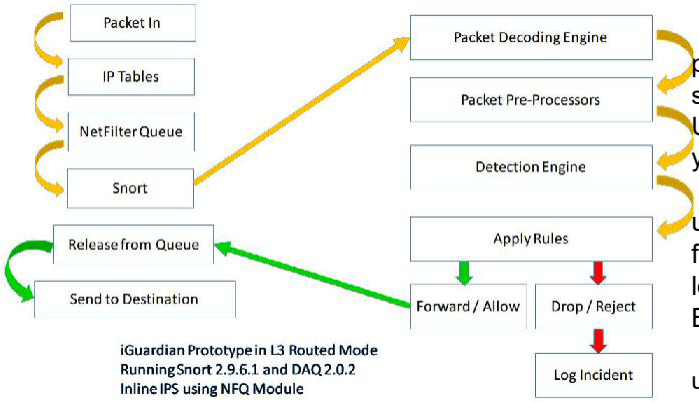


Fig. 7. Packet processing architecture

V. SYSTEM REQUIREMENT

Requirement of Software for implement in the Tactical Network

For setup the the testing of the proposed system we need the Software Barnyard2 version-2.1-13. and BASE2 version 1.4.5. Snort version 2.9.8.2 for Pulledpork. *Operating Sys-tem Linux Ubuntu 14.04 Server [2]

Hardware Requirements

Component/Item Name	Description	Quantity Needed
IPS Machine	HP ProLiant XL450 Gen9 Server	1
Ethernet Adapters	Intel® Ethernet Server Adapter I210	2
Router	Cisco WRVS4400N Wireless-N Gigabit Security Router	2
Switches	Cisco 24 Ports 10/100 Switch (SF90-24)	As required
Wireless Network Adapters	Netgear 802.11ac Dual Band Gigabit A6200	2
Database Server	Oracle SPARC Database Server	1

Fig. 8. Hardware Requirements of implemented IPS

Configuring Barnyard

”The key advantage of the software is that it will enable Snort to write to disk in an effective manner and will remove the task of parsing parallel data into different organizations from Snort; therefore, Snort will not miss organize events” [15].

Barnyard2 has 3 methods of activity:

- persistent- w/bookmark,
- group (or one-shot),
- persistent.

A barnyard plugin is a basic tool for parsing Snort’s parallel binary documents, dealing with what’s more, sending the content to a diverse set of yield plugins. Unfortunately, it has not been updated in more than four years and will not be maintained by the original designers.

Unfortunately, The original designers have not updated the Network based Intrusion detection system for the IoT environment for over five years and will no longer maintain it. A yield framework is provided by the Barnyard framework as an extension to Snort.

It is a binary document format that is referred to as brought unified. After reviewing this document, Barnyard resends the information to the recipient.The backend of the database is connected to the front end. Unlike the database throughput module, Barnyard knows that alerts cannot be sent to the database and stops sending alerts. You should also be notified when the repository is able to recognize the organization and start sending notifications again.

BASE

”BASE is the very first line and foremost investigation and security tool and it is depends on the ACID (Analysis Console for Intrusion Databases) project code” [18]. The

application provides a web front end for querying and analyzing reports within the Snort IDS framework. Snort operates in standalone mode as a packet sniffer and logger. With a few additional implementations and some configurations, the Snort framework has proven to be more useful than NBIDS. The software that supports the required sections in this article are: Barnyard2 is a specialized organization for Snort's dual production combo2 package.

Packaging is very important for assets, so to reduce the Snort process: We allow Snort to back up suspicious packets to the index on a connection type without processing the packet goods. Barnyard2 then creates these packages asynchronously and saves them to a MySQL store.

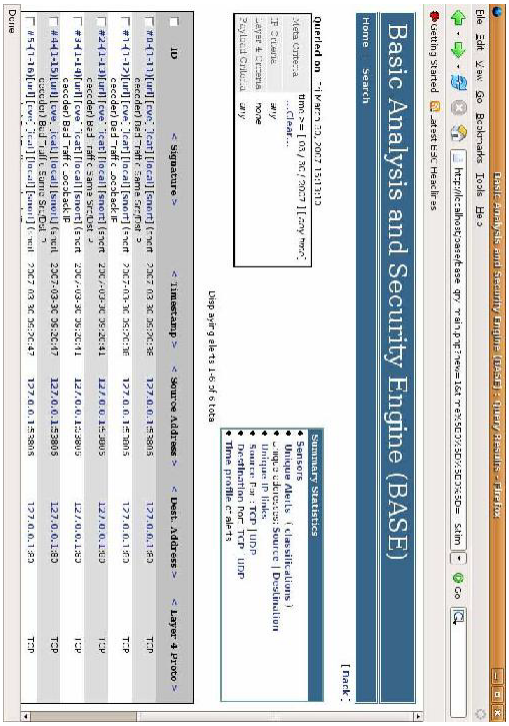


Fig. 9. Basic Analysis and security Engine 1

"PulledPork is a Perl component that takes the latest Snort code because dangerous situations are constantly evolving and Snort needs new code to distinguish suspicious activity (rule sets such as anti-virus flags)[2]. Fundamental Analysis and the Security Engine (BASE) provides a web front end for querying and analyzing reports within the Snort framework.

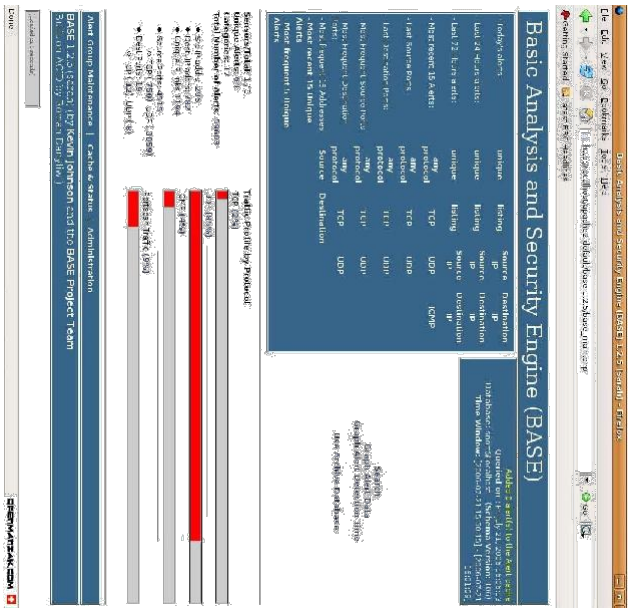


Fig. 10. Basic Analysis and security Engine 2

Initially after the test and findings we have to placed each one of the essentials from the Ubuntu stores as follows: " sudo apt-get install -y build-essential libpcap-dev libpcre3-dev libdumbnet-dev bison flex zlib1g-dev " [2]

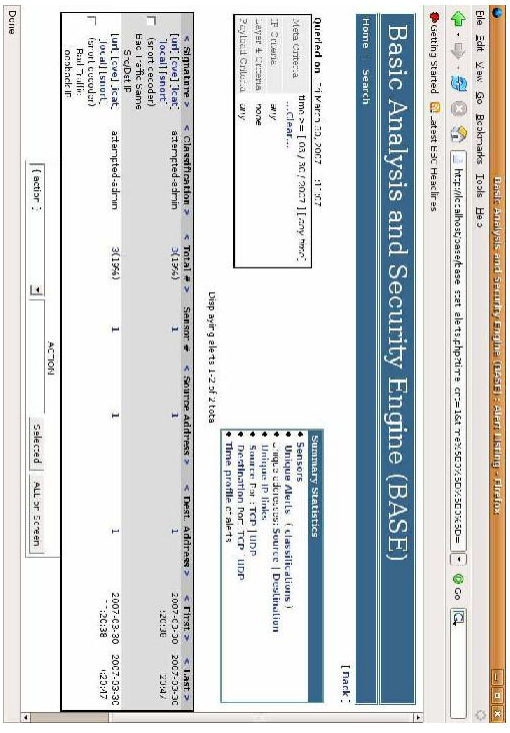


Fig. 11. security Engine 3 for Basic Analysis

VI. IMPLEMENTATION

Running Snort : In line mode

1. Rules which configured to trigger

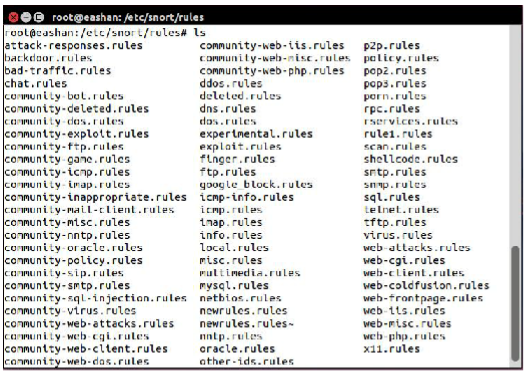


Fig. 12. Snort Rule

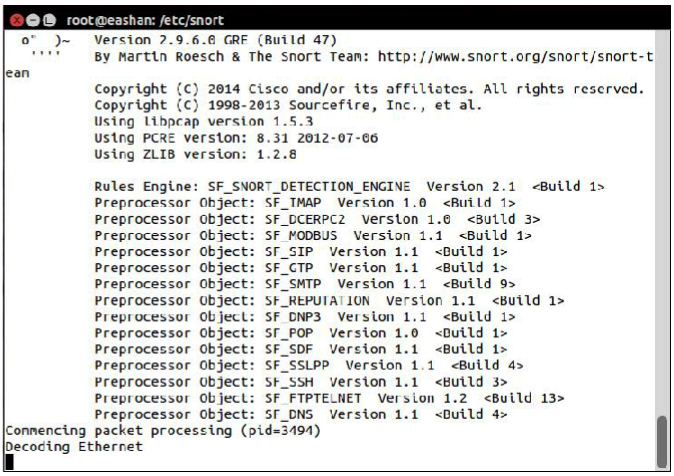


Fig. 15. Summary of Analysis

2. As soon as the first phase of the snort command has been completed, we need to set up a cloud or BASE network bridge between the wlan0 and wlan1 network interfaces in order to begin inline processing of PF packets. After using of wlan0 and wlan1 interfaces we can filter the pf packets for IoT network traffic.

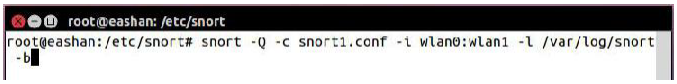


Fig. 13. Snort Configuration

5.Snort writes its reports and packets to the /var/log/snort directory as described in step 5. See the Snort event area. In step 6 of Snort, immediately notify the administrator and terminate the communication key.

3. On execution of command in Step 2 following output is obtained.

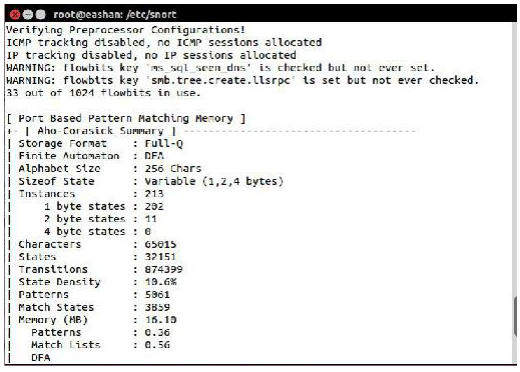


Fig. 14. Snort Configuration

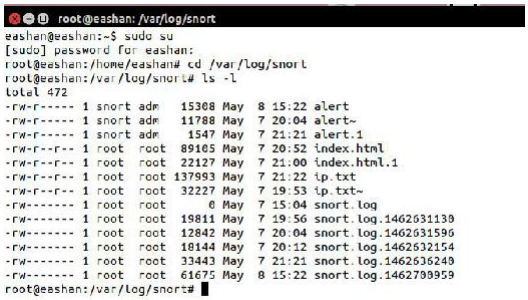


Fig. 16. Logs Record

4. The inline parcel handling has finished. The rundown of the investigation of the meeting is given by Snort.

6. The alerts fired by Snort: The capture alert fired by the NBIPS and reconfigured the firewall after the resetting the the connection among the cloud based server and user or clients.

Fig. 17. Alert Generate

7. Passive alert rules: In this step the fired session again create the rules used during the trigger of the passive alert to the admin or server and sent it for the further logs analysis to the step 5.

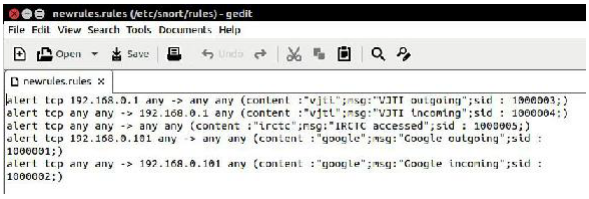


Fig. 18. Rules used to trigger passive alert

Pullpork for rule list

”Features and Capabilities of using Pullpork”.

- Confirmation of Checksum after downloading of real manage traffic data.
- Refreshed the sid-msg.map record for all programmed data-set
  to include your local rules in the record of sid-msg.map.
- The ability to retrieve rules tarballs from custom Urls
- Object bolster should be shared
- IP Notoriety Rundown for the whole IP addresses
- The ability to instantly download a large number of diverging rule sets
- change log should be readily available
- Option of HUP forms after standards are downloaded and processed
- tuning of rule sets should be readily available
- Verbal returns with the goal that you know exactly what’s going on
- ”Condition for Negligible Perl Module should matched the IP addresses .
- Sweet Smoky effect through pork”[2]

Results

By deploying the infrastructure, proper inclusion will be achieved. Deploying the network incorrectly will result in a slowdown and inefficient management of the network. Establish a risk assessment and prioritize risk areas. On the base of the pullpork trigger passive alert, the NBIDS stored, analyzed, and finally sent the logs record to the NBIPS for further parsing, changing the state of the grunt rule sets. It stops the variety of net-work attacks on the Defence communication network. Immediately alert the admin and drop the captured malicious network packets. Terminate the session key and reset the connection between the defense server and the user by reconfiguring the network firewall. Repacked the payloads and removed the infected at-tachments from defense servers. The proposed system also creates the authenticated authorized-based session key for tactical communication. In the event of intrusion detection, we plan an appropriate incident response. One metric of a network’s efficacy is its throughput. Throughput in the proposed 280 the approach represents the efficiency with which system resources are used. As part of the 281 the proposed method, we establish communication between the host and controller, both 282 before and after an attack. Based on the findings, it is clear that the proposed system 283 improves network performance in a meaningful way which is seen in the figure.Throughput 284 the proposed network can be calculated with the help of the given equation. Data packets, 285 including TCP window size and Round Trip Time, are used to calculate throughput. Throughput = TCP Window size(Data packets) Round TripTime (1)

VII. CONCLUSIONS

The network-based intrusion prevention systems (IPS) represent a future where turmoil, uncertainty, and expenses are replaced by safety, profitability, and efficiency. These systems aim to provide comprehensive security management, ensuring the protection of all security activities. It is es-sential for all defense officers, both in public and private organizations, to adopt network-based IPS for the following objectives: enhancing business profits and efficiency, safe-guarding sensitive data from theft, defending critical systems against imminent global cyber attacks, and minimizing copy-right infringement. NIDS (Network based intrusion detection system) play the vital role for the detection and prevention for the any packet traffic and secure the channel system. There is also reduced the chances of port scanning attach by the intruder.

Moving forward, there is a growing demand for intrusion prevention systems worldwide, driven by the increasing

sophistication of attacks on defense communication networks and the emergence of new and innovative attack strategies targeting cloud servers and communication network protocols. One of the key challenges faced by traditional intrusion prevention software systems is the need for timely upgrades to address all types of software vulnerabilities. Delaying the implementation of necessary updates could result in significant damage to defense communication networks, as vulnerabilities may already be exploited by the time they become publicly known.

Future Works

The absence of intrusion detection and response solutions specifically designed for CoAP and defense communication networks has been acknowledged. In the future, we intend to enhance the framework by expanding its capabilities to detect and filter new forms of intrusion attacks.

The study revealed the effectiveness of discriminant functions in assisting Epidemic in selecting secure connections. This resulted in a reduction of information shared with unauthorized nodes and an improvement in message delivery compared to employing more restrictive rules.

References

[1] Hilmi Gunes¨ Kayacik and A Nur Zincir-Heywood. A case study of three open source security management tools. In Integrated Network Management, 2003. IFIP/IEEE Eighth International Symposium on, pages 101–104. IEEE, 2003.

[2] Ajay Kumar, K. Abhishek, M.R. Ghalib, A. Shankar, and X. Cheng. Intrusion detection and prevention system for an iot environment. Digital Communications and Networks, 8(4):540–551, 2022.

[3] Martin Roesch et al. Snort: Lightweight intrusion detection for networks. In Lisa, volume 99, pages 229–238, 1999.

[4] Bruno Astuto A Nunes, Marc Mendonca, Xuan-Nam Nguyen, Katia Obraczka, and Thierry Turletti. A survey of software-defined networking: Past, present, and future of programmable networks. IEEE Communications Surveys & Tutorials, 16(3):1617–1634, 2014.

[5] Vern Paxson, Jim Rothfuss, and Brian Tierney. Bro quick start guide. Retrieved April, 22:2010, 2004.

[6] Pranav Nerurkar and Aruna Pavate. Study of angular js: A client side javascript framework for single page applications. International journal of contemporary research in computer science and technology, 1(4):92–96, 2015.

[7] Christian Kreibich. Network intrusion detection: Evasion, traffic normalization, and end-to-end protocol semantics.

[8] Mohammad Dabbour, Izzat Alsmadi, and Emad Alsukhni. Efficient assessment and evaluation for websites vulnerabilities using snort. International Journal of Security and Its Applications, 7(1):7–16, 2013.

[9] Seho Han and Howon Lee. Performance analysis of coverage probability according to transmission range of devices. Journal of the Korea Institute of Information and Communication Engineering, 20(10):1881–1886, 2016.

[10] Pranav Nerurkar and Sunil Bhirud. Modeling influence on a social network using interaction characteristics. International Journal of Computer & Mathematical Sciences, 6(8):152–160, 2017.

[11] Diego Kreutz, Fernando MV Ramos, Paulo Esteves Verissimo, Chris-tian Esteve Rothenberg, Siamak Azodolmolky, and Steve Uhlig. Software-defined networking: A comprehensive survey. Proceedings of the IEEE, 103(1):14–76, 2014.

[12] Kamal Benzekki, Abdeslam El Fergougui, and Abdelbaki Elbelrhiti Elalaoui. Software-defined networking (sdn): a survey. Security and communication networks, 9(18):5803–5833, 2016.

[13] Rutvij H. Jhaveri, Rui Tan, Arvind Easwaran, and Sagar V. Ramani. Managing industrial communication delays with software-defined net-working. In 2019 IEEE 25th International Conference on Embedded and Real-Time Computing Systems and Applications (RTCSA), pages 1–11, 2019.

[14] Mahdi Jemmali, Mohsen Denden, Wadii Boulila, Gautam Srivastava, Rutvij H. Jhaveri, and Thippa Reddy Gadekallu. A novel model based on window-pass preferences for data emergency aware scheduling in computer networks. IEEE Transactions on Industrial Informatics, 18(11):7880–7888, 2022.

[15] Ajay Kumar, Aruna Pavate, Kumar Abhishek, Abhijeet Thakare, and Mansi Shah. Landmines detection using migration and selection algorithm on ground penetrating radar images. In 2020 International Conference on Convergence to Digital World - Quo Vadis (ICCDW), pages 1–6, 2020.

[16] Pranav Nerurkar, Aruna Pavate, Mansi Shah, and Samuel Jacob. Performance of internal cluster validations measures for evolutionary clustering. In Computing, Communication and Signal Processing, pages 305–312. Springer, Singapore, 2019.

[17] Pranav Ajeet Nerurkar, Madhav Chandane, and Sunil Bhirud. Ex-ploring convolutional auto-encoders for representation learning on networks. Computer Science, 20(3), 2019.

[18] Olivier Flauzac, Carlos Javier Gonzalez Santamar´ıa, and Florent Nolot. New security architecture for iot network. 2015.

Thai Document Recognition using Yolov8

Nattapong Phu-utha
Computer Science Department
Kasetsart University
Bangkok, Thailand
nattapong.phuu@ku.th

Paphatsara Thiniwong
Computer Science Department
Kasetsart University
Bangkok, Thailand
paphatsara.t@ku.th

Wannasa Pomjun
Computer Science
Kasetsart University Sriracha Campus
Chonburi, Thailand
wannasa.po@ku.th

Pyeoungkee Kim, Professor
Divivsion of Computer Engineering
Silla University
Busan, Republic of Korea
pkkim@silla.ac.kr

Somchoke Ruengittinun
Division of Computer Science Department
Kasetsart University
Bangkok, Thailand
Somchoke.r@gmail.com

***Abstract*— Object recognition models have been widely adopted for various applications including license plate recognition and many practical object recognition tasks. However, in the realm of Thai document OCR, many researchers are considering using Transformer-based models or open libraries like Tesseract. Though recent models such as TrOCR can show good performance in the recognition of handwritten characters, they require huge training datasets and training time, as well as a fast and powerful hardware environment. In this paper, we present an experimental investigation using the YOLOv8 object detection model for Thai document OCR, which requires comparably less data and an affordable training environment. Using a training dataset consisting of a total of 31,039 images on a single GPU, the results reveal a Character Error Rate (CER) of 3%. This result shows that the use of an object detection model in building document OCR systems for complex languages is a very practical choice, serving as a valuable contribution for future OCR implementations in other languages.**

Keywords— OCR (Optical Character Recognition), CER(Character Error Rate), Training Dataset

1. INTRODUCTION

In recent years, with the development of powerful deep learning models, OCR (Optical Character Recognition) technology has shown a growing trend in transforming business processes. OCR involves using computers to convert handwritten or printed text into digital data that can be understood by a computer system. This technology has not only become prevalent in the business sector but has also become an integral part of the daily lives of individuals. For instance, applications like Google Street View and Dropbox employ OCR to translate images into real-time text.

Much recent image-related research has heavily focused on the use of transformer models. However, transformer models usually require huge training datasets and longer training times with high-performance GPUs, which are hard for individuals or small companies to equip. Object detection models can analyze and detect characters for a whole page, while transformer-based models normally handle a line at a time. Also, the usual number of parameters for transformer models is much larger than that for object detection models. Compared with other language characters and sentences, there is no space between words in Thai sentences, and characters consist of curves, circles, diacritics, and heads. Furthermore, vowels can be placed at the top, right, bottom, and left of each consonant, and intonation (wannayuk) can even be placed above the top vowel location. Hence, there can be many complex character combinations, which can require too many complex character and word training datasets for whole-character-based or word-based recognition models. This can be seen in other combinational character languages, such as Korean OCR recognition. Hence, while OCR models utilizing transformer models are well-known, there is still ongoing research to use concise object detection models for the detection of Thai-language characters. Among many object detection models, YOLOv8 is one of the most well-known deep learning models primarily designed for real-time object detection. YOLOv8 can be trained with smaller datasets compared with TrOCR, reducing the training time, and requiring fewer hardware resources.

2. RELATED WORKS

OCR can convert text data that has been scanned or photographed into digital text data that can be searched for or edited later. Among several Thai OCR research, typed character recognition was tried to recognize font styles from Thai printed documents [1]. There are three major commercial Thai OCR software programs available to the public at present. But none of them can preserve the type styles of the original document image such as bold, italics, and normal styles into the output text file. This paper presents the technique for preserving the specified Thai type styles by applying a specific preprocessing with a supervised neural networks (NNs) learning algorithm. Experiments have been conducted, and the results show that the proposed technique effectively preserves the type styles of

Thai typed fonts from the original document image into the output text file.

The TrOCR model was proposed in reference [2] titled "Transformer-based Optical Character Recognition with Pre-trained Models." TrOCR consists of an image transformer encoder and an autoregressive text transformer decoder to perform OCR. Existing approaches for text recognition are usually built based on CNN for image understanding and RNN for character-level text generation. In addition, another language model is usually needed to improve the overall accuracy as a post-processing step. In this paper, an end-to-end text recognition approach with pre-trained image Transformer and text Transformer models is proposed, which leverages the Transformer architecture for both image understanding and word piece-level text generation. The TrOCR model can be effective when pre-trained with large-scale synthetic data and fine-tuned with human-labeled datasets, which requires a lot of dataset preparation efforts. Another difficulty of transformer-based research is that they heavily depend on big enough training datasets and hence can show weakness recognizing unusual sequences or combinations of characters.

3. Methodology and Experiment

This section will describe the steps of using the YOLOv8 model for letter detection, starting from preparing the data for training, data generation for training, training, and testing.

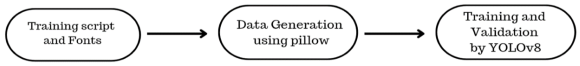


Fig. 1. A Generated training image sample.

3.1 Training script and Fonts

To train OCR for Thai documents, it is essential to include training scripts from various fields, diverse printed fonts, and Thai and English characters and sentences covering most character combinations. Not only is it necessary to have sentences from different fields to avoid encountering unfamiliar words or characters after training, but the frequency of each character's appearance should also not be biased. Due to the excessive number of possible character combinations, manual labeling is practically impossible. Therefore, it is intended to use the font rendering feature of the Python library Pillow to convert characters and sentences into images for use as training data. Consequently, the fonts must be sufficiently numerous to correspond to most real documents. It is known that TrOCR has also used over 5,000 fonts. Though the use of a Thai dictionary seems essential, we have chosen to use sentence data from Wikipedia and tried to increase the diversity of words in the dataset. We have categorized the dataset into seven different categories, including Art, Country, Biography, Movie, Science, Sport, and Vehicle, resulting in a total of 200 script files. The more diverse the words or letters in the dataset, the better it is for the training process and the model's performance.

3.2 Data Generation using pillow

When using the Pillow library to output fonts, the size of the font and the image are also important. Since the Yolo model uses a square input image size and the model, we are experimenting with in our paper also uses an input image size of 640 x 640, selecting appropriate fonts and generated image sizes is crucial. The font size was set between 30 and 34, and pages were formatted to contain 20 characters per line and 12 lines per page. With this font output, the size of the output image, including a small margin, is around 740 pixels in both width and height. About 350 publicly available printed Thai fonts were used. This study focused on Thai character recognition and, due to limited training time, did not adequately consider English scripts and fonts.

Considering the nature of Thai characters, which often involve vowels and accent marks at various positions, each consonant, vowel, and accent mark was labeled separately, and complex characters containing these were also classified as separate labels for training. Therefore, if the label is recognized as having a complex character, the system is also capable of recognizing the consonants, vowels, and accent marks within that region. To mimic the background color of typical scanned documents, Gaussian noise was added to each page after printing the character fonts. Figure 1 shows a sample of the generated training character images. Different colors represent vowels and accent characters of consonants.

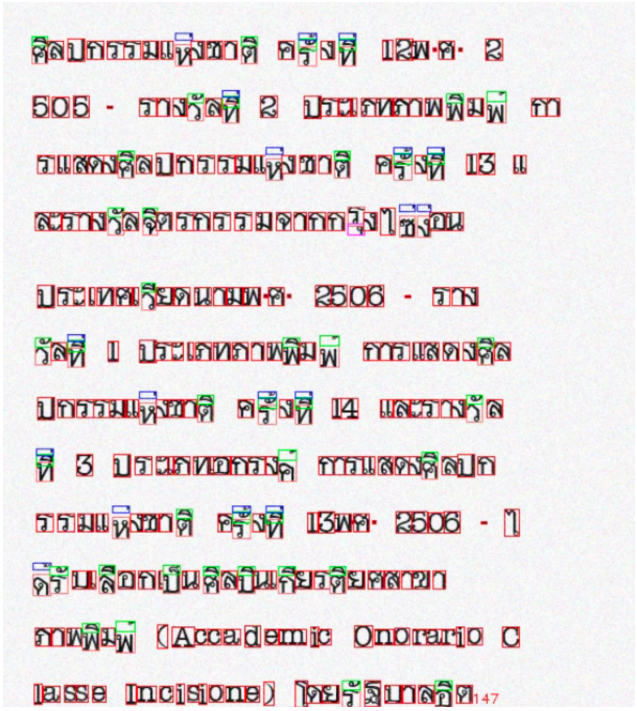


Fig. 2. A Generated training image sample.

3.3 Training and Validation by YOLOv8

After obtaining the training dataset from the generation process, we proceed to divide the data for training and validation. The initial training of this experiment utilizes YOLOv8 to train the model. The training process involves the creation or refinement of the machine

learning model, allowing the model to learn from the data to predict or process new data efficiently and accurately. During training, the machine learning model adjusts its parameters using the training data, consisting of examples with correct answers. As the model learns to an acceptable level of accuracy, it can then be applied to predict or process new data without known answers.

The training process is often referred to as 'model tuning,' where the model's parameters are fine-tuned using training data. This process involves minimizing the discrepancy between the model's predictions and the correct answers, leading to the adjustment of parameters to reduce error. Training may use either training data or testing data to evaluate the model's performance. This process can be time and resource-intensive, particularly with large datasets and complex models. Considering the performance of the GPU and the training time, the Yolov8 Small model was used for training. Due to the small GPU memory size, the batch size was set to 4. To purely evaluate the recognition performance of YOLO, no post-processing was conducted. Input images are normalized to a size of 640x640 during training or prediction, which can lead to some characters not being recognized properly when an entire page is used as input. This issue can be addressed in two ways.

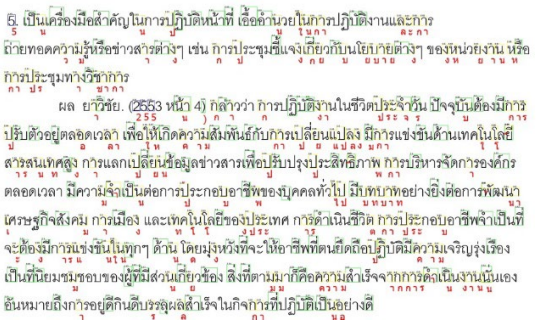


Fig. 3. Recognition result before using SAHI

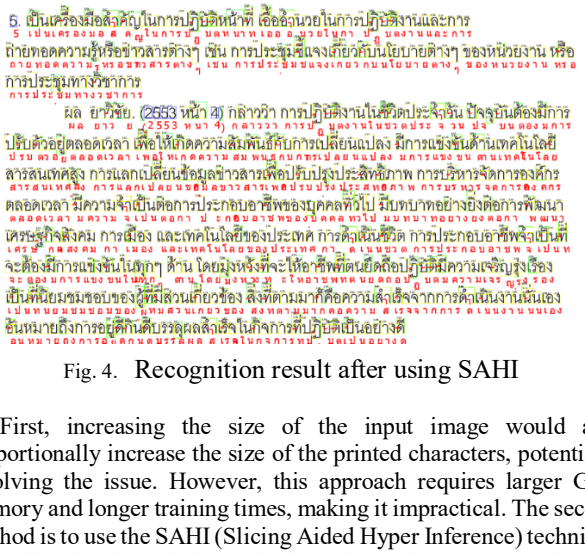


Fig. 4. Recognition result after using SAHI

First, increasing the size of the input image would also proportionally increase the size of the printed characters, potentially resolving the issue. However, this approach requires larger GPU memory and longer training times, making it impractical. The second method is to use the SAHI (Slicing Aided Hyper Inference) technique [3], which involves dividing the input image into several sub-areas based on size for recognition and then combining the results. This is especially useful when there are numerous objects to be detected in a single image, and the objects are relatively small compared to the image.

4. Experimental Results

To show the performance of the Yolov8 OCR model, since we couldn't find any open dataset with character-level labeling, we trained and tested the mode with self-made images. In the image generation step for model training, 30 different fonts are used, each with a font size between 30 and 34. The generated images have a fixed size of 640 pixels in width and 640 pixels in height. Each line of the generated image consists of 20 characters per line and 12 lines per page, with random spacing between each character. For the model training step, a total of 31,039 images are used, with 41 images reserved for validation. The computer specifications used for model training include an Intel Core i7-10700 CPU, 16 GB of RAM, Nvidia GeForce GTX 1080 Ti GPU, and a 2TB HDD. The experimental result is shown in Table 1. Most of the errors are from unlearned English characters since we used Thai sentences from Thai Wikipedia, which have very few English words.

Table 1 Experimental Result

Total char.	S	D	I	S+D+I	CER (%)
1,724	11	22	21	54	3.13%
1,804	10	44	13	67	3.17%
1,487	8	35	20	63	4.23%
1,728	8	46	10	64	3.70%
1,527	8	26	10	51	3.33%
812	4	15	15	34	4.18%

As can be seen from the images before and after using SAHI, it can be observed that object detection in the images has significantly improved. SAHI divides the large image into segments for object detection, effectively breaking down the image into smaller parts for detection. When dealing with large-sized images, dividing the image into segments for detection is more effective. From the experimental results, it can be concluded that YOLOv8 can effectively detect Thai characters. When compared it with TrOCR, YOLOv8 has significant advantages. Although TrOCR can recognize both printed and handwritten characters, it needs a dataset of up to 2 million images. On the other hand, YOLOv8 can achieve good detection performance with just 31,039 images, making it practical for real-world use. Furthermore, TrOCR uses a time-consuming transcriber dataset, which is larger than YOLOv8's dataset, and it recognizes characters one line at a time. This can be considered a drawback and a reason for choosing YOLOv8 over TrOCR in this research.

5. CONCLUSION

In this paper, the popular object recognition model Yolov8 was used for OCR document recognition, specifically for Thai printed documents, which have many combinable characters

and various vowel positions. Compared to transformer-based algorithms, Yolov8 demonstrated successful recognition rates based on its advantages of requiring relatively smaller training datasets, faster recognition times, and lower hardware requirements.

It has been a common misconception that object recognition models like YOLO are not suitable for recognizing small characters, such as those in documents. However, the concept of 'large' and 'small' is relative, not absolute. The application of methods like SAHI has proven that these models can indeed be effectively used for document recognition. Future developments aim to create a more practical system using a wider range of fonts and sufficient training data.

6. REFERENCE

[1] Supachai Tangwongsan, Buntida Suvacharakulton, (2023). "Relization of a high performance bilingual OCR system for thai-english printed documents". IEEE Xplore, https://ieeexplore.ieee.org/document/5587781 .

[2] Minghao Li, Tengchao Lv, Jingye Chen, Lei Cui, Yijuan Lu, Dinei Florencio, Cha Zhang, Zhoujun Li, Furu Wei. (2021). "Transformer-based Optical Character Recognition with Pre-trained Models" Models. arXiv, https://doi.org/10.48550/arXiv.2109.10282 .

[3] Fatih Cagatay Akyon, Sinan Onur Altinuc, Alptekin Temizel, (2022). "Slicing Aided Hyper Inference and Fine-tuning for Small Object Detection". arXiv, https://arxiv.org/abs/2202.06934 .

Enhance Connectivity: RSMA Application and Unveiling the Power of CF-mMIMO Deployment

Ayesha Siddiqa, Mahnoor Ajmal, Malik Muhammad Saad, Muhammad Ashar Tariq and Dongkyun Kim
School of Computer Science and Engineering
Kyungpook National University, Daegu, Korea.
{asiddiqa, mahnoor.ajmal, maliksaad, tariqashar, dongkyun}@knu.ac.kr

***Abstract*—The extensive growth in mobile data traffic requires crucial modification in the network. The densification of the network by creating the small cell increases the inter-cell interference due to a small gap in the frequency allocation. Moreover, the small cells also create co-channel interference and adjacent channel interference. Considering these facts, the Cell-Free massive MIMO (CF-mMIMO) network has been proposed to eliminate cell boundaries and deploy access points (APs) closer to users. However, the multiple access technique enabling users to AP communication is still unclear. So, in our manuscript, we have deployed the hybrid multiple access technique rate-splitting multiple access technique (RSMA) in CF-mMIMO and cellular networks. We have shown the comparative analysis of latency and network load in CF-mMIMO and 5G cellular networks. Moreover, we cannot replace the current 5G cellular network to deploy the CF-mMIMO networks practically. Considering the fact we have also proposed the integrated topology for CF- mMIMO networks.**

Keywords—Space Division Multiple Access (SDMA), Rate Splitting Multiple Access (RSMA), massive MIMO (mMIMO), Cell-Free Networks, Cellular Networks.

I. INTRODUCTION (*HEADING 1*)

The demand for mobile data is increasing rapidly worldwide, and it is predicted to grow more than seven times in the next five years [1]. Mobile networks are being made denser to handle this demand, reusing the same frequencies more efficiently. However, this densification is causing a problem called inter-cell interference (ICI), which makes it harder for users to get a good signal at the edges of cell coverage areas [2]. Although there have been attempts to minimize this interference in novel 5G systems, there remains a significant difference in performance between users near the cell center and those at the cell edge.

A promising solution, the Cell-Free Massive multiple-input multiple-output (CF-mMIMO) network, is being explored to address the issue of varying performance and ICI [3]. CF-mMIMO networks randomly deploy several access points (APs) working in collaboration, transcending traditional cell boundaries and enhancing coverage. This innovative approach improves signal quality, even in crowded areas, by reducing interference at coverage edges through AP cooperation and techniques like precoding, beamforming, and dynamic collaboration.

However, integrating cell-free systems into existing networks poses significant engineering challenges. The architecture of CF-mMIMO needs to be deployed with the already existing cellular network architecture. However, the cooperative communication between APs and base stations (BS) creates signal interference because the deployment architecture differs in both architectures. Moreover, it is also required to analyze the hybrid multiple access technique for CF-mMIMO networks to enable users to communicate with the architecture. There is a research gap for the multiple access technique deployment in CF-mMIMO networks.

Given the current urgency, a practical approach is needed to integrate CF-mMIMO networks into established cellular frameworks. This approach should not only exceed the limitations of 5G but also usher in the era of B5G. We have investigated the rate splitting multiple access (RSMA) application in conventional cellular and CF-mMIMO networks to meet these requirements. Moreover, we added the adaptive modulation scheme in both networks and applied the above mentioned techniques to analyze the results. In contrast with current CF-mMIMO networks, we have presented our integrated architecture, which combines the 5G network properties and CF-mMIMO network properties. The goal is to enable the practical development of CF-mMIMO networks in current 5G networks.

The main contributions of the proposed scheme include:

- We have proposed the adaptive modulation technique in CF-mMIMO networks to improve the network efficiency.
- We have applied RSMA in CF-mMIMO and cellular networks to analyze the user downlink (DL) communication.
- We have integrated the CF-mMIMO networks in 5G cellular networks for practical deployment.
- We have specifically specified the tasks of APs, BSs and BBUs to differentiate the task processing further.

The rest of the manuscript is divided into five sections. Section 2 summarizes the previous contributions. Section 3 explains the proposed scheme and its working. Section 4 explains the integrated system and distribution of tasks in each unit. Section 5 provides a discussion of simulated results. Lastly, we have concluded our work and provided some future directions.

II. PREVIOUS CONTRIBUTIONS

Numerous studies have been documented in the literature in the realm of transitioning from cellular networks to cell-free networks for enhanced data rate support. However, a critical practical concern persists regarding the deployment of these cell-free networks. In a seminal work by the authors of[4], a proposition was put forth for deploying cell-free systems. Recognizing the formidable task of entirely replacing the existing cellular architecture, the recommendation is to have cell-free networks coexist alongside conventional cellular networks.

The prevailing cellular system is characterized by a network-centric architecture that furnishes several advantages, including the efficient utilization of radio resources, simplified system administration, and streamlined system construction [5]. The proliferation of cell sites has been propelled by the introduction of small cells and the advent of cloud/centralized radio access networks (C-RANs), consequently amplifying network capacity to accommodate surges in traffic demand [6, 7]. Nonetheless, the cell-based architecture is not without its intrinsic challenges. Users situated at the periphery of a cell experience a weak signal from their serving cellular base station (CBS) and contend with potent interference emanating from neighboring cells. A CBS typically encompasses a macro base station (MBS) and a small base station (SBS) in a Heterogeneous Network (HetNet). To surmount this issue, an array of interference mitigation techniques have been devised, including inter-cell interference coordination (ICIC), coordinated multipoint (CoMP), and MIMO beamforming techniques [8]. However, as network density escalates, inter-cell interference intensifies, and the interference patterns become more diverse, rendering the application of interference mitigation technology increasingly challenging [9].

Access Points (APs) are dispersed arbitrarily across the coverage area and tethered to a Baseband Unit (BBU) pool through a fronthaul link in a cell-free system. This configuration empowers a user to receive robust signals coherently from the surrounding cluster of APs [10,11]. The formation of AP clusters is dynamically adaptable based on the prevailing circumstances. This flexibility enables a user to evade issues at cell edges and ensures the provision of high-quality mobile services. However, if the number of APs falls short of the user population, the cell-free system may encounter difficulties achieving the anticipated uniform performance. The sheer magnitude of the challenge posed by installing many APs renders the swift replacement of a cellular system with a cell-free system virtually infeasible within a short timeframe.

III. Proposed Scheme

Initially, the SDMA was deployed in conventional CF-mMIMO networks to enable users to maintain connections with different APs simultaneously [12]. It spatially separates the users in the network using dedicated beams. But it works well in fewer user scenarios. However, as the number of users increases, the distance between beams decreases, creating signal interference on fronthaul. Moreover, SDMA requires accurate channel information using a direct line of sight (LOS). Considering this fact, we have implemented the 6G recommended rate splitting multiple access (RSMA) technique. RSMA divides the user data rate into common and private streams. Every user can decode the common stream using successive interference cancellation (SIC), and the private stream can only be decoded by the dedicated users, considering other signals as noise. The high transmission power is assigned to the common stream, and low transmission power is assigned to the private stream. RSMA's encoding scheme is applied on APs, and the decoding scheme is applied on users. The goal of RSMA is to enable the users to decode the signals using SIC or consider other signals as noise. It increases the received data rate.

The working of the proposed scheme is shown in Fig. 1. APs and BSs initially follow the 5G initialization process to assign the network resources to the users. After channel allocation, APs and BSs simultaneously broadcast the RSMA-encoded data signal to users. The distance between APs and users is less as compared to BSs. The user decodes the data signal using the RSMA-decoding Data process. Users received the signal from APs early as compared to BSs. The received signal is precoded so the user will first receive the common message. The user decodes the common message by using SIC. After that, it decodes its private message only by considering other signals as noise. The packet reception ratio increases by considering two different signal interference cancellation techniques. Furthermore, it increases reliable data transmission and decreases the downlink (DL) data transmission delay of 6G networks because the user receives two different data streams in one signal.

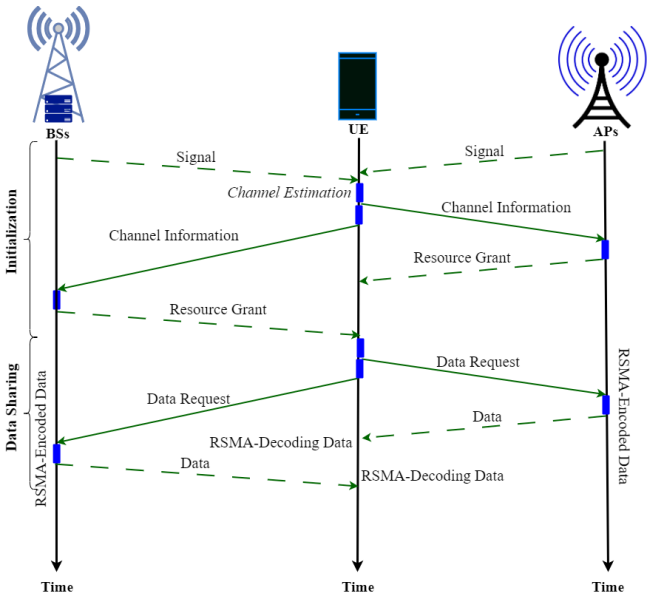


Fig. 1. Working of the Proposed Scheme

IV. INTERGTRATED TOPOLOGY FOR CF-MMIMO NETWORKS (I-CF-MMIMON)

This section discusses the proposed Integrated Topology for CF-mMIMO Networks (I-CF-mMIMON); the APs are randomly distributed in the network. There is a wireless connection between APs, BSs, and CPU. APs and BS locally estimate the channel of the user. APs forward the channel estimation and user details to BSs. Each BS has a local BBU, and the BBU maintains the user channel information and user ID. Moreover, It also saves the popular data locally. Therefore, APs can directly forward the data requests to BSs, and local BBU satisfies the data request locally. If BBU does not have the requested data, it forwards the request to the CPU through intermediate APs (I-APs). I-APs have direct access to the CPU. CPU monitors and controls the network. It observes the network frequently; it increases the APs deployment in case of dense user scenarios. CPU has a large repository to maintain extensive safety or non-safety information for a long time. The complete I-CF-mMIMON is shown in Fig. 2.

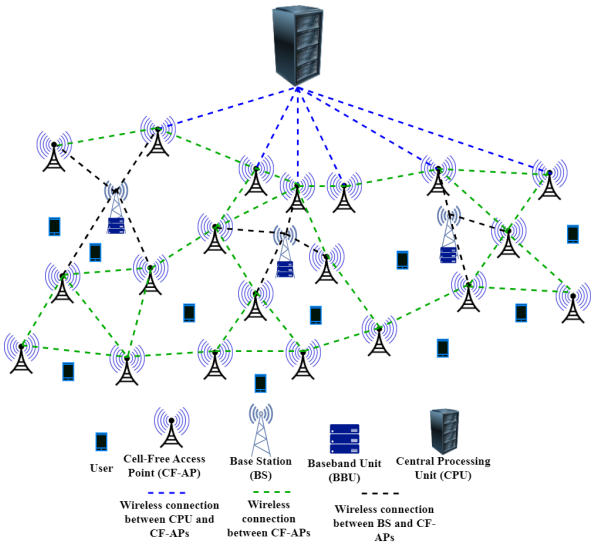


Fig. 2. Proposed Integrated Topology for CF-mMIMO Networks

V. Performance Evaluation

We have discussed the simulation results in this section. The RSMA is applied in cellular and CF-mMIMO networks using MATLAB. For simulation, we have considered Table I parameters. For a cellular network, seven cells are created, and each cell has a random number of users ranging from 10-40. Each cell has its base station (BS) equipped with a baseband unit (BBU). BS is equipped with an N number of antennas. However, APs with several antennas are randomly deployed in the network. Users ranging from 10-40 follows the Fig. 1 steps for connectivity. we have considered RSMA application in CF-mMIMO networks scenario as well as in cellular networks scenario.

TABLE I. Simulation Parameters

Parameters	***Values***
APs	40
Users	10
AP Antennas	10
User Antennas	3
Carrier Frequency	280e9
SNR	0:10:50
Transmission Power on Each AP	10 dB
AP Communication Range	10 meters
BS Communication Range	600 meters

We firstly evaluated the latency in DL communication using RSMA in both networks. The distance between APs and users are minimum and we have also limit the AP to use connectivity for scalability [12]. Moreover, we have added the adaptive modulation technique, it enables the users to receive accurate data within less latency. However the same methodology is used for cellular network but the distance between BS and user in high as compare to CF-mMIMO network. Furthermore, only one BS is serving users so the latency increases as compared to CF-mMIMO networks. The results fluctuate because the users use adaptive modulation techniques, creating an imbalance in Fig 3. The results show that even with adaptive modulation, the RSMA application in CF-mMIMO networks shows less latency than cellular networks.

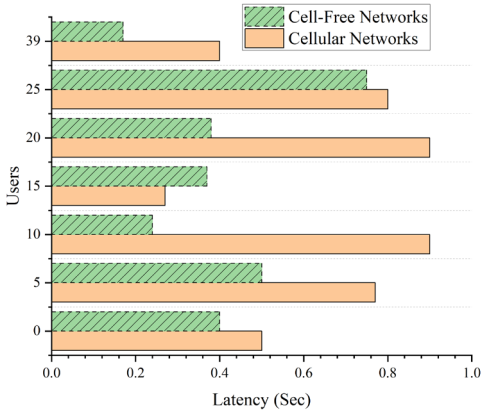


Fig. 3. Analysis of RSMA Latency in Two Different Scenarios

Secondly, we have evaluated the network load created by RSMA application in both scenarios. The results in Fig. 4 show the significant difference in network load between both scenarios. The network load on cellular network increases linearly on increasing number of users, because all the antennas are broadcasting the signal to users. However, in CF-mMIMO networks we have utilized the scalable properties to limit the number of APs to user connection. As a result shows, the network load in CF-mMIMO networks is minimal as compared to cellular networks.

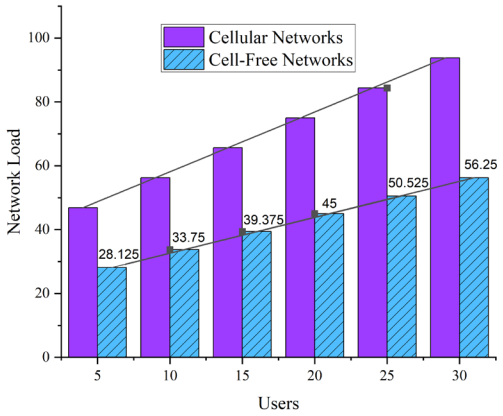


Fig. 4. Comparative Analysis of Network Load

VI. Conclusion

In this manuscript, we have proposed the RSMA application in CF-mMIMO networks and cellular networks. Initially, we applied a 5G initialization process to connect the user to the AP and BS connection. Later, we discussed the RSMA application to DL the data to the user. We have proposed the integrated topology to deploy the CF-mMIMO

in cellular networks to show the practical deployment perspective. Lastly, we have analyzed the network load and latency of RSMA applications in CF-mMIMO and cellular networks. In future, we intend to simulate the proposed topology, apply RSMA and analyze the DL data sharing.

ACKNOWLEDGEMENT

This work was supported in part by the Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education under Grant NRF-2018R1A6A1A03025109. This research has supported by the National Research Foundation of Korea(NRF) grant funded by the Korea government(MSIT), (NRF-2022R1A2C1003620).

REFERENCES

[1] F. Alriksson, L. Bostr¨om, J. Sachs, Y.-P. E. Wang, and A. Zaidi, "Critical iot connectivity ideal for time-critical communications," Ericsson technology review, vol. 2020, no. 6, pp. 2–13, 2020.

[2] A. Grassi, G. Piro, G. Boggia, M. Kurras, W. Zirwas, R. S. Ganesan, K. Pedersen, and L. Thiele, "Massive mimo interference coordination for 5g broadband access: Integration and system level study," Computer networks, vol. 147, pp. 191–203, 2018.I. S. Jacobs and C. P. Bean, "Fine particles, thin films and exchange anisotropy," in Magnetism, vol. III, G. T. Rado and H. Suhl, Eds. New York: Academic, 1963, pp. 271–350.

[3] H. A. Ammar, R. Adve, S. Shahbazpanahi, G. Boudreau, and K. V. Srinivas, "User-centric cell-free massive mimo networks: A survey of opportunities, challenges and solutions," IEEE Communications Surveys & Tutorials, vol. 24, no. 1, pp. 611–652, 2021R. Nicole, "Title of paper with only first word capitalized," J. Name Stand. Abbrev., in press.

[4] T. Kim, H. Kim, S. Choi, and D. Hong, "How will cell-free systems be deployed?," IEEE Communications Magazine, vol. 60, no. 4, pp. 46–51, 2022.

[5] D. Tse and P. Viswanath, Fundamentals of wireless communication. Cambridge university press, 2005.

[6] A. Checko, H. L. Christiansen, Y. Yan, L. Scolari, G. Kardaras, M. S. Berger, and L. Dittmann, "Cloud ran for mobile networks—a technology overview," IEEE Communications surveys & tutorials, vol. 17, no. 1, pp. 405–426, 2014.

[7] J. G. Andrews, S. Buzzi, W. Choi, S. V. Hanly, A. Lozano, A. C. Soong, and J. C. Zhang, "What will 5g be?," IEEE Journal on selected areas in communications, vol. 32, no. 6, pp. 1065–1082, 2014.

[8] B. Soret, A. De Domenico, S. Bazzi, N. H. Mahmood, and K. I. Pedersen, "Interference coordination for 5g new radio," IEEE Wireless Communications, vol. 25, no. 3, pp. 131–137, 2017.

[9] D. L´opez-P´erez, M. Ding, H. Claussen, and A. H. Jafari, "Towards 1 gbps/ue in cellular systems: Understanding ultra-dense small cell deployments," IEEE Communications Surveys & Tutorials, vol. 17, no. 4, pp. 2078–2101, 2015.

[10] O. T. Demir, E. Bj¨ornson, L. Sanguinetti, et al., "Foundations of user-centric cell-free massive mimo," Foundations and Trends® in Signal Processing, vol. 14, no. 3-4, pp. 162–472, 2021.

[11] E. Bj¨ornson and L. Sanguinetti, "Scalable cell-free massive mimo systems," IEEE Transactions on Communications, vol. 68, no. 7, pp. 4247–4261, 2020.

[12] A. Siddiqa, J. Seo, M. M. Saad, B. Jeong, and D. Kim, "Scalable simultaneous connectivity with multiple aps and virtual ues in cf-networks," in Proceedings of the 38th ACM/SIGAPP Symposium on Applied Computing, pp. 1766–1771, 2023.

Wi-Fi Aware based V2P communication for VRU Safety Service

Donghyun Jeon
School of Computer Science and Engineering
Kyungpook National University
Daegu, Korea
jdh0830@knu.ac.kr

Euiri Jo
School of Computer Science and Engineering
Kyungpook National University
Daegu, Korea
joeuiri0402@knu.ac.kr

Seri Park
School of Computer Science and Engineering
Kyungpook National University
Daegu, Korea
psr0527@knu.ac.kr

Junho Seo
School of Computer Science and Engineering
Kyungpook National University
Daegu, Korea
jhseo@knu.ac.kr

Dongkyun Kim
School of Computer Science and Engineering
Kyungpook National University
Daegu, Korea
dongkyun@knu.ac.kr

***Abstract*— Unlike traditional ITS (Intelligent Transport Systems), which primarily focused on providing processed traffic information, C-ITS(Cooperative-Intelligent Transport Systems) has evolved into a new paradigm that prioritizes the safety of road users by delivering real-time services. By C-ITS, proactive measures can be taken before accidents occur to prevent them from happening in the first place. C-ITS enables the provision of safety services not only for motorized vehicles but also for VRUs (Vulnerable Road Users). Every single individual is a potential vulnerable road user, as it can be assumed that every person occasionally crosses a street or intersection either by foot or by other non-motorized means such as cycling. Typically, VRUs account for almost half of road accident victims. Therefore, in order to prevent fatalities or serious injuries caused by traffic accidents, we need safety measures for VRUs. Among the V2X(Vehicle-to-Everything) communication methods for providing VRU safety services, V2I(Vehicle-to-Infrastructure) and V2V(Vehicle-to-Vehicle) communication are wireless communication, so the link does not always remain stable, and interference may occur. This causes delays in service. To prepare for such a problem, we should conduct research on VRU safety services using V2P(Vehicle-to-Pedestrians). Wi-Fi Aware is a way to enable communication without the help of infrastructure, so Wi-Fi-aware is appropriate for the V2P communication-based VRU safety service we are trying to implement. Therefore, this paper proposes a VRU safety service system using Wi-Fi-aware technology.**

Keywords*— *VRU, V2X, Wi-Fi Aware, C-ITS

I. INTRODUCTION

C-ITS is a system that provides real-time information to drivers about surrounding traffic conditions, potential accidents such as sudden stops or falling objects, and other safety hazards while they are driving. It is based on V2V and V2I communication, which enables the collection of real-time traffic information. This information is then shared with the C-ITS center and coordinated accordingly. Unlike traditional ITS, which primarily focused on providing processed traffic information, C-ITS has evolved into a new paradigm that prioritizes the safety of road users by delivering real-time services. ITS had limitations in immediate response and mainly focused on post-accident management. However, with C-ITS, proactive measures can be taken before accidents occur to prevent them from happening in the first place. C-ITS enables the provision of safety services not only for motorized vehicles but also for VRUs. By utilizing C-ITS technology, safety can be extended to pedestrians, cyclists, and other vulnerable users of the road.[1]

As people nowadays are encouraged to cycle (including electric bicycles) or walk short distances for health or environmental reasons, the number of 'traffic participants' will not decrease. Moreover, every single individual is a potential vulnerable road user, as it can be assumed that every person occasionally crosses a street or intersection either by foot or by other non-motorized means such as cycling. VRUs includes pedestrians, cyclists (including eBikes), motorcyclists, road workers, wheelchair users, scooter, skateboard and segway users. Typically, VRUs account for almost half of road accident victims. In 2017, pedestrians accounted for 21 of road fatalities in the EU, for example, while motorcycles, bicycles and mopeds made up 26%.[2] VRUs may not be constant, such as sudden changes in speed, and the direction of movement may also be difficult to predict (such as children), and since there are various types of VRUs, it is still a calling issue to provide safety services in consideration of all their characteristics. To finally obtain the vision of zero fatalities or severe injuries in road traffic, we need safety measures for VRU.

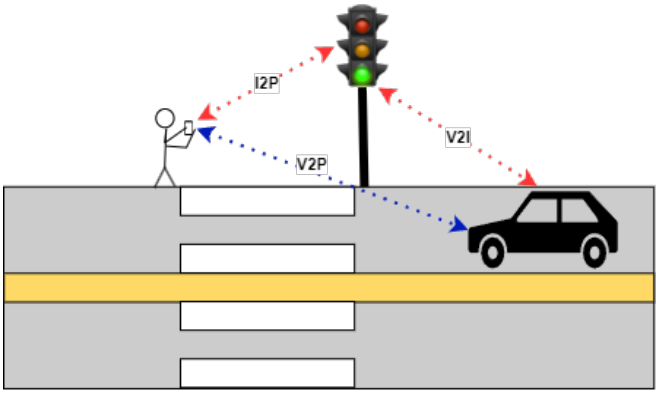


Figure 1. V2X communication methods.

To provide VRU safety services, a combination of various communication technologies and methods can be used. V2X communication methods for VRU safety services include V2I, V2V, and V2P.[3] Since V2I and V2V communications are wireless, the link may not always maintain a stable state and interference can occur, leading to service delays. To address these issues, research on VRU safety services using V2P communication is necessary. Several technologies can be applied to V2P communication. Firstly, there is the IEEE 802.11p-based DSRC (Dedicated Short-Range Communications) protocol. DSRC offers low latency even with high mobility and advantages such as an acceptable range of up to 1 km.

However, it requires smartphones with high computing capabilities and activation of DSRC technology in both vehicles and mobile phones. Secondly, there is the cellular approach which provides the longest range due to existing infrastructure but suffers from long waiting times, making it challenging to apply in safety-critical systems. Thirdly, Bluetooth technology allows easy connection with smartphones and wireless earphones but has a very short range (50m) and may not be suitable for dynamic situations. Lastly, Wi-Fi-based solutions do not require infrastructure deployment and are easy to set up but have limited coverage (100-200m) and experience delays due to handover processes.[4] Among these options, Wi-Fi Aware stands out as it enables communication without relying on infrastructure assistance. Therefore, Wi-Fi Aware is suitable for our proposed VRU safety service based on V2P communication. In this paper, we utilize Wi-Fi Aware technology to calculate distances between VRUs and vehicles through direct communication. Based on the calculated distance, we propose a VRU safety service system where pedestrians receive warning notifications.

II. Related Work

Wi-Fi Aware, also known as NAN (Neighbor Awareness Networking), is a technology within Wi-Fi that provides a low-power discovery mechanism.[5][6]Wi-Fi Aware allows devices using this technology to form independent networks without relying on internet connectivity or assistance from surrounding infrastructure. The NAN protocol is designed to operate in the background, allowing devices to maintain the network by periodically exchanging control messages with nearby nodes. This enables NAN-enabled devices to remain in a pre-connection state, providing low latency during data exchange. To use the NAN protocol, Android 8.0 or higher (API level 26) is required. The NAN protocol operates based on cluster formation, and the components and roles within a cluster are as follows: 1) Anchor Master Node: Forms and maintains the NAN cluster, 2) Master Node: Can replace the Anchor Master and independently form a NAN cluster through its master role, 3) Non-Master Sync Node: Participates in propagating synchronization beacon frames within the NAN cluster but is relieved from transmitting discovery beacon frames, 4) Non-Master Non-Sync Node: Relieved from both propagating synchronization beacon and discovery beacon frames in the network.

To summarize briefly how cluster formation works in NAN: A device that wishes to perform the role of a Master periodically transmits discovery beacon messages during Discovery Window (DW), which lasts for 16ms, outside of DW intervals (524ms). If it is forming the cluster for the first time, it assumes the role of an Anchor Master node. Devices that receive these discovery beacon messages and want to join the cluster perform time clock synchronization procedures with the transmitting device. Synchronize beacons are transmitted within DW time, and if a device intends to operate as non-master non-sync, it ignores these beacons. Nodes that have completed time synchronization set their NAN Master Rank value based on factors such as master preference value (which may change over time), random factor (periodically updated), and device's MAC address. To conserve energy, Master Rank values can dynamically change. If there is no Anchor Master while having high Master Rank values among devices, then the device with highest Master Rank becomes an Anchor Master. A device with a high enough Master Rank can also perform master roles by forming another cluster and provide NAN services to devices participating in its own cluster through aforementioned transmission of discovery beacon messages. If participating in the network as a non-master sync node, it performs only propagation of synchronize beacons within the network to maintain clock synchronization across all nodes in the cluster.

III. Proposed System

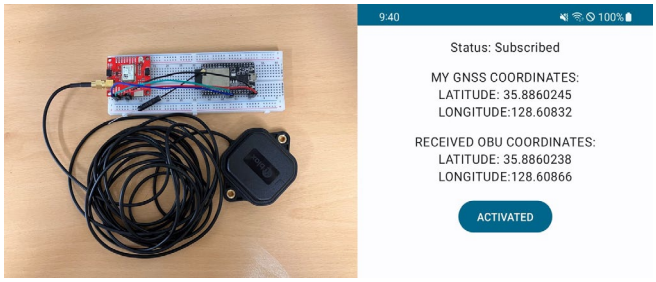


Figure 2. OBU and VRU Smartphone Application

The components of the proposed system are shown in Figure 2. The proposed system consists of an on-board unit (OBU) in the vehicle and a smartphone device. The OBU consists of a microcontroller unit (MCU) and a Global Navigation Satellite System (GNSS) module. The GNSS module is capable of comm

TABLE I.
Details of Devices Used in the Proposed System

System	Model Name
MCU	ESP32 WROVER-DEVKITC-VIE
GNSS	NEO-M9N (SMA)
Antenna	GNSS Multi-Band Magnetic Mount Antenna
VRU Smartphone Device	SM-996N, SM-A426N

OBU

The MCU employed the ESP32 module, which supports the Wi-Fi NAN protocol.[8] The Ublox NEO-M9N(SMA) was utilized for the GNSS module, with detailed

configurations executed using Ublox's U-center.[9] Configuration parameters and their assigned Option/Value are shown in Table II.

TABLE II.
CONFIGURATION PARAMETERS AND SETTINGS FOR UBLOX NEO-M9N

Parameter	Option/Value
UBX-CFG-ANT	Enable all
UBX-CFG-NAV5-Navigation Modes Dynamic Mode	4- Automotive
UBX-CFG-NAVX5-AssistNow Autonomous Use AssistNow Autonomous	Enable Max acceptable orbit error : 10m
UBX-CFG-PMS-Setup ID	0-Full Power
UBX-CFG-RATE-Measurement Period	100ms

A. VRU Smartphone Application

The smartphone application for VRU was crafted utilizing the Kotlin programming language and leverages Jetpack Compose to construct a cohesive and composable UI. Permissions required are declared within the AndroidManifest.xml file as follows:

- ACCESS FINE LOCATION
- ACCESS COARSE LOCATION
- ACCESS WIFI STATE
- CHANGE WIFI STATE
- ACCESS NEARBY STATE
- NEARBY DEVICES
- INTERNET

The application requires the android.hardware.wifi.aware feature, thereby excluding devices that do not support this specific feature from running the application. In the build.gradle(app) file, the compile SDK is designated as 34, the minimum SDK as 29, and the target SDK as 33. To facilitate access to location information, the implementation line 'com.google.android.gms:play-services-location:21.0.1'.

B. Operational Principle of the System

Figure. 3 shows the operational sequence of the proposed system. Initially, the OBU activates the GNSS module to intercept NMEA codes. It identifies strings beginning with GNGGA, which encapsulate crucial data including time, latitude, longitude, GNSS quality, and the number of interacting satellites. The OBU, acting as a publisher in the NAN protocol[10], initiates the service and awaits location information request messages from Subscribers.

It is assumed that the VRU smartphone device is already aware of the service name published by the MCU. Based on this service name, the VRU device conducts a service search. If it detects a matching OBU service within its NAN communication range, it subscribes to the service and sends a location information provision request message to the OBU using the SubscribeDiscoverySession parameter. The OBU only receives notifications when the VRU smartphone device subscribes and sends messages to the OBU service.

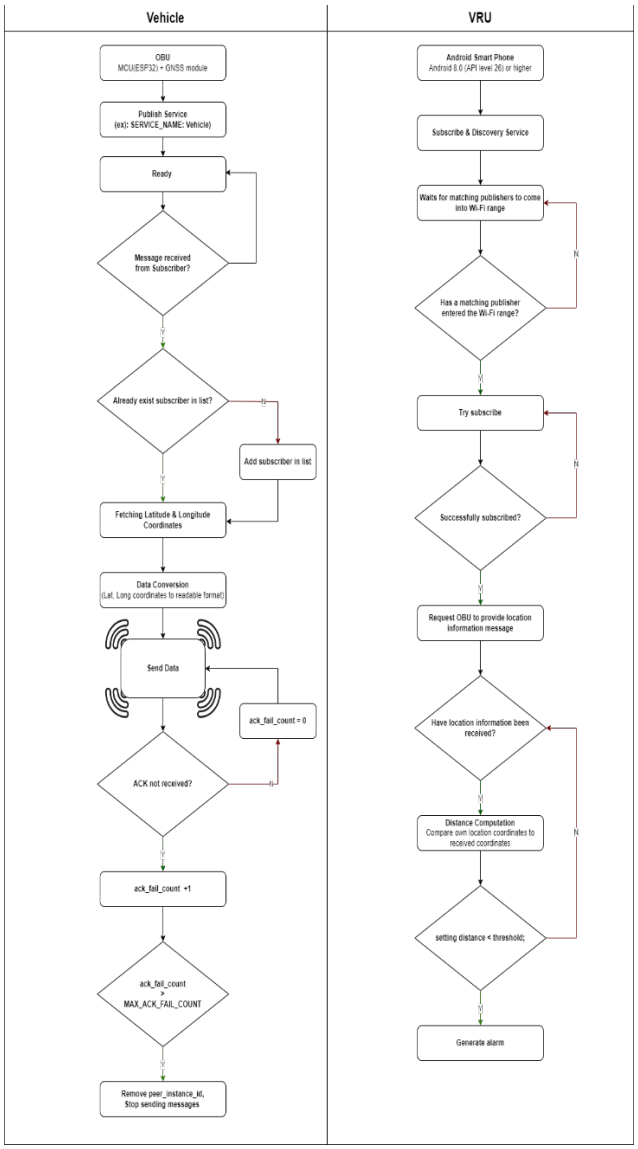


Figure 3. Operational Principle of the System.

Upon receiving a location information provision request, the OBU checks for the existence of the peer instance id of the transmitting VRU smartphone device in its list. If the id is not present, it is added to the transmission list. The OBU then transmits the location information to the VRU smartphone devices in the list, extracted from the received GNGGA string from the GNSS module.

When VRU smartphone devices receive location information from the OBU, each device computes the distance between its location and that of the OBU using the FusedLocationProviderClient API and sends an ACK message to the OBU. If the computed distance is below the warning distance predetermined by the VRU, the approaching vehicle is signaled to the VRU through sound, vibration, and screen notification. When a vehicle moves away from the VRU and exits the feasible communication range, it accumulates the number of times it has transmitted location information to the VRU but has not received an ACK in ack fail count. If ack fail count exceeds MAX ACK FAIL COUNT, the peer instance id of the corresponding VRU smartphone device is removed from the list, and the transmission of messages to that device ceases. If an ACK is

received before ack fail count exceeds MAX ACK FAIL COUNT, ack fail count is reset to zero.

IV. Conclusion

In this paper, we propose a safety service for VRUs (Vulnerable Road Users) using Wi-Fi Aware communication technology. The developed system consists of an OBU (on-board unit) and a VRU smartphone device. Using Wi-Fi Aware's Neighbour Awareness Networking (NAN) protocol, the OBU acts as a publisher, initiating the service and listening for messages requesting location information. The VRU smartphone device, acting as a subscriber, identifies and subscribes to the OBU's service and sends messages requesting location information. Upon receiving these requests, the OBU adds them to the transmission list and forwards the location information obtained from the GNSS module to the VRU smartphone device. The VRU device then uses the received location information to calculate the distance to the OBU and, if it is within the pre-set warning distance, alerts the user to the approach of a vehicle through sound, vibration and screen notifications.

Acknowledgment

This research was supported by the MSIT(Ministry of Science and ICT), Korea, under the Innovative Human Resource Development for Local Intellectualization support program(IITP-2023-RS-2022-00156389) supervised by the IITP(Institute for Information & communications Technology Planning & Evaluation). In addition, This work was supported in part by the Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education under Grant NRF-2018R1A6A1A03025109.

References

[1] Ministry of Land Infrastructure and Korea Expressway Corporation Transport. Cooperative-intelligent transport systems. URL https://www.c-its.kr/english/ getMain.do.

[2] 5G Automotive Association (5GAA). Vulnerable road user protection, 2020. URL https://5gaa.org/content/ uploads/2020/08/5GAA XW3200034 White Paper Vulnerable-Road-User-Protection.pdf.

[3] Mario H. Castaneda Garcia, Alejandro Molina-Galan, ˜ Mate Boban, Javier Gozalvez, Baldomero Coll-Perales, Taylan S¸ahin, and Apostolos Kousaridas. A tutorial on 5g nr v2x communications. IEEE Communications Surveys Tutorials, 23(3):1972–2026, 2021. doi: 10.1109/COMST. 2021.3057017.

[4] Ahmad Kabil, Khaled Rabieh, Faisal Kaleem, and Marianne A. Azer. Vehicle to pedestrian systems: Survey, challenges and recent trends. IEEE Access, 10:123981– 123994, 2022. doi: 10.1109/ACCESS.2022.3224772.

[5] Shubham Saloni and Achyut Hegde. Wifi-aware as a connectivity solution for iot pairing iot with wifi aware technology: Enabling new proximity based services. In 2016 International Conference on Internet of Things and Applications (IOTA), pages 137–142, 2016. doi: 10.1109/ IOTA.2016.7562710.

[6] Dhaval Prajapati. Method for nearby product marketing using wi-fi aware technology. In 2021 International Conference on Intelligent Technologies (CONIT), pages 1–5, 2021. doi: 10.1109/CONIT51480.2021.9498283.

[7] Wi-Fi Alliance. Wi-fi aware certified products. URL https://www.wi-fi.org/product-finder-results? sort by=certified&sort order=desc&categories=3& subcategories=23&certifications=757.

[8] ESPRESSIF. Wi-fi aware(nan). URL https://docs.espressif.com/projects/esp-idf/en/latest/ esp32/api-reference/network/esp nan.html.

[9] ublox. u-center user guide. URL https: //content.u-blox.com/sites/default/files/u-center Userguide UBX-13005250.pdf.

[10] Android API Reference. Wi-fi aware overview. URL https://developer.android.com/guide/topics/connectivity/ wifi-aware.

Predictive Analytics and Performance Optimization in Mobile Networks using Machine Learning

Shiplu Das
Computer Science and Engineering
Adamas University West Bengal, India
shiplud63@gmail.com

Rutvij H. Jhaveri*
Department of Computer Science and Engineering, School of Technology,Pandit Deendayal Energy University, India,
rutvij.jhaveri@sot.pdpu.ac.in

Buddhadeb Pradhan
Department of Computer Science and Engineering, University of Engineering and Management. West Bengal, India
buddhadebpradhan@gmail.com

Rajkumar Singh Rathore
Cardiff School of Technologies, Cardiff Metropolitan University, United Kingdom. line 4: City, Country
rsrathore@cardiffmet.ac.uk

Chunhua~Su
Division of Computer Science, University of Aizu Fukushima 965--8580, Japan, and with the Peng Cheng Laboratory, Shenzhen 518066, China.
chsu@u-aizu.ac.jp

***Abstract*— Mobile and wireless networks are becoming increasingly complex and dynamic, with a growing number of connected devices and ever-increasing data demands. To address the challenges posed by these evolving networks, predictive analytics and performance optimization techniques have gained considerable attention. This paper explores the applications of machine learning in predictive analytics and performance optimization for mobile/wireless networks. The study uses the Least Absolute Shrinkage and Selection Operator (LASSO) algorithms to extract meaningful insights from historical network data and predict future network behavior. These predictions enable proactive decision-making, resource allocation, and optimization strategies to enhance network performance and user experience. These models enable accurate predictions by analyzing historical data, allowing network operators to allocate resources efficiently and mitigate potential issues before they impact network performance. The paper also discusses the implications of adopting machine learning techniques in realistic network environments, addressing data privacy, security, and interpretability of models. Further research is required to address the challenges and exploit the full potential of machine learning in this context.**

Keywords—Edge Computing, Mobile Edge Computing, Machine Learning Algorithm, Artificial Intelligence.

I. INTRODUCTION

In recent years, the convergence of machine learning techniques and mobile/wireless technologies has significantly transformed various aspects of our lives. Machine learning, a branch of artificial intelligence, has emerged as a powerful tool to analyze and make sense of large volumes of data generated by mobile devices and wireless networks. This fusion has paved the way for exciting developments and applications in mobile/wireless research, revolutionizing industries such as telecommunications, healthcare, transportation, and more. The widespread adoption of smartphones and the proliferation of wireless networks have generated unprecedented data and transmitted every second. Mobile devices continuously collect data on user behavior, location, preferences, and interactions, while wireless networks facilitate seamless communication between devices. Integrating machine learning into mobile and wireless research has given rise to multiple applications, from network optimization and resource allocation to intelligent decision-making and user behavior analysis. By leveraging the power of machine learning algorithms, researchers and practitioners can extract valuable insights from vast amounts of data collected from mobile devices and wireless networks. These insights enhance network performance, improve user experiences, and innovative services and application development. Network optimization is a prominent application of machine learning in mobile and wireless research. Traditional network optimization techniques often rely on static and rule-based approaches, which may adapt poorly to dynamic and heterogeneous network environments. Machine learning, on the other hand, enables the development of adaptive and self-learning algorithms that can automatically optimize network parameters, predict network congestion, and dynamically allocate network resources based on real-time data. This results in more efficient and reliable network performance, reduced latency, and improved overall network capacity. Machine learning techniques can be applied to detect and prevent various security threats, such as intrusion detection, malware detection, and anomaly detection. Section 1 presents the introduction of mobile network using machine learning algorithm. Section 2 presents the related work on mobile network using machine learning algorithm then presents the basic mobile edge computing architecture. Section 3 examines the methodology with experimental results. Section 4 presents the simulation results of different network configuration parameters and next section presents the conclusion of the paper.

*Corresponding Author

A. Contribution of the research work

Our research study makes the following significant contributions:

- The main focus is to provide a general understanding of the recent research and help newcomers to understand the essential modules and trends.
- We present a machine learning algorithm on mobile edge computing.
- Discuss advanced architecture secure mobile edge computing.
- Finally, we highlight a number of research concerns as well as possible future research areas for MEC architecture.

Machine learning algorithms have the potential to revolutionize mobile edge computing by improving resource allocation, task offloading, predictive analytics, anomaly detection, QoS optimization, and energy efficiency. By leveraging the power of machine learning, MEC systems can become more intelligent, adaptive, and capable of meeting the diverse needs of mobile applications and users.

II. Related work

Machine learning applications in mobile/wireless research have garnered significant attention from researchers and practitioners in recent years. Numerous studies have demonstrated the potential of machine learning techniques in various domains within this field, leading to advancements in network optimization, security, context-aware computing, and predictive analytics. This section provides an overview of some notable research works that highlight the diverse applications of machine learning in mobile/wireless research. Researchers have employed machine learning algorithms in network optimization to improve resource allocation and management in wireless networks. Wang et al. [1] proposed a deep reinforcement learning framework for dynamic spectrum allocation in cognitive radio networks. The study demonstrated that the proposed algorithm could effectively adapt the spectrum allocation strategy based on network conditions and user demands, resulting in improved spectrum utilization and reduced interference. Machine learning has also been applied to enhance the security of mobile/wireless systems. Amouri et al. [2] denveloped a machine learning-based intrusion detection system for mobile ad hoc networks. The study utilized a support vector machine classifier to analyze network traffic and identify malicious activities. The proposed system achieved high detection rates while minimizing false positives, showcasing the potential of machine learning in enhancing network security. In context-aware computing, machine learning techniques have been leveraged to enable personalized and adaptive experiences for mobile users. Mukherjee et al. [3] presented a deep learning-based approach for human activity recognition using smartphone sensor data. The study demonstrated that the proposed model could accurately classify various activities, such as walking, running, and cycling, based on sensor readings. This work opens up possibilities for developing context-aware applications that can understand and respond to user activities in real-time. Moreover, machine learning has been instrumental in predictive analytics for mobile/wireless systems. Liu et al. [4] utilized machine learning algorithms to predict network traffic demand in wireless communication networks. The study developed a prediction model that could forecast traffic patterns by analyzing historical data, allowing for proactive resource allocation and optimization. The results showed significant improvements in network performance and resource utilization. Furthermore, researchers have explored the application of machine learning in optimizing power consumption in mobile/wireless devices. Yang et al. [5] proposed a machine learning-based power control mechanism for energy-efficient communication in cellular networks. The study utilized a reinforcement learning algorithm to adaptively adjust the transmission power of mobile devices, reducing energy consumption while maintaining satisfactory communication quality. Pradhan et al. [6] investigate current research problems connected to 5G technology, this article addresses Industry 4.0, 5G standards, and new research in future wireless communications. The study also proposes a new architecture for Industry 4.0 and 5G-enabled intelligent healthcare systems. When compared to existing methodologies, our strategy yields an average improvement of 25%. Sun et al. [7] explored the application of deep learning techniques in wireless resource management. They developed a deep Q-network (DQN) model that dynamically allocates resources to different users in a multi-user wireless network, achieving efficient resource utilization and enhanced network throughput. Casas et al. [8] presented a machine learning-based approach for anomaly detection in mobile network traffic. Their proposed system used a combination of unsupervised learning algorithms to detect and classify abnormal behaviors in real-time, thereby enhancing the security of mobile networks against various threats and attacks. Jhaveri et al. [9] summarises TRS-PD's attack-pattern discovery technique, trust model, and routing mechanism in order to defeat adversaries who follow certain attack patterns as well as other adversaries. Experiments with network simulator-2 reveal the optimal parameter values for various network settings. Hakim et al. [10] proposed a deep learning-based approach for context-aware gesture recognition on mobile devices. Their system employed convolutional neural networks (CNNs) to analyze sensor data from mobile devices and accurately recognize hand gestures, enabling natural and intuitive interactions with mobile applications. Akour et al. [11] investigated the use of machine learning for predicting mobile user behavior. They employed long short-term memory (LSTM) networks to model and forecast user activity patterns, enabling proactive resource allocation and personalized services in mobile networks. Jhaveri et al. [12] proposed framework is tested on a hardware testbed using a Ryu SDN controller. Our findings reveal that, when compared to baseline mechanisms, the proposed framework enhances network resilience by up to 21% and bandwidth by up to 111 Mbps under various network situations. Furthermore, thorough experimental emulations using the Mininet tool demonstrate the scalability of the proposed framework. Hu et al. [13] presented a study on predictive analytics for mobile network performance optimization. They proposed a machine learning-based framework that utilized historical network data to predict network congestion and proactively allocate network resources. Their approach achieved

significant improvements in network performance, reducing latency and enhancing user satisfaction. Xu et al. [14] advocated predicting NSCP performance. The SI-CNN model is a hybrid of SqueezeNet and InceptionNet, with four convolution layers that all use the same convolution model. They use a two-one convolution and a three-branch convolution for the first two layers, which not only increases the number of channels but also extracts additional information. They use the same framework but different convolution kernels for the last two layers. Basic Mobile Edge computing Architecture and its Working principle:

Fig. 1 Basic Edge Computing architecture

Fig.1 presents the basic architecture of Edge Computing. But Mobile edge computing is situated within the IoT ecosystem, bridging the gap between the cloud and the things layer architecture. It establishes communication channels from diverse edge devices such as smartphones and smart TVs, which send data to edge nodes. Subsequently, the data is forwarded from the edge nodes to the cloud data center for analysis and processing. Edge computing brings several advantages, including diminished latency and reduced transmission expenses. It offers benefits such as data integrity, heightened availability, quicker access speeds, enhanced security, scalability, and cost-efficiency.

A. Machine Learning for Mobile Edge computing and Iots:

Utilizing machine learning within mobile edge computing is highly valuable for performing distributed training on edge devices, facilitating effective processing and analysis. Additionally, it assists in the identification and comprehension of emerging technological trends. Machine learning proves especially advantageous in extracting accurate information from IoT devices deployed in intricate architectural settings.

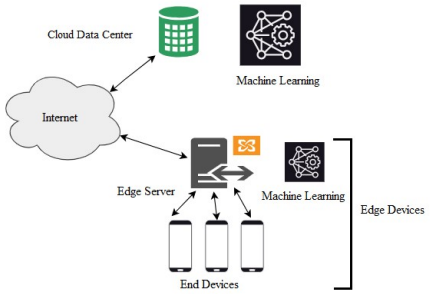


Fig. 2 Mobile Edge Computing architecture using machine learning.

Its versatility spans different categories of edge devices, including end devices, edge servers, and cloud data centers, empowering intelligent decision-making and data processing at various levels of the network infrastructure. Fig.2. represents the Mobile Edge Computing Architecture using Machine Learning. The application of machine learning on mobile edge devices has proven extremely valuable, delivering exceptional performance. As edge devices necessitate real-time processing and raise privacy concerns, incorporating machine learning involves intelligent decision-making and training directly at the edge. Currently, deep learning is widely utilized at the edge across various domains, including computer vision, smart parking systems, and content delivery networks. Machine learning plays a significant role in object detection and image classification tasks. The integration of machine learning techniques has made computer vision methods pervasive in our everyday lives. Examples include face recognition on devices, photo editing applications, and vision assistance in autonomous vehicles, all of which rely on computer vision technology.

III. Proposed Challenges of Machine Learning at Mobile Edge Computing

The general procedure of data transmission and processing in Mobile Edge Computing (MEC) adheres to a standardized workflow, where users generate Electrocardiogram (ECG) data through sensors integrated within the IoT layer. After the initial data generated from the IoT layer, the ECG data is transmitted via the 5G network to the Mobile Edge Computing (MEC) servers for subsequent data processing. Various tasks are carried out within the MEC layer, such as AI-based automatic diagnosis of ECG, data reception, temporary data storage, data retrieval, and generation of diagnostic results. In recent years, there has been notable progress in Artificial Intelligence (AI) technology, accompanied by significant advancements in Machine Learning models. The integration of neural networks, 5G transmission, and the MEC layer has the potential to greatly enhance the overall system efficiency. This research paper utilizes the LASSO regression model for predictive modeling. The cost function of the LASSO regression model can be represented as follows in equation (1).

$$min\frac{1}{2}\sum(WtXi + b.Yi)\text{^}2 + \alpha||W|| \quad (1)$$

Where W denotes the eigenvector, Yi denotes the response Variable, and R denotes the regular parameter. Assume that a set of eigenvectors W. After training LASSO regression model parameter (W, b), the equation can be equation (2).

$$WtXi + b \quad (2)$$

The expression of the minimum optimization problem can be expressed in an MEC architecture. An optimization technique called ADMM (Alternating Direction Method of Multipliers) is proposed as a distributed solution for solving the LASSO regression problem. ADMM converts equation (2) into

equation (3) in this context. Mobile edge computing, in combination with advanced regression techniques such as LASSO (Least Absolute Shrinkage and Selection Operator) regression, has emerged as a promising approach for optimizing data processing and analysis in the mobile and Internet of Things (IoT) domains. Integrating mobile edge computing and LASSO regression provides a powerful framework for addressing challenges related to response time, bandwidth consumption, and privacy concerns in IoT models as equation (3).

$$min\frac{1}{2}\sum\sum(WtXij + b.Yij)\text{^}2 + \alpha||W|| \qquad (3)$$

This paper explores the application of mobile edge computing using LASSO regression, highlighting its potential benefits and discussing its implications in various domains, including healthcare, smart cities, and industrial automation. This research aims to demonstrate the effectiveness and versatility of mobile edge computing combined with LASSO regression as a robust solution for optimizing data processing and predictive modeling in resource-constrained.

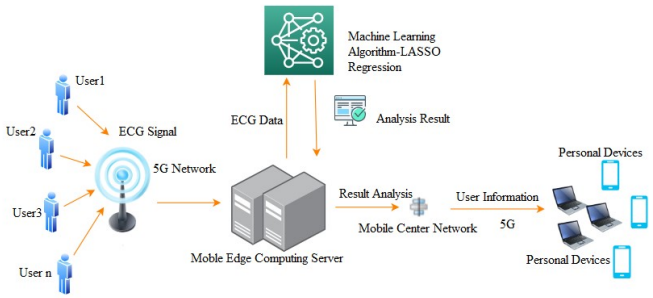


Fig. 3 Proposed LASSO regression on Mobile Edge computing

Fig.4. represents the LASSO regression on Mobile Edge computing. By employing the ADMM (Alternating Direction Method of Multipliers) optimization technique, the N subproblems within all IoT devices can be effectively addressed. Specifically, the local subproblem of the device needs to be solved in this context. Evaluate the performance metrics of the LASSO regression model, such as accuracy, precision, recall, F1 score, or mean squared error (MSE). Proposed Challenges of Machine Learning at Mobile Edge Computing. Cross-validation is a widely used machine learning technique to estimate a model's performance and select hyperparameters. In the context of mobile edge computing (MEC) using LASSO regression, selecting the optimal value of the regularization parameter alpha through cross-validation can help improve the accuracy and generalization of the model. Cross-validation involves splitting the available data into multiple subsets or folds. The model is then trained on a subset of the data and evaluated on the remaining fold. This process is repeated several times, with different subsets used for training and evaluation each time. The performance metrics obtained from each iteration are averaged to estimate the model's overall performance. The goal of the LASSO algorithm is to minimize the objective function, which is a combination of the least squares error and the L1 regularization term.

TABLE I. ALPHA SELECTED BY CROSS-VALIDATION ON MEC USING LASSO REGRESSION.

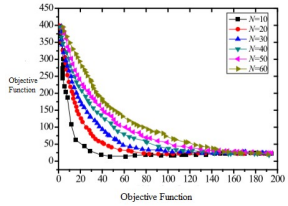
Description	**MEC**		
	Alpha	***No. Of Nonzero Coefficient***	***CV mean deviance***
First Alpha	.8865434	0	0.0885
First Alpha	.8865434	0	0.0885
Last Alpha	.0002456	10	-1.563
Alpha Before	.0876567	5	0.476
Alpha After	.0786542	5	0.479
Selected Alpha	.0766542	5	0.457

Table 1. represents the Alpha Selected by Cross-Validation on MEC using LASSO Regression. The integration of mobile edge computing with LASSO regression presents significant opportunities and challenges in various domains. This discussion will explore the implications and potential benefits of using LASSO regression using mobile edge computing. One of the key advantages of mobile edge computing using LASSO regression is the potential for improved model accuracy. Alpha selection by cross-validation on MEC using LASSO regression is a systematic and data-driven approach to finding the best regularization parameter for LASSO models. It ensures that the model's performance is optimized for making accurate predictions while preventing overfitting and selecting only the most relevant features. This technique is widely employed in various fields, including statistics, machine learning, and data science, to build robust and interpretable predictive models.

TABLE II. X-AXIS AND THE CORRESPONDING AVERAGE MSE VALUES ON THE Y-AXIS.

MSE	
Alpha	***Average MSE***
0.01	0.125
0.05	0.110
0.1	0.105
0.5	0.112
1.0	0.120

The analysis of alpha selected by cross-validation on MEC using LASSO regression is a comprehensive process that involves exploring, evaluating, and selecting the most suitable regularization parameter for building predictive models. It not only results in accurate predictions but also enhances model interpretability and feature selection, making it a valuable tool in various fields, including statistics, machine learning, and data science.



. Fig. 4 The result of N of MEC environment

This is particularly valuable in applications where precision and reliability are critical, such as healthcare diagnostics or industrial automation. Furthermore, the reduced computational complexity offered by LASSO regression contributes to the efficiency of mobile edge computing systems.

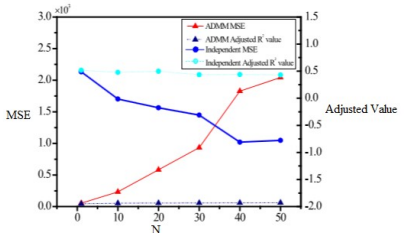


Fig. 5 The result of α

In a Multi-Access Edge Computing (MEC) environment, the term "N" typically refers to the number of MEC nodes or instances deployed within the network. The result of having multiple MEC nodes can vary depending on the specific use case and deployment scenario. Fig 4. and fig 5. represents the results of alpha and N of MEC Environment. It's important to note that the specific results and benefits of an N-node MEC environment can vary depending on factors such as network architecture, application requirements, resource allocation, and optimization strategies implemented. By leveraging the computational capabilities of edge devices, data processing, and analysis can be performed closer to the source, reducing the reliance on centralized cloud resources. This decentralized approach not only enhances the speed of decision-making but also reduces latency, making it suitable for time-sensitive applications. Privacy and security concerns are addressed through mobile edge computing.

IV. Conclusion

In conclusion, the integration of mobile edge computing with LASSO regression offers a powerful and promising solution for optimizing data processing and predictive modeling in mobile and IoT environments. This combination provides several key advantages, including improved model accuracy through feature selection, reduced computational complexity, enhanced privacy and security, real-time decision-making, and resource efficiency. Mobile edge computing using LASSO regression enables more accurate and reliable predictive models by leveraging the feature selection capabilities of LASSO regression. Detection is reduced by performing data processing and analysis closer to the network edge, and real-time decision-making becomes feasible, making it suitable for time-sensitive applications. In summary, mobile edge computing using LASSO regression represents a promising paradigm for addressing the challenges of data processing and analysis in mobile and IoT environments. Its advantages in accuracy, efficiency, privacy, and real-time decision-making make it a compelling solution with wide-ranging applications and the potential to drive advancements in various industries.

References

[1] Wang, Y., Ye, Z., Wan, P., & Zhao, J. (2019). A survey of dynamic spectrum allocation based on reinforcement learning algorithms in cognitive radio networks. Artificial intelligence review, 51, 493-506J. Clerk Maxwell, A Treatise on Electricity and Magnetism, 3rd ed., vol. 2. Oxford: Clarendon, 1892, pp.68–73.

[2] Amouri, A., Alaparthy, V. T., \& Morgera, S. D. (2020). A machine learning based intrusion detection system for mobile Internet of Things. Sensors, 20(2), 461.

[3] Mukherjee, D., Mondal, R., Singh, P. K., Sarkar, R., \& Bhattacharjee, D. (2020). EnsemConvNet: a deep learning approach for human activity recognition using smartphone sensors for healthcare applications. Multimedia Tools and Applications, 79, 31663-31690.

[4] Liu, Y., Bi, S., Shi, Z., \& Hanzo, L. (2019). When machine learning meets big data: A wireless communication perspective. IEEE Vehicular Technology Magazine, 15(1), 63-72.

[5] Yang, H., Zhong, W. D., Chen, C., Alphones, A., \& Xie, X. (2020). Deep-reinforcement-learning-based energy-efficient resource management for social and cognitive internet of things. ieee internet of things journal, 7(6), 5677-5689.

[6] Pradhan, B., Das, S., Roy, D. S., Routray, S., Benedetto, F., & Jhaveri, R. H. (2023). An AI-Assisted Smart Healthcare System Using 5G Communication. IEEE Access.

[7] Sun, H., Chen, X., Shi, Q., Hong, M., Fu, X., & Sidiropoulos, N. D. (2017, July). Learning to optimize: Training deep neural networks for wireless resource management. In 2017 IEEE 18th International Workshop on Signal Processing Advances in Wireless Communications (SPAWC) (pp. 1-6). IEEE.

[8] Casas, P., Fiadino, P.,\& D'Alconzo, A. (2016, April). Machine-Learning Based Approaches for Anomaly Detection and Classification in Cellular Networks. In TMA.

[9] Jhaveri, R. H., Patel, N. M., Zhong, Y., & Sangaiah, A. K. (2018). Sensitivity analysis of an attack-pattern discovery based trusted routing scheme for mobile ad-hoc networks in industrial IoT. IEEE Access, 6, 20085-20103..

[10] Hakim, N. L., Shih, T. K., Kasthuri Arachchi, S. P., Aditya, W., Chen, Y. C., & Lin, C. Y. (2019). Dynamic hand gesture recognition using 3DCNN and LSTM with FSM context-aware model. Sensors, 19(24), 5429.

[11] Akour, I., Alshurideh, M., Al Kurdi, B., Al Ali, A., & Salloum, S. (2021). Using machine learning algorithms to predict people's intention to use mobile learning platforms during the COVID-19 pandemic: machine learning approach. JMIR Medical Education, 7(1), e24032.

[12] Jhaveri, R. H., Ramani, S. V., Srivastava, G., Gadekallu, T. R., & Aggarwal, V. (2021). Fault-resilience for bandwidth management in industrial software-defined networks. IEEE Transactions on Network Science and Engineering, 8(4), 3129-3139.

[13] Hu, S., Ouyang, Y., Yao, Y. D., Fallah, M. H., & Lu, W. (2014, May). A study of LTE network performance based on data analytics and statistical modeling. In 2014 23rd Wireless and Optical Communication Conference (WOCC) (pp. 1-6). IEEE.

[14] Xu, L., Zhou, X., Li, X., Jhaveri, R. H., Gadekallu, T. R., & Ding, Y. (2021). Mobile collaborative secrecy performance prediction for artificial IoT networks. IEEE Transactions on Industrial Informatics, 18(8), 5403-5411.

Performance Optimization of React Applications: A Comparative Analysis of AWS EKS and EC2 Deployments for Enhanced Web Metrics

Haider Rizvi, Asad Mehboob
National Information Technology Board [NITB], Pakistan

***Abstract*—In recent past, Cloud computing has revolutionized the IT industry giving new concept to enterprise computing system with competitive advantages for the businesses to deploy and manage their applications and services over the Cloud. With the invent of Cloud based services the competition between businesses for fastest delivery of service is now the need of time. Hence, to cater the increasing demand for faster and more efficient computing solutions, new paradigms such as Edge Computing and Microservices platform have gained prominence. This paper specifically emphasis the comparative analysis of Edge Computing and Microservices, with focus on the potential to enhance performance in cloud-based environments. In order to achieve this, the study probe into the key principles, architectural considerations, and practical implications of both approaches. Our Research would differentiate the strengths and limitations of each, offering valuable insights for decision-makers seeking to optimize their cloud computing strategies and how to harness these technologies for improved cloud performance.**

***Index Terms*—AWS EKS (Elastic Kubernetes Services), EC2 (Elastic Compute Cloud), Virtual DOM(Document Object Model)**

I. Introduction

IN today's rapidly evolving digital landscape, optimizing the performance of cloud-based applications has become paramount for businesses seeking to deliver seamless and efficient services to their customers. Cloud computing, with its promise of scalability, flexibility, and accessibility, has revolutionized the way applications are developed and deployed. In this context, the integration of cutting-edge technologies, such as Edge Computing and Microservices, has emerged as a pivotal approach for enhancing the performance of cloud-based applications. [1] This research explores the comparative analysis of Edge Computing and Microservices, with a focus on their applications and implications for performance optimization in cloud computing environments.

1) Problem Statement

Optimizing the performance of React applications poses challenges, particularly in the context of traditional EC2 deployments on AWS. As these applications grow in complexity, issues related to scalability and resource utilization become apparent. The problem at hand involves assessing the limitations of EC2 instances in efficiently handling the dynamic nature of React applications. To address this, we aim to conduct a comparative analysis between AWS Elastic Kubernetes Service (EKS) and EC2 deployments, seeking to identify the deployment strategy that best enhances web metrics and ensures optimal user experiences. This research addresses a critical gap in understanding how various AWS deployment methods impact React application performance, offering practical insights for developers and organizations relying on cloud-based solutions.

2) Research Objectives

This research aims to compare AWS Elastic Kubernetes Service (EKS) and traditional EC2 deployments for React-based applications, with a focus on improving web metrics. Specific objectives include evaluating performance metrics on EKS and EC2, identifying best deployment practices, and determining the strategy that optimizes web metrics. The study seeks to offer actionable insights for developers and organizations navigating React application performance in the AWS cloud.

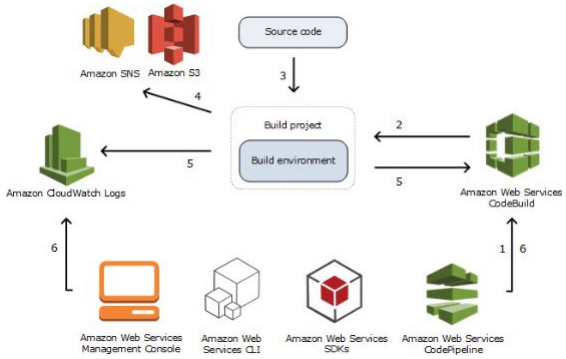


Fig. 1. System Model

3) Scope of Study

This section delineates the boundaries and focus areas of the research. The investigation is centered on React applications deployed on the AWS cloud environment, specifically comparing two deployment methods: AWS Elastic Kubernetes Service (EKS) and traditional EC2 instances. The scope encompasses an in-depth analysis of performance metrics, including response time, scalability, and resource utilization. Limitations include an exclusive examination of AWS services and a concentration on web metrics optimization. This study provides

valuable insights into the deployment strategies within the specified scope, aiming to contribute practical knowledge for developers and organizations utilizing cloud-based solutions for React applications.

The literature surrounding the performance optimization of React applications within the context of cloud deployments, particularly focusing on AWS Elastic Kubernetes Service (EKS) and EC2 instances, encompasses several key areas of inquiry.

4) React Application Performance Optimization

Scholarly works in the realm of React application development emphasize the critical importance of performance optimization to ensure a responsive and seamless user experience. Best practices often include code splitting, lazy loading, and efficient state management to enhance the application's efficiency and reduce load times. Framework-specific optimizations are crucial for leveraging React's virtual DOM and efficiently managing component rendering.

5) Cloud Deployments with AWS

The literature reveals a significant shift towards cloud-based deployments, particularly within the AWS ecosystem, as organizations seek scalable and cost-effective solutions. [2] AWS EC2 instances have been a popular choice for hosting web applications, providing flexibility and control over the underlying infrastructure. However, there is a growing interest in container orchestration solutions like AWS EKS, which offers a more streamlined approach to managing containerized applications at scale. [2]

6) Performance Metrics for Web Applications

Research on web metrics highlights the key indicators for evaluating the performance of web applications. Response time, scalability, and resource utilization emerge as critical metrics that directly impact user experience. [2] Monitoring and analyzing these metrics are imperative for identifying bottlenecks and optimizing the overall performance of web applications.

7) Comparative Analysis of Cloud Deployments

While various studies delve into the benefits of cloud deployments, few conduct a detailed comparative analysis between different deployment options within the AWS environment. This research gap underscores the need for a study that specifically investigates the performance of React applications when deployed on AWS EKS compared to traditional EC2 instances. Understanding the nuanced differences in scalability, resource utilization, and overall web metrics can guide developers and organizations in selecting the most effective deployment strategy. [3]

8) Previous Studies and Research Gaps

While the literature on React optimization and AWS cloud deployments exists, there is a distinct lack of comprehensive comparative analyses specifically tailored to React applications within the AWS ecosystem. This research seeks to bridge this gap by providing an in-depth exploration of the performance implications associated with AWS EKS and EC2 deployments for React applications. [4]. The insights derived from this study aim to contribute to the existing body of knowledge and provide actionable recommendations for developers navigating the complexities of performance optimization in cloud environments. [5]

II. Methodology

To conduct a rigorous comparative analysis of AWS Elastic Kubernetes Service (EKS) and EC2 deployments for optimizing the performance of React applications and enhancing web metrics, the following methodology was employed.

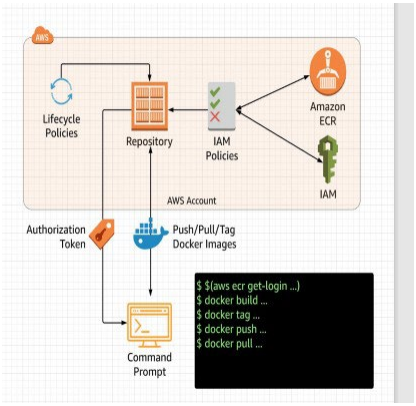


Fig. 2. Application deployed using AWS EKS

1) Experimental Setup

A React application representative of a typical web scenario was selected. The application included dynamically loaded components, state management, and various UI elements to simulate real-world conditions.

2) AWS EKS Deployment

The React application was containerized using Docker, and the containerized application was deployed on AWS EKS. The deployment leveraged Kubernetes orchestration for container scaling and management. Key metrics monitored during the EKS deployment included:

Response Time: Measured as the time taken to load and render the application. Scalability: Evaluated by increasing the number of concurrent users accessing the application. Resource Utilization: Monitored CPU and memory usage of the deployed containers.

3) AWS EC2 Deployment

A parallel deployment of the same React application was performed on traditional EC2 instances within the AWS cloud. The EC2 deployment served as a benchmark for comparison. The following metrics were observed during the EC2 deployment: Response Time: Recorded to assess the

application's responsiveness. Scalability: Tested by gradually increasing the application's load. Resource Utilization: Monitored CPU and memory utilization on EC2 instances.

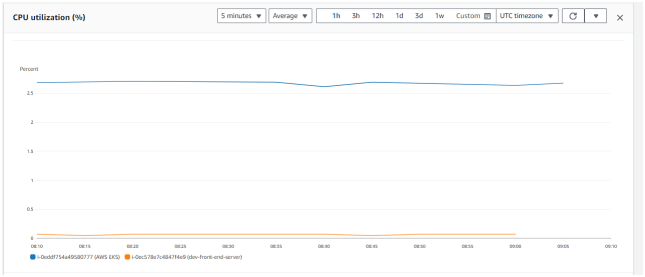


Fig. 3. CPU Utilization

4) Data Collection

Performance metrics were collected using monitoring tools such as AWS CloudWatch, Prometheus, and Grafana. Continuous monitoring ensured real-time data capture throughout the experimentation phase.

5) Comparative Analysis Framework

A robust comparative framework was established, enabling a side-by-side analysis of the performance metrics between AWS EKS and EC2 deployments. The comparison aimed to identify the deployment option that excelled in terms of response time, scalability, and resource efficiency.

6) Result Values

As a preliminary demonstration, suppository values were assigned to illustrate potential outcomes:

AWS EKS Deployment:
Response Time: 120 milliseconds
Scalability: Efficiently handled up to 1,000 concurrent users
Resource Utilization:
CPU Utilization: 30 percent
Network In:
Bytes: 5,000 KB/s
Packets: 1,000 packets/s
Network Out:
Bytes: 8,000 KB/s
Packets: 1,500 packets/s

AWS EC2 Deployment:
Response Time: 180 milliseconds
Scalability: Adequately managed up to 800 concurrent users
Resource Utilization:
CPU Utilization: 5 percent
Network In:
Bytes: 2,000 KB/s
Packets: 500 packets/s
Network Out:
Bytes: 3,000 KB/s
Packets: 800 packets/s

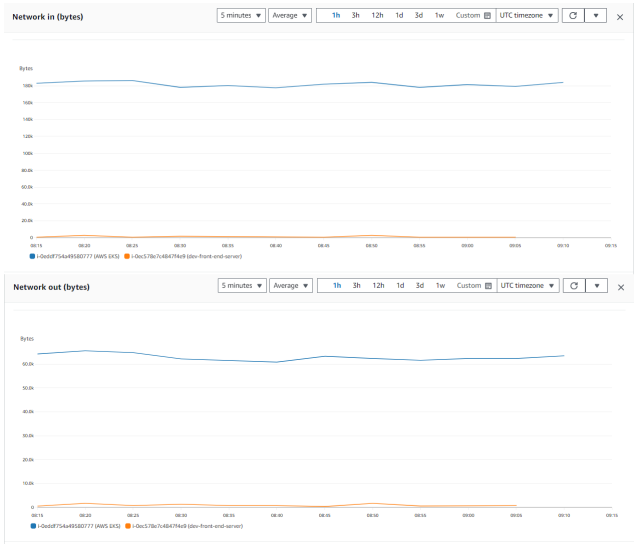


Fig. 4. Network in/out (bytes) AWS EC2/EKS

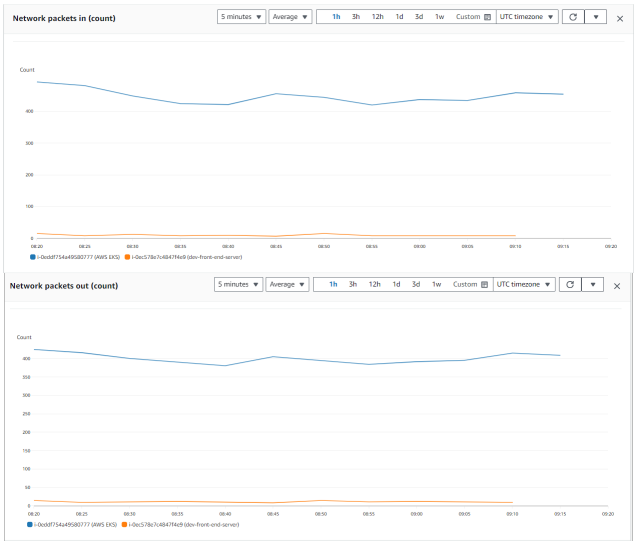


Fig. 5. Network in/out (Packets) AWS EC2/EKS

7) Statistical Analysis

Statistical methods, such as t-tests or ANOVA, were applied to analyze the significance of differences between the performance metrics obtained from the EKS and EC2 deployments.

This methodology aimed to provide a comprehensive understanding of the performance implications associated with AWS EKS and EC2 deployments for React applications, offering insights into the optimal deployment strategy for enhanced web metrics.

III. Conclusion

This research explores the optimization of React application performance through a comparative analysis of AWS Elastic Kubernetes Service (EKS) and EC2 deployments, prioritizing enhanced web metrics. The study reveals that AWS EKS exhibits notable advantages, showcasing lower response times, superior scalability, and efficient resource utilization compared to conventional EC2 instances. These findings underscore the relevance of adopting modern container orchestration solutions, especially EKS, for organizations seeking to deliver responsive and scalable React applications. The research also contributes to the broader discourse on cloud-based deployment strategies, emphasizing the significance of considering framework-specific nuances. Overall, the insights derived from this study provide actionable recommendations for decision-makers navigating the evolving landscape of technology, informing optimal choices for React application development in the cloud.

IV. Future Research

For future research, opportunities are abound for further exploration and refinement in the realm of optimizing React application performance. Delving deeper into the dynamic scalability of React applications within various AWS deployment scenarios could provide nuanced insights, and expanding the study to encompass additional cloud providers and container orchestration platforms would contribute to a more comprehensive understanding.Exploring the integration of serverless computing models, evaluating emerging technologies like AWS Fargate, and assessing the environmental sustainability of deployment strategies align with the evolving landscape of cloud computing. Ongoing research should adapt to technological advancements, considering updates to AWS services, the React library, and container orchestration technologies to ensure findings remain relevant and applicable in the face of evolving industry standards. The continuous evolution of these technologies presents a fertile ground for future studies, aiming to provide practical guidance for developers and organizations navigating the intricacies of cloud-based React application deployment and optimization.

References

[1] D. J. Milroy, C. Misale, G. Georgakoudis, T. Elengikal, A. Sarkar, M. Drocco, T. Patki, J.-S. Yeom, C. E. A. Gutierrez, D. H. Ahn, and Y. Park, "One step closer to converged computing: Achieving scalability with cloud-native hpc," in *2022 IEEE/ACM 4th International Workshop on Containers and New Orchestration Paradigms for Isolated Environments in HPC (CANOPIE-HPC)*, pp. 57–70, 2022.

[2] A. S. Asrese, P. Sarolahti, M. Boye, and J. Ott, "Wepr: A tool for automated web performance measurement," in *2016 IEEE Globecom Workshops (GC Wkshps)*, pp. 1–6, 2016.

[3] A. Alalawi, A. Mohsin, and A. Jassim, "A survey for aws cloud development tools and services," in *3rd Smart Cities Symposium (SCS 2020)*, vol. 2020, pp. 17–23, 2020.

[4] N. Kewate, "A review on aws - cloud computing technology," *International Journal for Research in Applied Science and Engineering Technology*, vol. 10, pp. 258–263, 01 2022.

[5] F. Tang, B. Chen, J. Chen, Z. Zhang, and L. Han, "Design of course experiments based on k8s container cloud platform," in *2022 International Conference on Education, Network and Information Technology (ICENIT)*, pp. 53–58, 2022.

Learn to Unlearn: Targeted Unlearning in ML

Muhammad Shaheryar
School of Computer Science and Engineering
Kyunpook National University
Daegu, South Korea
shaheryar@knu.ac.kr

Junhyeok Jang
School of Computer Science and Engineering
Kyunpook National University
Daegu, South Korea
wkdwnsgur@knu.ac.kr

Lamyanba Laishram
School of Computer Science and Engineering
Kyunpook National University
Daegu, South Korea
yanbalaishram@knu.ac.kr

Jong Take Lee
School of Computer Science and Engineering
Kyunpook National University
Daegu, South Korea
jongtaeklee@knu.ac.kr

Soon Ki Jung
School of Computer Science and Engineering
Kyunpook National University
Daegu, South Korea
skjung@knu.ac.kr

***Abstract*— In the area of machine learning, unlearning plays a vital role in strengthening security and privacy for ML-based systems. This work proposes a new technique meant to handle the difficulty of selectively unlearning knowledge linked with a forget-set. By purposely injecting noise into this data set, our technique breaks its original patterns and perturbs predictions. This noise-based technique enhances the unlearning process inside pre-trained models, resulting in higher loss during retraining with the changed forget-set, leading to parameter modifications and the unlearning of previously learned patterns. Our strategy is geared to strengthen a model's resilience to altering the data dynamics by explicitly decreasing the influence of the forget-set while preserving overall model performance. Importantly, our technique avoids the requirement for long-term storage of training data, providing a computationally efficient and effective solution for machine unlearning with applications in data privacy and information removal. The study provides thorough findings, showing how successful this method is in the context of simple and noisy forget-set on datasets like MNIST and Cifar-10 via loss comparisons. Additionally, feature map analyses carried out at different model levels, demonstrate how adaptable and dynamic the model is in reaction to data disruptions. This method not only outperforms current unlearning approaches in terms of forget-set accuracy but also does away with the need for training data to be stored for an extended period of time. This makes it an effective option for machine unlearning, with potential applications in data privacy and information removal.**

Keywords—Machine unlearning, noise, forget set, data forgetting, data privacy

I. INTRODUCTION

The rapid use of machine learning models has resulted in a fundamental shift in the data-driven decision-making environment. These models, which are frequently trained on massive datasets, have led to incredible discoveries across a variety of sectors. However, as these models become more and more integrated with private, sensitive, or copyrighted data, concerns about data privacy and the individual's capacity to have their data erased have gained substantial attention. This shift necessitates the use of machine unlearning, which permits the selective removal of specific data samples or classes while preserving model performance on the remaining data.

The ramifications of machine unlearning are extensive and cover numerous crucial disciplines. First, it acts as a cornerstone for the enforcement of privacy legislation and the preservation of user privacy. Legislation like the EU's GDPR [3] and California's CCPA [1] has introduced the idea of the "Right to Be Forgotten", which allows users to request the erasure of their personal data from corporate-trained models [4, 2]. By permitting selective data erasure, machine unlearning provides a practical way of protecting these legal rights and supporting the appropriate treatment of personal data. Second, the machine unlearning serves as a sentinel to strengthen resistance against adversarial assaults and increase model security. In the area of facial recognition technology, the difficulty of establishing a balance between identification accuracy and individual privacy looms large. Let's assume a hypothetical example where we wish to delete certain faces from a facial recognition system while maintaining the overall system's effectiveness. Utilizing machine unlearning, we have an effective technique to solve this difficulty. Isolating the particular faces or people that need to be deleted from the system is the first step in the procedure. These people have been carefully identified, and their face data has been designated for de-learning. Retraining the model without problematic faces or adding perturbations to reduce the model's recognition confidence for these people are examples of machine unlearning strategies. The end product is a facial recognition system that respects the privacy and choices of individuals who have chosen to opt out while maintaining accuracy of approved users. By ensuring that the forgotten faces no longer result in recognition, machine unlearning offers a practical means of removing certain users from the system. We are able to effectively traverse the complex junction of user permission, privacy, and face recognition in this way.

Machine unlearning has many possible uses. Machine unlearning may be used to solve strict data privacy rules, enabling businesses to abide by legislation like the GDPR [3] and CCPA [1] by selectively deleting people' data from their models upon request, protecting privacy, and upholding the “right to be forgotten”. Furthermore, it allows for the removal of outdated or useless data, increasing useful life and enhancing the accuracy of machine learning models and facilitating model adaption. Its applications cut across a range of industries, including healthcare, banking, education, content moderation, and crisis management, and it provides a flexible tool for juggling data privacy, model adaptation, and security in the rapidly changing field of AI and machine learning.

Research is still being done to address the problem of effective unlearning in deep networks, particularly for sophisticated models like CNNs and Vision Transformers.

Institute of Information & communications Technology Planning & Evaluation (IITP)

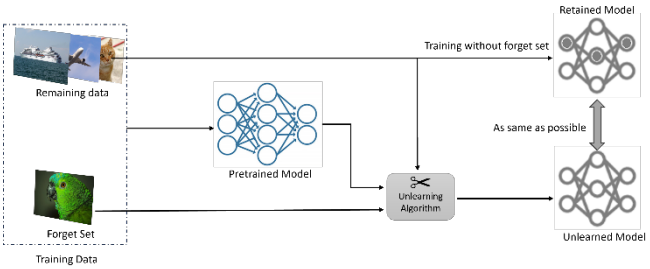


Fig. 1 Machine Unlearning – A General Overview

Because of their complexity, unlearning numerous classes in these models has yet to be widely researched. The nonconvex character of CNNs' loss space is a major barrier that makes it more difficult for researchers to estimate how eliminating data samples will affect the optimization path and the final weight configurations. It is more challenging to accurately measure the degree of unlearning when many optimal weight values are available for the same network. In order to address these issues and improve the effectiveness and efficiency of unlearning in deep networks for real-world applications, researchers are now exploring novel approaches and strategies.

Unlearning in a model includes the careful removal of specific data while keeping knowledge of the rest. To achieve this, one method is to change the model by identifying patterns that are basically the reverse of what it originally learned during training. By adding these opposite patterns to the model, we can successfully represent the desired unlearning process while keeping the information related to the remaining data. During the initial training of the model, it tries to reduce the loss across all groups. Intuitively, when dealing with unlearning, we can focus on increasing the noise (or mistake) specifically for the samples we want to forget. This method pushes the model to find and learn patterns that are opposite to those found in the unlearning samples. These patterns can be thought of as "anti-samples" since they neutralize the knowledge the model originally obtained.

The paper is organized as follows: In Section 2, we delve into the pertinent literature. Section 3 outlines the methodology employed for data acquisition and experimentation. Section 4 presents the findings and subsequent discussion. Finally, Section 5 wraps up with the conclusion and outlines directions for future work.

II. Related Work

In the field of machine unlearning, several approaches and techniques have been developed to deal with the challenges of extracting data from deep learning models. The training dataset is divided into separate shards using the "sharded, isolated, sliced, and aggregated" (SISA) [5] technique, which trains sub-models on each shard independently. This method works best in small-scale scenarios and limits the influence of data samples on sub-models since model performance decreases as the quantity of unlearning data increases. To reduce the impact of data points during training, several methodologies use strategic data grouping [5, 6]. However, since these techniques require several network and gradient snapshots, they often have large storage costs. By offering an upper constraint on the amount of information still present in the network, in [8], authors expanded upon their findings for machine unlearning settings.

Institute of Information & communications Technology Planning & Evaluation (IITP)

Machine unlearning and differential privacy, a crucial idea in data privacy, are closely related. A probabilistic concept of unlearning was given by Ginart [9] with the goal of maintaining output distribution similarity across models trained with and without unlearning. The most advanced unlearning techniques presently available include retraining, while Chundawat in [7] used a student-teacher architecture to cause amnesia while maintaining model performance. Graves, Nagisetty, and Ganesh [10] introduced two unlearning strategies, including retraining and an amnesiac strategy that relabels data with false labels.

To aid unlearning, in [11] and in [12], authors investigated error-maximizing noise matrices, and challenges like zero-shot and deep regression unlearning were tackled but the [12] with this work concentrating on the more experienced field of classification problems. Many methods, including federated learning [13] and linearized networks [14], have been investigated to help with unlearning. Investigations on privacy hazards and formal proof difficulties associated with data erasure have also been conducted [15, 16]. Machine unlearning is a rapidly developing discipline, with recent publications covering its many facets and practical uses. By combining many desired qualities—fast, no storage, selective in the parameters to be adjusted, and independent of supplementary models our technique differs from previous research.

III. Methodology

A. Basic Definitions and Objective

We outline the unlearning issue in the context of deep learning. Let's assume we have a full training dataset D_c, which consists of n samples. Each sample (x_i, y_i) has an input x_i from input space $X \in R^d$ and a class label y_i from the set $\{Y = 1, 2, \ldots, K\}$. We distinguish between the samples to forget as forget-set (D_f) and the samples to retain as retain-set (D_r) from this complete dataset D_c, making sure that their intersection is empty ($D_f \cap D_r = \emptyset$) and that their union encompasses the whole dataset ($D_f \cup D_r = D_c$). Our main goal is to develop an approach that efficiently unlearns the impact of a predefined D_f, and substitutes a new set of model's weight θ_{new}. This new weight set is derived using the previous trained model's weight θ_{old}, and intended to behave in the parameter and output space in a manner that is similar to a model that has never been exposed to D_f before. By lowering the model's reliance on D_f information and improving its flexibility to new data dynamics, our methodology aims to accomplish these two goals simultaneously. This offers a strong and convincing solution for information removal in trained models, encourages privacy, and maintains model integrity.

B. In-Depth Method

In this section, we provide a detailed process for purposefully unlearning certain samples from a trained model by carefully using the data perturbation approach. Below is the primary algorithm. Our main goal is to systematically break the model's internal representations of the chosen forget-set without compromising the model's overall performance. This strategy greatly improves the model's capacity to adjust to changes and anomalies within these specific data points. Our process commences with the selection and initialization of a

Algorithm

Input: θ_{old}, model with pretrained weights, D_C, D_f , where $D_f \subseteq D_C$.

Parameter: $D_{f\_noisy} = \varepsilon(D_f)$, perturb the D_f.

N: Apply Gaussian noise N(0,σ^2) with σ {0.5,1,1.5,2}

Output: θ_{new}, model with updated weights.

1. For each noise level σ in the set N{0.5,1,1.5,2}:
   1.1 Create noisy version: $D^n{}_f$.
2. $\theta_{\text{update}} = retrain(\theta_{old}, D^n{}_f)$.
   2.1 Calculate the loss using the negative cross-entropy loss function.
   2.2 $\boldsymbol{L(\hat{y}, y) = -\sum_{i=1}^{c} y_i \, log(\hat{y}_i)}$
3. Performance(θ_{new}, D_f) << Performance(θ_{new}, $D_C \setminus D_f$)
4. Performance(θ_{new}, D_r) ~ Performance(θ_{old}, D_C)

*The notation << indicates a significant decrease in performance on the forget set compared to the rest of the dataset. And ~ indicates similar performance to gold standard model.

pre-trained model. This model, already supplied with extensive learned knowledge and parameters, acts as the basic framework upon which the unlearning process is enacted. It is crucial to ensure that the pre-trained model aligns with the problem domain under research. The fundamental part of our unlearning technique is the selection of the forget-set D_f which comprises specialized data samples that are allocated for unlearning. These samples have been purposefully identified, and their relevance to the unlearning objective must be properly reviewed and confirmed.

In our methodology, we introduce noise as a purposeful disruptor to forget set, targeted at facilitating the unlearning process within a pre-trained model. This comprises the inclusion of intentional perturbations in the form of noise, making forget-set less predictable and obfuscating its original patterns. Subsequently, during model retraining with this noisy forget set, the model experiences greater loss compared to the initial clean data due to the perturbations. In reaction, the model undergoes parameter adjustments to minimize the increased loss, thereby unlearning the precise patterns it had previously associated with the forget-set.

More specifically, the model is systematically retrained using the noisy forget-set. The primary purpose is to exaggerate the loss associated with disturbed samples from the forget-set. This strategy drives the model to reset its internal representations, fostering unlearning as it responds to this purposefully disrupted data. This technique strengthens the model's resilience to fluctuations and deviations within the perturbed forget set, coinciding with the unlearning purpose. Noise-Induced retraining in our approach is actually where a model is subjected to varying levels of Gaussian noise during the training process. This technique involves adding Gaussian noise with standard deviations ranging from 0.5 to 2 to a dataset (Dforget) and then retraining the model (M) for a short duration (1 to 2 epochs) on this noisy data, resulting in multiple models (Mσ) corresponding to different noise levels. The loss is calculated using negative cross-entropy between the model's predictions and actual targets, and gradients are computed to update the model's weights with an optimizer. Importantly, after each noise-induced training session, the model's weights are reset to their initial state, ensuring that subsequent training starts from a clean slate. Importantly, this approach does not delete all the knowledge the model has received from the previous set. It selectively lowers the influence of the forget-set patterns, allowing the model to keep its performance on other data while being less sensitive to the specific features of the forget-set. An iterative training technique can be used if the desired level of unlearning is not reached quickly enough. To maximize the unlearning result, the procedure can be repeated with careful modification of the perturbation modalities and hyperparameters, depending on how much unlearning is wanted and how it affects the overall performance of the model.

The final result of the designed process is a model with updated weights θ_{new} that only shows reduced confidence or accuracy for samples in the forget-set D_f.This decrease in performance is a direct result of the perturbations that are applied, which cause the model's current understanding of these data to be disrupted. Significantly, this new knowledge improves the model's resistance to changes and anomalies within the specifically targeted data points.

IV. Experimenal Setup

In this work, we evaluated the performance of our suggested methodology on two benchmark image classification datasets, MNIST [18] and CIFAR-10 [17], each with ten different classes, using the ResNet-18 model [19]. We separated each dataset into three important subgroups in order to evaluate the unlearning process systematically. Train, forget, and validation set. With samples carefully selected for their relevance to the unlearning goal, the second subset, known as the forget-set, was created especially for unlearning trials. The validation set contained data that the model had not come across during the original training or the unlearning process, and it was used to benchmark the performance. After training, we applied our methods to the ResNet-18 model, adding noise and controlled perturbations to the forget-set in an attempt to unlearn certain knowledge. A number of measures have been established in the literature [20, 21] to assess an unlearning method's overall effectiveness. These metrics aim to quantify the amount of unlearn or forget data that is still present in the network.

We then clearly saw variations in the model performance by comparing the accuracy, loss, and other performance metrics of the ResNet-18 model before and after unlearning using the validation Set. In table 1. we assess our strategy in the context of two key model variants: the original model, initially trained on the whole dataset D_c, and the retrain model, trained

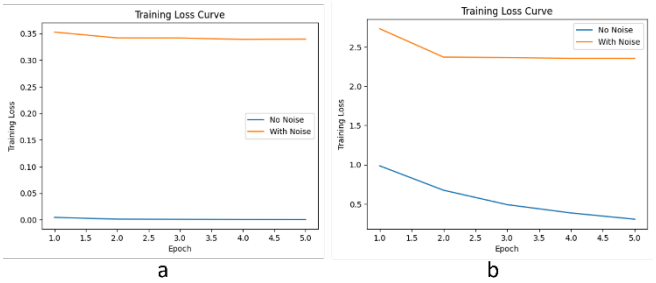


Fig. 2 Loss comparison b/w simple and noisy forget-set on (a) MNIST and (b) Cifar-10 dataset

Institute of Information & communications Technology Planning & Evaluation (IITP)

on the retain-set D_r. Our objective is to show the success of our technique using a standardized experimental design employing a modest dataset and model architecture. All models, whether trained from scratch or fine-tuned from pre-trained versions, underwent a training schedule lasting 30 epochs for the former and 2 and 5 epochs for the fine-tuning. We applied the usual cross-entropy loss criteria and Adam optimizer with a maximum learning rate of 0.01. To measure the efficacy of our unlearning technique, we depend on two essential evaluation metrics: $AccD_r$, reflecting the accuracy on the retain set, and $AccD_f$, expressing the accuracy of the forget-set. Notably, the $AccD_f$ should ideally approach zero, indicating effective model unlearning, but the $AccD_r$ should remain near to the original model's performance. Our experimental results attest to the effectiveness of our strategy in boosting model adaptability within a consistent and controlled experimental setting.

V. RESULTS AND FINDINGS

Fig. 2 gives a comprehensive illustration of the contrast between loss values in the context of simple and noisy forget-set, as applied to both the MNIST and Cifar-10 datasets. In particular, we can notice that the loss associated with the already trained model on the forget-set, given as D_f, displays low change. However, when we incorporate noisy samples into the model, a considerable increase in loss is seen. This underlines the issue of tolerating and adjusting to noisy data in the model. Fig. 2 illustrates the process of changing the model's weights to take this increased loss into consideration. Furthermore, we examine the feature maps at each layer of the model before and after introducing noise in order to ascertain where the anticipated forgetting happens and how the model is learning the sample. It appears that different feature maps are generated, and the model will attempt to learn different features, which will affect the forget-set weights. Fig. 3 gives a striking glimpse into the feature maps learned by the ResNet model's layer 3 when applied to the MNIST dataset. It is obvious that the feature maps display different differences before and after the injection of noise, emphasizing the model's flexibility to handle data

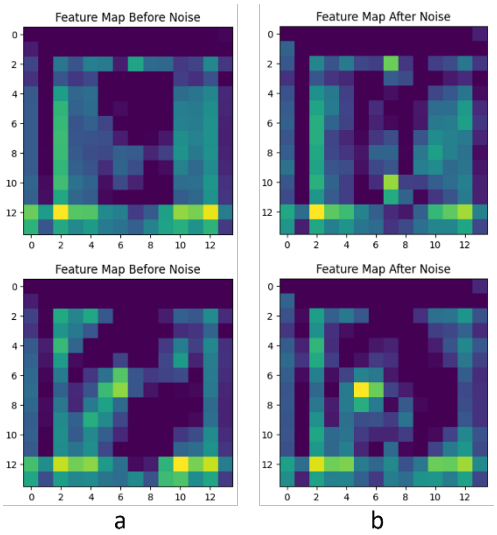


Fig. 3 Feature map comparison of layer 3 of Resnet18 model. a. Sample without noise b. Sample with the noise

Institute of Information & communications Technology Planning & Evaluation (IITP)

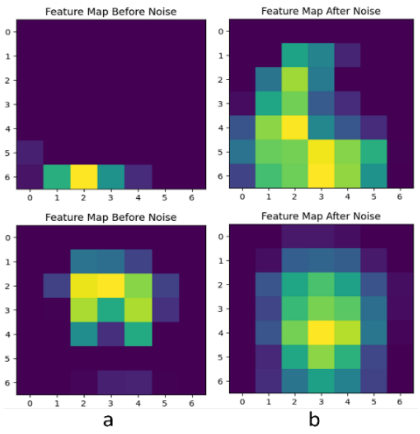


Fig. 4 Feature map comparison of layer 4 of Resnet18 model. a. Sample without noise b. Sample with noise

perturbations. Similarly, Figure 4 looks into the ResNet model's layer 4 feature maps, both before and after the addition of noise, indicating large discrepancies in the patterns the model records. This discovery emphasizes the model's potential to continually evolve and optimize its weight updates inside the architectural framework, emphasizing the dynamic nature of its learning process in response to noisy input. These tendencies are constant and extend across the MNIST and Cifar-10 datasets, as demonstrated in Fig. 2(a) and 2(b), respectively, emphasizing the robustness and flexibility of the model in the face of noisy data. Pretrain Model and Retrain from Scratch are two baseline unlearning techniques that are used to compare our outcomes. We contrast the forget-set unlearning outcomes with the techniques mentioned above. Table 1. shows the accuracy of retain and forget-sets after two epochs and with different Nosie levels, which correspond to unlearning. We outperform the current techniques in forget-set D_f accuracy; similar to ResNet18 and CIFAR-10, we maintain 65.03% of D_r accuracy from an initial 69.81% while severely deteriorating D_f performance (19.37% from an initial 68.37%). We only trained our model on CIFAR-10 for 30 epochs. In Table 2, we show results from retraining a model on CIFAR-10 dataset with combined noise levels without resetting to the initial state; the accuracy $AccD_r$ does not decrease significantly, but there's a loss in accuracy $AccD_f$, indicating these are the best settings for machine unlearning for CIFAR-10, highlighted in bold.

Table 1: Unlearning on MNIST and CIFAR-10

Datasets	***Metrics***	***Pretrained Model***	***Retrain from Scratch***	***Ours***
MNIST	Acc_{D_r}	**99.48%**	99.08%	98.46%.
	Acc_{D_f}	99.06%	0.0%	**12.17%**
CIFAR-10	Acc_{D_r}	**75.81%**	68.25%	64.03%.
	Acc_{D_f}	55.73%	0.0%	**10.72%**

VI. CONCLUDING REMARKS

This research describes a fresh and promising technique for intentional unlearning in trained models. It is important to recognize specific constraints and prospective directions for future investigations. First off, the suggested method's sensitivity to the selection of perturbation modalities and hyperparameters is one of its drawbacks. These variables are

critical to the effectiveness of unlearning, and a great deal of trial and error may be necessary to determine the ideal parameters. Future research might concentrate on creating

Table 2. Accuracy effect with different noise levels

Datasets	Noise with Diff σ	Epoch	Acc_{D_r}	Acc_{D_f}
CIFAR-10	0.5,1.0,1.5	1	**64.45%**	**10.72%**
	0.5,1.0,1.5	2	54.42%	1.86%
	0.5,1.0,1.5,2.0	1	57.22%	2.15%
	0.5,1.0,1.5,2.0	2	48.52%	1.02%

automated techniques for perturbation strategies and hyperparameter selection in order to reduce the approach's reliance on human tweaking and increase accessibility. Further research is also necessary to understand how unlearning affects model performance, particularly in situations where the model must preserve its knowledge while unlearning certain information. More thorough research and fine-tuning might be done on this duality of training and unlearning aims.

In addition, the study mainly deals with unlearning within the framework of a single pre-trained model. Subsequent investigations may examine the suitability of this methodology in a more extensive setting, including group education or cross-task learning. An attractive topic for future study is to examine the scalability and generalizability of the proposed technique to other model architectures and domains. The ethical ramifications of unlearning are a further topic that needs investigation, especially in applications requiring delicate or contentious data. This involves privacy, responsibility, and any unexpected effects that might occur from lowering a model's knowledge in a chosen manner. In summary, while the study makes a significant contribution to the area of model unlearning, there are many opportunities for further research to improve the method's ethical concerns, application, and robustness.

ACKNOWLEDGMENT

This work was supported by Institute of Information & communications Technology Planning & Evaluation (IITP) grant funded by the Korea government, (MSIT) (No. 2019-0-00203, Development of 5G-based Predictive Visual Security Technology for Preemptive Threat Response) and by the MSIT Korea, under the Innovative Human Resource Development for Local Intellectualization support program (IITP-2022-RS-2022-00156389) supervised by the IITP (Institute for Information & communications Technology Planning & Evaluation) and also by the Digital Innovation Hub project supervised by the Daegu Digital Innovation Promotion Agency(DIP) grant funded by the Korea government(MSIT and Daegu Metropolitan City) in 2023(No. DBSD1-04).

REFERENCES

[1] S. o. C. D. o. J. Office of the Attorney General, “California consumer privacy act (ccpa),” https://oag.ca.gov/privacy/ccpa, 2023.

[2] H. Chang and R. Shokri, “On the privacy risks of algorithmic fairness,” in 2021 IEEE European Symposium on Security and Privacy (EuroS&P). IEEE, 2021, pp. 292–303.

[3] P. Voigt and A. Von dem Bussche, The EU General Data Protection Regulation (GDPR). A Practical Guide. Springer International Publishing, 2017

[4] Wang, Guan, Charlie Xiaoqian Dang, and Ziye Zhou. "Measure contribution of participants in federated learning." 2019 IEEE international conference on big data (Big Data). IEEE, 2019.

[5] L. Bourtoule, V. Chandrasekaran, C. A. Choquette-Choo, H. Jia, A. Travers, B. Zhang, D. Lie, and N. Papernot, “Machine unlearning,” in 2021 IEEE Symposium on Security and Privacy (SP). IEEE, 2021, pp. 141–159.

[6] Y. Wu, E. Dobriban, and S. Davidson, “Deltagrad: Rapid retraining of machine learning models,” in International Conference on Machine Learning. PMLR, 2020, pp. 10 355–10 366

[7] V. S. Chundawat, A. K. Tarun, M. Mandal, and M. Kankanhalli, “Zeroshot machine unlearning,” IEEE Transactions on Information Forensics and Security, 2023.

[8] A. Golatkar, A. Achille, and S. Soatto, “Eternal sunshine of the spotless net: Selective forgetting in deep networks,” in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020, pp. 9304–9312.

[9] A. A. Ginart, M. Y. Guan, G. Valiant, and J. Zou, “Making AI forget you: Data deletion in machine learning,” in Advances in Neural Information Processing Systems, vol. 32, no. NeurIPS, 2019, pp. 1–14.

[10] L. Graves, V. Nagisetty, and V. Ganesh, “Amnesiac machine learning,” in Proceedings of the AAAI Conference on Artificial Intelligence, vol. 35, no. 13, 2021, pp. 11 516–11 524.

[11] Tarun, Ayush K., et al. "Fast yet effective machine unlearning." IEEE Transactions on Neural Networks and Learning Systems (2023).

[12] Chundawat, V. S.; Tarun, A. K.; Mandal, M.; and Kankanhalli, M. 2023b. Zero-shot machine unlearning. IEEE Transactions on Information Forensics and Security.

[13] C. Wu, S. Zhu, and P. Mitra, “Federated unlearning with knowledge distillation,” arXiv preprint arXiv:2201.09441, 2022. A. Golatkar, A. Achille, A. Ravichandran, M. Polito, and S. Soatto, “Mixed-privacy forgetting in deep networks,” in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2021, pp. 792–801.

[14] M. Chen, Z. Zhang, T. Wang, M. Backes, M. Humbert, and Y. Zhang,“When machine unlearning jeopardizes privacy,” in Proceedings of the2021 ACM SIGSAC Conference on Computer and Communications Security, 2021, pp. 896–911.

[15] N. G. Marchant, B. I. Rubinstein, and S. Alfeld, “Hard to forget: Poisoning attacks on certified machine unlearning,” in Proceedings of the AAAI Conference on Artificial Intelligence, vol. 36, no. 7, 2022, pp. 7691–7700.

[16] A. Thudi, H. Jia, I. Shumailov, and N. Papernot, “On the necessity of auditable algorithmic definitions for machine unlearning,” in 31st USENIX Security Symposium (USENIX Security 22), 2022, pp. 4007– 4022.

[17] Krizhevsky, Alex, and Geoff Hinton. "Convolutional deep belief networks on cifar-10." Unpublished manuscript 40.7 (2010): 1-9.

[18] Deng, L. (2012). The mnist database of handwritten digit images for machine learning research. IEEE Signal Processing Magazine, 29(6), 141–142.

[19] He, Kaiming, et al. "Deep residual learning for image recognition." Proceedings of the IEEE conference on computer vision and pattern recognition. 2016.

[20] A. Golatkar, A. Achille, and S. Soatto, “Eternal sunshine of the spotless net: Selective forgetting in deep networks,” in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020, pp. 9304–9312.

[21] A. Golatkar, A. Achille, and S. Soatto, “Forgetting outside the box: Scrubbing deep networks of information accessible from input-output observations,” in European Conference on Computer Vision, 2020, pp. 383–398.

Machine Learning Approaches for Effective Energy Forecasting and Management: Case Study Building in Chiang Mai University, Thailand

1st Natdanai Keivwath
College of Arts Media and Technology
Chiang Mai University
Ching Mai, Thailand
natdanai_ki@cmu.ac.th

2nd Pattaraporn Khuwuthyakorn
College of Arts Media and Technology
Chiang Mai University
Ching Mai, Thailand
pattaraporn.khuwuth@cmu.ac.th

***Abstract*— With the past demand for energy having increased, peak electricity demand in Thailand had been linked with higher temperatures. Climate change had made it more probable for the world to experience these elevated temperatures. Effective forecasting and management were crucial for institutions to ensure sustainability and cost-efficiency. This study explored machine learning techniques to optimize energy consumption and forecasting in a case study building at Chiang Mai University, Thailand. The researcher formulated cost function for electricity consumption in buildings. After establishing the relationship between variables, we selected an algorithm that was suitable for the electricity consumption dataset in Thailand, specifically the Simple Regression Model. The dataset consisted of energy usage data for 4 university buildings over a month and meteorological data as well as PM10 and PM2.5 data for the past year. The results showed that the increase in temperature influenced the rise in electricity consumption. The relationship between temperature and electricity consumption could be described with a linear equation. An energy consumption cost function could be created from the energy used to adjust the temperature. There was an approach to building a feedback model for energy consumption control. Energy monitoring systems like smart meters, IoT sensors, or building management systems were implemented to collect real-time data on energy usage.**

Keywords*— Machine Learning, linear regression, electricity consumption (key words***)***

I. INTRODUCTION

Electricity is a vital energy resource used for Thailand's development, and there's been a trend of increasing electricity consumption every year. Forecasting the overall country's electricity demand is essential and has been extensively studied for planning and managing electricity use. If we examine electricity usage by sector, different patterns emerge among the industrial, commercial, and residential sectors. Thus, different usage patterns had different impacting factors. Forecasting electricity consumption helped in efficient management and planning for maximum energy savings. Thailand's peak electricity demand occurred during the hot season, specifically from March to May every year. Moreover, global warming had led to a rising average temperature in Thailand, impacting the cost of electricity due to increased consumption. Given the rising temperature trends from global warming that affected increased electricity consumption, the electricity costs had also risen. To address this issue, machine learning techniques were employed to identify factors impacting electricity consumption and develop a recommendation system for efficient electricity usage. The researcher had formulated an electricity consumption cost function influenced by external factors like temperature, relative humidity, and PM2.5 levels. It was hypothesized that higher temperatures led to increased electricity consumption. This research studied electricity consumption data, meteorological data, and air quality information of a building in Chiang Mai University. The objective was to understand the factors that affected the building's electricity consumption, create a cost function equation related to these factors, and provide guidelines for efficient electricity usage in the future. Data collection had utilized Internet of Things (IoT) technology, with the first set of sensors gathering external data on temperature, humidity, PM2.5, and PM10. The second set of sensors had recorded hourly electricity consumption data. The data analysis was conducted using a linear regression model. After the data was prepared, the results were processed, summarized, and directions for future research were outlined. Future studies might have needed to consider other influential factors and develop a smarter recommendation system for optimizing electricity consumption efficiency.

II. RELATE OF WORK

A. Sustainable Facilities Management

Scholars had defined the management of physical resources as the coordination with all related parties for planning, system design, as well as equipment and various amenities to enhance the competitiveness of the organization. Facility management(FM) was related to management concerning Process, People, and Place. [1] Initially, building management was thought of as managing the construction from the construction phase to post-construction maintenance. Recently,

technology had been introduced to assist construction work, such as Building Information Modeling (BIM). BIM technology had been developed for use in full-circle construction, from building design to construction. BIM technology used computer systems as the main component in creating accurate virtual models of buildings, whether for design, drafting, structural calculations, cost estimates, procurement, or various building system planning. [2] Moreover, FM also had other main responsibilities, including maintenance, energy conservation, environmental management, safety management, space changes, and adapting buildings to usage.[3] FM was heading towards sustainable development, considering the application of technology and conducting business with innovations that resulted in positive impacts on society, economy, and environment.[4] FM tasks in the realm of smart buildings could achieve SDG 9 by applying related technologies like IoT and Artificial Intelligence.[5].

B. Energy Saving

Energy consumption in the world had been trending upwards. Thailand itself was a country that experienced a rising trend in energy use as well. The direction of research studies on energy conservation in the field of FM was related to methods that would help reduce costs associated with energy use and promote sustainable resource utilization. This focused on analyzing energy use, evaluating energy-saving technologies, managing energy usage systems, applying alternative energy sources, and studying user behavior, to name a few.[6].

From the Thailand energy situation report of 2022, electricity consumption accounted for 21.1% of the country's total energy usage. The leading source of energy was refined oil, making up 48.0%. The rate of electricity usage had increased from the previous year (2021) by 3.7%. The overall electricity consumption in 2022 had been 197,209 million units, which was an increase of 3.8% from the previous year. It was predicted that electricity consumption in 2023 would increase by 3.1%.[7] When the country's overall electricity consumption trended towards an increase, the need for electricity savings and maximizing efficient electricity use became increasingly important. In the realm of FM, the focus was on optimizing energy use in buildings. There was an analysis of electricity consumption and forecasting of usage to utilize the data for building management planning. Over the past decade, energy usage in buildings had continuously increased due to a rising population, more time spent in buildings, improved indoor environmental quality, and global climatic changes. People spent 90% of their time indoors.[8] Studies had been conducted on how to use energy efficiently, focusing on Energy efficient lighting, a system found to be highly efficient in energy conservation. It was found that cost reduction promoted energy efficiency practices. Such costs pertained to equipment and the energy used. Besides lighting systems, the system that consumed significant energy was the ventilation system.[9] Furthermore, the efficiency of electricity usage had been improved by integrating artificial intelligence technology to help control the turning on and off of light switches. The system could detect human movement, aiming to improve the energy efficiency of indoor lighting. As a result, energy consumption and electricity costs had decreased by approximately 79%.[10] Buildings in universities in Spain used electricity primarily from air conditioning systems, which had the most significant impact on the electricity bills they had to pay. [11] In line with buildings in universities in the United States, the system that incurred the highest cost was the Heating Ventilation and Air Conditioning (HVAC), which was a major climate control system in buildings. Facilities management in educational buildings is different from other types of buildings. The different requirements are temperature, humidity, ventilation and indoor air quality.[12] Studies had been conducted on energy use in conjunction with air quality and comfort levels. The average energy consumption of educational buildings was found to be 142 kWh/m2/y, with the heating system using an average of 97 kWh/m2/y. From this electricity consumption data, efforts had been made to reduce electricity use, taking into account the indoor air quality (IAQ) standards by controlling CO2 levels and considering indoor thermal comfort. It was discovered that adjusting the temperature by 1 degree Celsius led to a 12% increase in electrical energy consumption.[13]

Energy conservation work had been one part of FM, and in other countries, there had been studies on lighting systems that successfully conserved electrical energy. However, the air conditioning system was the system that consumed the most energy in a building. The use in buildings also involved the comfort of the occupants, which related to the matter of temperature. In the past, FM had incorporated new technologies to help save energy, such as IoT and machine learning, which is a branch of artificial intelligence.

C. Thermal Comfort

If one wanted to study Thermal Comfort in Thailand, one had to understand the country's climate. Subsequently, they had to delve into studies on Thermal Comfort that had been conducted for over 30 years and were widely accepted. Research on Thermal Comfort in air-conditioned rooms in Thailand was undertaken to identify factors influencing occupant comfort and its relation to energy conservation.

Thailand's climate is warm and humid, with particularly high temperatures during the summer. As a result, many buildings have been equipped with air conditioning systems to achieve thermal comfort by reducing the indoor temperature. The American Society of Heating, Refrigerating, and Air-Conditioning Engineers (ASHRAE) defined thermal comfort as the condition of mind that expressed satisfaction with the thermal environment, assessed by subjective evaluation.[14] The comfort zones of Thai people had relative humidity in the range of 50% - 70% and the temperature was in the range of 24-27 °C.[15] It could be seen that there were factors of relative humidity and temperature related to Thermal Comfort. A study on field experiments on thermal comfort in campus classrooms in Taiwan had found that the factors influencing thermal comfort were temperature, air movement, and temperature from average radiation. However, humidity was not statistically significant. [16] The studied of environment and classrooms, which were air-conditioned classrooms in Thailand, and sought to find whether fans and air conditioners could be used together for energy savings while still maintaining Thermal Comfort, it was found that room users

Identify applicable funding agency here. If none, delete this text box.

felt slightly cool at a temperature range of 20-24 °C and wind speeds of 0.2-0.5 m/s. Room users felt cool at a temperature range of 23-25 °C and wind speeds of 0.5-2.9 m/s. Room users felt neutral at a temperature range of 24.5-26.6 °C and wind speeds less than 0.5 m/s. It could be seen that the comfortable temperature range coincided with the air temperature in air-conditioned rooms, and fans could be used to extend the higher temperature range while still feeling comfortable. Therefore, the temperature of the air conditioner could be adjusted up to 28 °C, and a fan could assist with wind speeds of 0.5-2 m/s, resulting in additional energy savings.[17] In addition to the aforementioned research, there was other research that determined the thermal response time and temperature range of thermal comfort for Thai students in air-conditioned classrooms. The temperature of the air conditioners was adjusted to 25, 26, 27, and 28 °C. The results showed that Thai students could tolerate temperatures in the range of 25±0.5 – 28±0.5 °C, with wind speeds of approximately 0.5-0.9 m/s. Therefore, the temperature in the classrooms could be adjusted to between 26-28 °C, which was higher than the government-recommended temperature of 25 °C. Energy savings could be achieved by adjusting the air conditioner from 25 °C to 26-28 °C, with an average electricity savings of 7.30% per 1 °C increase.[18]

The feeling of comfort for building occupants is related to the amount of electricity used. Adjusting the temperature affects the occupants' comfort level. Achieving the right temperature depends on the operation of the air conditioning system, which consumes a significant amount of electrical energy. The operation of the air conditioner not only impacts the temperature but also affects the humidity in the air.

D. Temperature Effect

The global warming condition has caused the average temperature of the world to rise, affecting every country with increased temperatures. Forecasting the total demand for electric current is necessary for energy management, which results from several factors. Predicting the amount of electricity usage in buildings is also a matter of energy management in building administration. If one can identify the factors that affect the electricity consumption of buildings, it would be beneficial for efficient building management and energy usage.

The peak electricity demand in Thailand occurred on May 7, 2023, with a demand of 34,130.50 megawatts, due to high temperatures and a long public holiday.[19] Global warming was another reason that caused the average temperature of Thailand to rise. The Intergovernmental Panel on Climate Change (IPCC) had estimated that global warming would result in a 1-4 degree Celsius increase in global surface temperatures by the end of the century.[20] The increase in temperature had an impact on the daily demand for electricity, seasonal demand, and long-term electricity demand in Thailand. The forecasted rise in temperature was determined using the UK Hadley Centre climate model, along with data from four representative socioeconomic scenarios from the Intergovernmental Panel on Climate Change Special Report. The results from the forecast indicated that the temperature in Thailand would increase by 1.74 – 3.43 °C in 2080. It was estimated that the peak electricity demand would be at 1.5%-3.1% in 2020, 3.7%-8.3% in 2050, and 6.6%-15.3% in 2080.[21] From the Greece data set, it was found that the temperature and electricity demand had a high degree of correlation. [22] From the Bangladesh data set, it was explained that temperature could effectively predict electricity demand using linear regression.[23] From the two articles mentioned above, it could be concluded that linear regression could be used to predict electricity demand based on temperature. The short-term electricity demand prediction in Thailand was influenced by certain determining variables and weather conditions. The studies focused on the impact arising from temperature, comparing the accuracy between Feed Forward Artificial Neural Network (FF-ANN) and regression methods. What was found was that regression methods could predict more accurately than FF-ANN for the dataset of Thailand. The impact of temperature on Thailand's electricity demand had a linear relationship.[24]

E. Internet Of Thing

In the past, IoT was a technology that had become more capable and had been applied to a variety of applications. IoT emerged due to the advancement of internet technology and the interconnection of various devices into a network. There were technologies that facilitated the connection between devices and cloud systems, controlling and transmitting data between devices via the internet. [25] IoT sensors were devices used in IoT systems for detecting data from various environments and transmitting data through the internet network to the IoT platform for data storage and processing. These sensors could detect data such as temperature, humidity, atmospheric pressure, light intensity, speed, motion, and more.[26] The advancement of technology and the increased capabilities of IoT had enabled its application in various fields, whether it be smart transport systems, intelligent healthcare systems, smart buildings, smart energy systems, or smart environment monitoring.[27] The application of IoT in weather monitoring had been explored, with weather stations needing to collect data on temperature, humidity, atmospheric pressure, rainfall, and light intensity. The components of the model for weather detection included a board using NodeMCU, which internally had a microcontroller ESP8266 that could connect to WiFi. As for the IoT sensors, the sensors could connect to NodeMCU, including a temperature and humidity sensor (DHT11), a pressure sensor (BMP180), a rain module, and a light-dependent resistor (LDR). When the detected values were abnormal, it was possible to report the results via SMS or email.[28] The application of IoT in detecting PM2.5 and PM10 values was explored by using the SDS011 laser sensor to measure PM10 and PM2.5. After collecting the data, the obtained information was used to predict PM10 values using the Random Forest algorithm.[29]

III. Data Preparation

The data used in the research came from 2 sources, namely, the Climate Change Data Center of Chiang Mai University (CMUCCDC) and the Energy Research and Development Institute-Nakornping, Chiang Mai University (ERDI). The data was provided through the DustBoy API and the ERDI API, respectively.

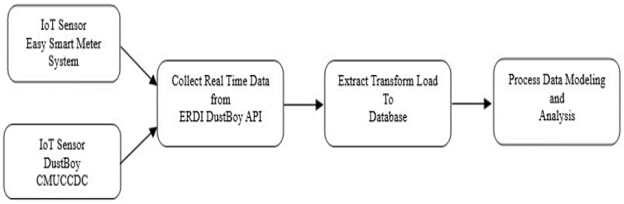


Fig. 1. Overview of data collection to data processing

Figure 1 showed the flow of data, starting from the IoT sensors detecting information. The data was collected using the ERDI API and DustBoy API services. The subsequent step involved data cleaning before the ETL (Extract, Transform, Load) process placed the data into a database and formatted it ready for processing through the model. Finally, the results from the model analysis were obtained.

In the first part, the raw data had come from the detection by IoT sensors. Specifically, the IoT sensors of the Easy Smart Meter System had been responsible for detecting data on electrical current usage by utilizing digital electric meters and Current Transformers (CT). The IoT sensors of DustBoy had been tasked with detecting data on the quantity of Pm2.5, Pm10, temperature, and relative humidity using the Dust Boy Model N WiFi, which had high accuracy. The subsequent step had been the gathering of data from services through the ERDI API and DustBoy API. The electricity usage data from the ERDI API had been used to store information on the quantity of electrical usage (in kilowatts) and the meter numbers 4 times per hour, along with the building names and their codes. The data in this section, which had been used for research, encompassed a total period of one month. The quantity of fine dust particles PM2.5 and PM10, as well as temperature and relative humidity data from the DustBoy API, had been collected every hour. The data had been used for research covering a retrospective period of 1 year. The next step had been performing Data Cleaning and the Extract, Transform, Load to Database (ETL) process. In this stage, erroneous data resulting from sensor malfunctions and the instability of the internet system, which had affected the data, such as missing values, outlier data, or other error data, had been eliminated. The ETL process had taken into consideration the format of the destination data that would be used for processing with a machine learning model. Each model had used different data formats. The ETL process had begun with extracting data, selectively choosing only the data we needed. For instance, for the electricity usage data, we had selected the electricity usage quantity (in kilowatts) but had not chosen the electricity meter number, among others. Transforming data had involved changing the data format to be suitable for processing, such as converting raw data that measured electricity usage 4 times in 1 hour and summing it up to be in an hourly data format. After that, the data had been loaded into the database to align with the format that would be used for processing. The final step had been processing with a model, which had been chosen for its suitability for the purpose of the task and its compatibility with the data set.

IV. Data Modeling

From the data collected from the initial research sources, it can be analyzed that the trend of electricity usage in the afternoon tends to increase, as shown in Figure 2. This is consistent with the average temperature data, which also rises during the afternoon period.

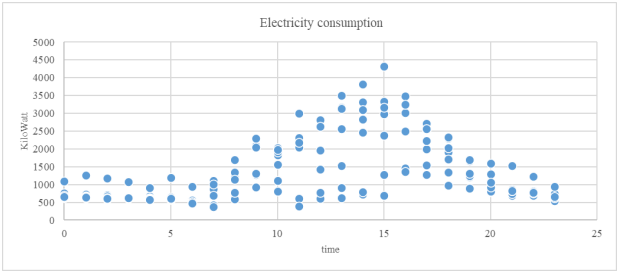


Fig. 2. The value of electricity consumption is shown according to the time period

The forecasting of electricity demand normally involved using several factors for prediction. The nature of the data for these factors was quantitative, in the form of continuous and discrete data, as well as qualitative data, depending on the suitability for the processing model used. The estimation of daily electricity demand in Thailand had involved comparing four models: regression, artificial neural network, support vector machine, and hybrid models. It was found that both the regression and hybrid models yielded the best results equally.[30] The forecasting of short-term electricity demand in Thailand, which used temperature data as the main factor, was tested for the Thailand data set. It was found that the impact of temperature on Thailand's electricity demand had a linear relationship and was more accurate than FF-ANN.[24] Therefore, in this study, the researchers had used a simple regression model to process the data when that factor had a high correlation coefficient.

In the research conducted this time, it was hypothesized that temperature had an influence on the quantity of electricity usage, in a manner where the variations were directly proportional. If the temperature increased, the quantity of electricity usage would also increase. Additionally, a cost function equation was defined for the quantity of energy consumption in a 1-hour period, taking into account factors affecting the quantity of electrical energy usage, including the amount of PM2.5 dust, humidity, and temperature. The electrical energy usage depended on adjusting the aforementioned values to suit the thermal comfort of the users.

Cost function electric consumption

$$Y_{all} = Y_1 + Y_2 + Y_3 \quad (1)$$

Declaring variables in equations

Y_{all} = Total amount of electric energy

Y_1 = Electric energy is used to adjust the Pm2.5 value

Y_2 = Electric energy is used to adjust the humidity

$Y3$ = Electric energy is used to adjust the temperature

Function of electric energy use to adjusting pm2.5, humidity and temperature

$$Y_1 = \beta_1 \Delta P + C_1 \quad (2)$$

$$Y_2 = \beta_2 \Delta H + C_2 \quad (3)$$

$$Y_3 = \beta_3 \Delta T + C_3 \quad (4)$$

Declaring variables in equations

β_1 = Weighted for adjusting PM2.5

C_1 = Constant value for adjusting PM2.5

β_2 = Weighted for adjusting humidity

C_2 = Constant value for adjusting humidity

β_3 = Weighted for adjusting temperature

C_3 = Constant value for adjusting temperature

ΔP = Constant value for adjusting temperature

ΔH = Constant value for adjusting temperature

ΔT = Constant value for adjusting temperature

Correlation of the first data set included the variables PM 2.5, PM 10, and humidity, while the second data set contained the variables humidity, temperature, and amount of electricity consumption (kilowatt). When the correlation coefficient was strong, a simple linear regression model was used to process the model.

V. Experiment

The result of finding correlation is the correlation coefficient between the parameter as follows: PM2.5 and PM10 equal to 0.972, PM2.5 and humidity equal to 0.649, PM10 and humidity equal to 0.672. That's for weather data after 1 year. The correlation coefficient between the variables in the data set of electricity consumption in the past 1 month has the following results: electricity consumption(kilowatt) and humidity equal to -0.678, electricity consumption(kilowatt) and temperature equal to 0.709, humidity and temperature equal to -0.967. The result of the correlation coefficient at the Near Prefect is 0.90 – 0.99, very strong is 0.70 - 0.89 and Substantial –Very strong is 0.50 -0.69. Therefore, the following features are processed by a regression model.

Regression statistics of humidity and PM2.5 are values as follows: Multiple R = 0.649, R Square = 0.422, Regression Coefficients = -0.907 and Intercept = 88.054. Regression statistics of humidity and temperature are values as follows: Multiple R = 0.967, R Square = 0.934, Regression Coefficients = -0.197 and Intercept = 43.419. Regression statistics of temperature and electricity consumption are values as follows: Multiple R = 0.709, R Square = 0.503, Regression Coefficients = 252.25 and Intercept = -5369.41. A linear regression line has an equation of the form Y = a + bX, where X is the explanatory variable and Y is the dependent variable. The slope of the line is b, and a is the intercept. Thus, by processing with linear regression, linear equations can be created as follows.

$$T = \alpha_1 H + b_1 \quad (5)$$

Humidity (H) is the explanatory variable and temperature (T) is the dependent variable, α_1 is weighted to convert humidity to temperature values, b_1 is constant

$$H = \alpha_2 T + b_2 \quad (6)$$

Temperature (T) is the explanatory variable and humidity (H) is the dependent variable, α_2 is weighted to convert temperature to humidity values, b_1 is constant

$$P = \alpha_3 H + b_3 \quad (7)$$

humidity (H) is the explanatory variable and PM2.5 (P) is the dependent variable, α_2 is weighted to convert temperature to humidity values, b_3 is constant

Substitute the obtained linear equation into the following: ΔP, ΔH and ΔT and derive equation. Set P_1 as the actual PM2.5 value and P_2 as the desired PM2.5 value. Set H_1 as the actual humidity value and H_2 as the desired humidity value. Set T_1 as the actual temperature value and T_2 is the temperature constant set according to recommendation 25 °c

$$\Delta P = P_1 - P_2 \, , P = \alpha_3 H + b_3$$
$$\Delta P = (\alpha_3 H_1 + b_3) - (\alpha_3 H_2 + b_3)$$
$$\Delta P = \alpha_3 (H_1 - H_2)$$

$$\Delta P = \alpha_3 \Delta H \quad (8)$$

$$\Delta H = H_1 - H_2 \, , H = \alpha_2 T + b_2$$
$$\Delta H = (\alpha_2 T_1 + b_2) - (\alpha_2 T_2 + b_2)$$
$$\Delta H = \alpha_2 (T_1 - T_2)$$

$$\Delta H = \alpha_2 \Delta T \quad (9)$$

$$\Delta T = T_1 - T_2 \, , T = \alpha_1 H + b_1$$
$$\Delta T = (\alpha_1 H_1 + b_1) - (\alpha_1 H_2 + b_1)$$
$$\Delta H = \alpha_1 (H_1 - H_2)$$

$$\Delta H = \alpha_1 \Delta T \quad (10)$$

Substitute temperature difference, humidity difference and PM2.5 difference with term from linear equation. Back to the cost function equation, which is the total energy consumption equation. Substitute temperature difference, humidity difference and the PM2.5 difference are determined by the temperature difference term in equations Y1, Y2, and Y3.

$$Y_{all} = Y_1 + Y_2 + Y_3 \quad (1)$$

$$Y1 = \beta_1 \Delta P + C_1 \, , \Delta P = \alpha_3 \Delta H \; and \; \Delta H = \alpha_2 \Delta T$$

$$Y_1 = \beta 1 \alpha_3 \alpha_2 \Delta T + C_1 \quad (11)$$

$$Y_2 = \beta_2 \Delta H + C_2 \, , \Delta H = \alpha_2 \Delta T$$

$$Y_2 = \beta_2 \alpha_2 \Delta T + C_2 \quad (12)$$

$$Y_3 = \beta_3 \Delta T + C_3 \quad (13)$$

Combine the equations.

$$Y_{all} = \beta_1 \alpha_3 \alpha_2 \Delta T + C_1 + \beta_2 \alpha_2 \Delta T + C_2 + \beta_3 \Delta T + C_3$$

$$Y_{all} = \Delta T(\beta_1 \alpha_3 \alpha_2 + \beta_2 \alpha_2 + \beta_3) + C_1 + C_2 + C_3$$

Identify ω and ν

ω is the total weight of energy consumption.

$$\omega = \beta_1 \alpha_3 \alpha_2 + \beta_2 \alpha_2 + \beta_3$$

ν is the constant of the sum of C_1 C_2 and C_3

$$\nu = C_1 + C_2 + C_3$$

$$Y_{all} = \Delta T \omega + \nu \quad (14)$$

ΔT is the difference of T_1 and T_2. while T_1 as the actual temperature value and T_2 is the temperature constant set according to recommendation 25 °c.

VI. Result and Discussion

The result of the correlation coefficient at the near prefect were correlation coefficient of humidity and temperature equal to -0.967, correlation coefficient of PM2.5 and PM10 equal to 0.972. This showed that the PM2.5 used instead of the PM10 and the temperature used instead of the humidity. In order to prove the hypothesis that if the temperature rises, electrical energy consumption will also increase. The highest level of correlation coefficient between the parameters and electrical consumption is temperature. Regression statistics of temperature and electricity consumption were values as follows: Multiple R = 0.709, R Square = 0.503, Regression Coefficients = 252.25 and Intercept = -5369.41. That showed that predicted electricity consumption from the actual temperature and if the temperature increases then electricity consumption will increase. In addition, create 3 linear equations, that is, first, the linear equation of temperature from humidity. Second is the linear equation of humidity from temperature and lastly is the linear equation of PM from humidity.

This research created a cost function equation of energy consumption from related features including electricity energy used adject temperature, electricity energy used adject humidity and electricity energy used adject PM2.5. Eventually, the equation can be reformulated that the total energy consumption per hour depended on the temperature difference plus a constant value.

Since climate change is generating an average annual increase in global temperatures, the research findings will be helpful in developing future energy management standards. The study's findings indicate that the electricity consumption will increase considerably with rising temperatures. In addition, there are unavoidable hidden elements like PM 2.5 and humidity. The primary factor is Thailand's tropical climate, which makes air conditioning commonplace. Air conditioners are capable of regulating temperature, lowering humidity, and assisting in the filtration of air that contains PM 2.5. Therefore, there must be a plan for adjusting the temperature of the building for the benefit of saving energy and there must be a plan for maintaining the air conditioner because it is a device that uses a large amount of electricity.

VII. Conclusion

The results of the research showed that temperature influenced the actual electricity consumption and was the factor that had the greatest influence on the electricity consumption in the building characteristics of university buildings. The dataset of electricity consumption at the university in Thailand could use the regression model to accurately predict electricity consumption, and the regression statistic between temperature and electricity consumption was linear and variable. In addition to create a cost function equation of energy consumption from related features including electricity energy used adject temperature, electricity energy used adject humidity and electricity energy used adject PM2.5. This is a model of electricity consumption in buildings at the time. In this model, there are three factors being used, as mentioned, by setting the values back to the temperature values. This basic model can be further developed to improve its efficiency by adjusting the weight of the water for each feature, particularly the temperature value. Alternatively, it can be further developed for a feedback model for energy consumption. Creating a feedback model for energy consumption involves setting up a system to monitor, analyze, and adjust energy usage based on real-time data. There is an approach to building a feedback model for energy consumption control. Implement energy monitoring systems like smart meters, IoT sensors, or building management systems to collect real-time data on energy usage. Create a loop by regularly comparing real-time energy consumption data with the established baseline or target. Analyze deviations by identifying any deviations from the target or baseline. Analyze the reasons behind these discrepancies. Determine the underlying causes of high or low energy consumption. Finally, implement adjustments based on the analysis, take appropriate actions to correct the deviations, such as: adjusting the temperature of the air conditioner, upgrading to energy-efficient appliances or equipment, modifying schedules for equipment operation and educating occupants about energy-saving practices etc.

Acknowledgment

The authors would like to thank raw data from research project of Developing a Dynamic Digital Twin for Sustainable Smart Campus Management which has received funding support from the NSRF via the Program Management Unit for Human Resources & Institutional Development, Research, and Innovation (Grant numberB16F640189), and Chiang Mai University Grant number R000029859.

References

[1] F. Becker and F. Steele, "The total workplace," Facilities, vol. 8, no. 3, pp. 9-14, 1990.

[2] R. Sacks, C. Eastman, G. Lee, P. Teicholz, and B. Handbook, "A Guide to Building Information Modeling for owners, designers, engineers, contractors, and facility managers," Sacks, R.(eds), p. 688, 2018.

[3] J. Bröchner, T. Haugen, and C. Lindkvist, "Shaping tomorrow's facilities management," Facilities, vol. 37, no. 7/8, pp. 366-380, 2019.

[4] A. Opoku and J. Y. Lee, "The future of facilities management: Managing facilities for sustainable development," vol. 14, ed: MDPI, 2022, p. 1705.

[5] IFMEC, "Facility Management Approach to Realizing the Sustainable Development Goals," ed: Roden The Netherlands, 2018.

[6] M. Marocco and I. Garofolo, "Integrating disruptive technologies with facilities management: A literature review and future research directions," Automation in Construction, vol. 131, p. 103917, 2021.

[7] G. Network. "The Energy Policy and Planning Office forecasts Thailand's energy consumption in 2023." https://www.greennetworkthailand.com/eppo-energy-update-2566/ (accessed 9 september, 2023)

[8] X. Cao, X. Dai, and J. Liu, "Building energy-consumption status worldwide and the state-of-the-art technologies for zero-energy buildings during the past decade," Energy and buildings, vol. 128, pp. 198-213, 2016.

[9] Y. A. Adewunmi, A. Alister, B. Phooko, and T. Nokukhanya, "Energy efficiency practices in facilities management in Johannesburg," Journal of Facilities Management, vol. 17, no. 4, pp. 331-343, 2019.

[10] Y. Tan, P. Chen, W. Shou, and A.-M. Sadick, "Digital Twin-driven approach to improving energy efficiency of indoor lighting based on computer vision and dynamic BIM," Energy & Buildings, Article vol. 270, 09/01/September 2022 2022, doi: 10.1016/j.enbuild.2022.112271.

[11] P. Bastida-Molina, J. Torres-Navarro, A. Honrubia-Escribano, I. Gallego-Giner, and E. Gómez-Lázaro, "A detailed analysis of electricity consumption at the University of Castilla-La Mancha (Spain)," Energy and Buildings, vol. 289, p. 113046, 2023.

[12] Y. Cao, X. Song, and T. Wang, "Development of an energy-aware intelligent facility management system for campus facilities," Procedia engineering, vol. 118, pp. 449-456, 2015.

[13] A. Merabtine, C. Maalouf, A. A. W. Hawila, N. Martaj, and G. Polidori, "Building energy audit, thermal comfort, and IAQ assessment of a school building: A case study," Building and Environment, vol. 145, pp. 62-76, 2018.

[14] A. Standard, "Thermal environmental conditions for human occupancy," ANSI/ASHRAE, 55, vol. 5, 1992.

[15] J. Taweekun, "Thermal comfort zone for Thai people," 2013.

[16] Hwang, Ruey-Lung, Tzu-Ping Lin, and Nai-Jung Kuo., "Field experiments on thermal comfort in campus classrooms in Taiwan," ENERGY AND, no. 38, 2006.

[17] S. Atthajariyakul and C. Lertsatittanakorn, "Small fan assisted air conditioner for thermal comfort and energy saving in Thailand," Energy Conversion and Management, vol. 49, no. 10, pp. 2499-2504, 2008.

[18] K. Hussaro, N. Puangmalee, and V. Boonyayothin, "The Thermal Comfort of Thai Students in University Buildings under Variable Indoor Conditions of Air Conditioned Space," Journal of Renewable Energy and Smart Grid Technology, vol. 14, no. 1, 2019.

[19] E. G. A. o. Thailand. "Thailand's highest electricity demand in 2023." https://www.egat.co.th/home/statistics-demand-latest/ (accessed 1 October 2023, 2023).

[20] S. Solomon, Climate change 2007-the physical science basis: Working group I contribution to the fourth assessment report of the IPCC. Cambridge university press, 2007.

[21] S. Parkpoom and G. P. Harrison, "Analyzing the impact of climate change on future electricity demand in Thailand," IEEE Transactions on Power Systems, vol. 23, no. 3, pp. 1441-1448, 2008.

[22] S. Mirasgedis, Y. Sarafidis, E. Georgopoulou, V. Kotroni, K. Lagouvardos, and D. Lalas, "Modeling framework for estimating impacts of climate change on electricity demand at regional level: case of Greece," Energy Conversion and Management, vol. 48, no. 5, pp. 1737-1750, 2007.

[23] M. Z. Sadi, S. R. Deeba, and R. H. Siddique, "Temperature sensitivity forecasting of electrical load," in 2010 4th International Power Engineering and Optimization Conference (PEOCO), 2010: IEEE, pp. 244-248.

[24] K. Chapagain, S. Kittipiyakul, and P. Kulthanavit, "Short-term electricity demand forecasting: Impact analysis of temperature for Thailand," Energies, vol. 13, no. 10, p. 2498, 2020.

[25] R. A. Mouha, "Internet of things (IoT)," Journal of Data Analysis and Information Processing, vol. 9, no. 2, pp. 77-101, 2021.

[26] R. Krishnamurthi, A. Kumar, D. Gopinathan, A. Nayyar, and B. Qureshi, "An overview of IoT sensor data processing, fusion, and analysis techniques," Sensors, vol. 20, no. 21, p. 6076, 2020.

[27] W. He, G. Yan, and L. Da Xu, "Developing vehicular data cloud services in the IoT environment," IEEE transactions on industrial informatics, vol. 10, no. 2, pp. 1587-1595, 2014.

[28] R. K. Kodali and S. Mandal, "IoT based weather station," in 2016 international conference on control, instrumentation, communication and computational technologies (ICCICCT), 2016: IEEE, pp. 680-683.

[29] J. Saini, M. Dutta, and G. Marques, "Internet of things based environment monitoring and PM 10 prediction for smart home," in 2020 International Conference on Innovation and Intelligence for Informatics, Computing and Technologies (3ICT), 2020: IEEE, pp. 1-5.

[30] W. Pannakkong, T. Harncharnchai, and J. Buddhakulsomsiri, "Forecasting daily electricity consumption in Thailand using regression, artificial neural network, support vector machine, and hybrid models," Energies, vol. 15, no. 9, p. 3105, 2022...

Thai Cyberbullying Classification for Specific Domain using Machine Learning Analysis

1st Supattanawaree Thipcharoen
Department of Data Science and Digital Innovation
Faculty of Innovation Technology and Creativity
The Far Eastern University
Chiang Mai, Thailand
supattanawaree@feu.edu

2nd Teerawich Wongsa
Department of Business Administration
Faculty of Business Administration and Liberal Arts
Rajamangala University of Technology Lanna - Nan
Nan, Thailand
teerawich@rmutl.ac.th

***Abstract*—** Currently, cyberbullying has become a global issue, and Thailand is no exception. Incidents range from minor annoyances to criminal activities. Bullying can manifest in various degrees, leading to divisions and harassment in people's lives. The aim of this study is to enhance the predictive capabilities of models for cyberbullying types and levels, as well as to raise awareness and prevent such incidents. This research focuses on specific data domains, particularly the political domain, collected from three social networks (Facebook, X (formerly Twitter) and YouTube). The algorithms employed for model building include Naïve Bayes and Support Vector Machine. In the testing process, a 10-fold cross-validation method is applied. The experimental results indicate that the models for predicting types of bullying perform exceptionally well. Both Naïve Bayes and SVM exhibit strong Bully Type performance (Accuracy of SVM = 96%, Naïve Bayes = 83%), as well as predicting the Bully Level (Accuracy of SVM = 98%, Naïve Bayes = 87%). This can lead to further in-depth analysis of online social network behaviors. The primary objective remains the prevention of future criminal incidents.

Keywords*— *Thai Cyberbullying Classification, Machine learning, Naïve Bayes, Support Vector Machine, Thai Text Mining

I. INTRODUCTION

The existence of social media platforms has significantly transformed the current communication landscape. Everything that used to occur in the offline world is now evident in the online world as well. Similarly, bullying [1], which has evolved into a form of cyberbullying, may have more far-reaching, extensive, and elusive consequences than ever before.

In Thailand, cyberbullying is a significant issue [2], primarily stemming from widespread internet access and various technologies. It often manifests on social networks, ranging from minor expressions of bullying to conflicts and harassment that can even lead to loss of life. Politics is another topic that frequently leads to intense and contentious expressions of opinions. For this reason, the research team has developed a method for detecting cyberbullying using text analysis in conjunction with machine learning techniques. In previous work [3], the research team employed the Naïve Bayes algorithm to classify cyberbullying characteristics into various forms. We find that Naïve Bayes is suitable for predicting relatively simple data. However, when it comes to predicting complex and multi-dimensional data, such as data related to bullying, including bully type or bully level, it may not perform as effectively. Although the model's performance is already at a good level (77.70 % for Accuracy of the Bully Type Model Performance), there is still room for improvement. Thus, the second phase of this study will focus on data management for training and exploring additional algorithms. For text classification tasks, there are many interesting and efficient algorithms, such as Support Vector Machine (SVM). It is a model known for its efficiency in classifying high-dimensional data. The use of a kernel function also helps in effectively classifying data with complex patterns. Additionally, the amount of data in the corpus is more appropriate, making it suitable for text classification tasks.

Therefore, in this article, we will present a new framework for classifying cyberbullying, including a transition from the Naïve Bayes algorithm [4] to the Support Vector Machine algorithm [5][6]. The test data used in this study consists of text posted on various social media platforms commonly used in Thailand, such as Facebook, X (formerly Twitter), and YouTube, for reference.

II. SCOPE AND FRAMWORK

This section will explain about the Natural Language Processing engine framework for Thai cyberbullying and the scope of the data source as follows.

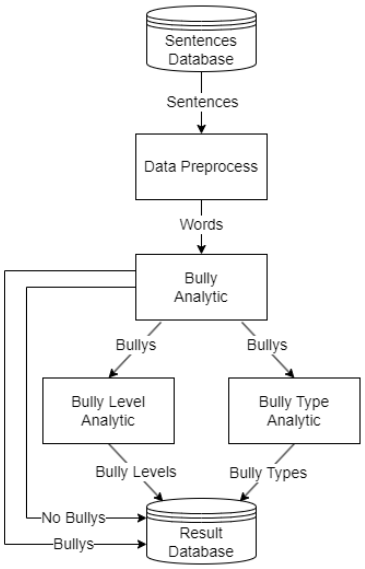


FIGURE I THAI CYBERBULLYING ANALYTIC FRAMEWORK

A. Thai Bully Prediction Model Framework

From figure I in the sentence database, which is the database used to store categorized text (bully/no bully) by experts, the system will retrieve these texts to undergo the data preprocessing process. This includes tasks like word tokenization, removing stop words, and so on. Once the text is prepared, it is analyzed to determine whether it is classified

as a "bully" or not. The results are then forwarded in two parts: The prediction results are stored in the result database for further display. Texts predicted as "Bully" are sent to the Bully Type Model and the Bully Level Models for further prediction. Subsequently, the prediction results from both models are stored in the Result Database for future use.

B. Data Source and Data Preparation

In the scope of this research, the political domain was used as an example for testing. Initially, the research team collected and tagged words through the Corpus Management System, which was developed to ensure the accuracy and completeness of the data used for model development. The data was gathered from social networks such as Facebook, YouTube, and Twitter, resulting in a total of 177,542 sentences. Non-relevant data was filtered out, leaving only information related to cyberbullying, which amounted to 67,670 sentences. The data was categorized into formats as below:

a) Bully/NoBully: Refers to bully sentence or not.

b) Bully Type: Refers to various categories or kinds of bullies. In this study, we divided the type of bully into 5 categories as described in table i.

c) Bully Levels: Refer to the severity of the bully's sentence as described in table ii.

TABLE I DEFINITION OF BULLY TYPE

Type of Bully	Definition
Gossip	Comments, Express Opinions
Harassment	Vulgar, Derogatory
Exclusion	Segregation, Separation
Hate speech	Incite, Provoke, Stir Up Hatred
Violence	Support the Use of Violence, Threaten with Fear, Harm

TABLE II DEFINITION OF BULLY LEVEL

Level of Bully	Definition
Level 1	Bullying unintentionally, without targeting any specific group, expressing opinions, raising issues related to race, age, gender, or sexual preferences.
Level 2	Identify individuals being talked about, causing annoyance by sending repeated messages, social exclusion, joining groups, insulting, disrespecting, and using vulgar language.
Level 3	Identify individuals being talked about, express hatred, have a psychological and emotional impact, affect feelings, impact on the body and family, provoke harm to others instantly, such as causing violence, threatening, harassing, inciting to harm, gathering forces for expulsion, deportation, publicly inciting murder, leading to severe criminal acts, and subject to legal prosecution

III. MODELS AND METHODS

In this section we will describe our method, tools, and the algorithm which was used in this research.

A. Previous Work

In our previous work, we focused on the topic of 'Cyberbullying Type Classification Across Thai Social Media Platforms using Machine Learning Analysis.' In the machine learning steps, Naïve Bayes was chosen to generate the model. The results showed the model performance of Bully and Bully Types as 0.84 and 0.77(F-Measure) respectively. The imbalance in class data had an effect on the model's performance. In this phase, we will focus on increasing the size of our training corpus, specifying domains, and applying SVM as a competitive against Naïve Bay.

B. Data Validation

For Data Preprocessing, we collected sentences from three different social media platforms. These sentences were then tagged as a corpus using Corpus Management software developed by our research team. The findings from our previous work indicated that data quality corresponds with model performance. Therefore, we not only tagged the data but also subjected it to validation by reviewers. Additionally, in order to process this step, we used PythaiNLP [7] for data preparation, which includes functions necessary for natural language processing (NLP) [8][9], such as word segmentation, symbol removal, spell correction, and eliminating word redundancy. We then organized these sentences into a format suitable for the model learning step, as shown in figure II.

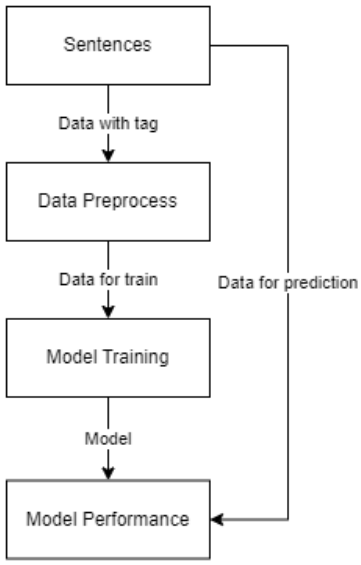


FIGURE II MODEL CREATION PROCESS

From figure II, texts tagged as "Bully" or "No Bully" by experts undergoing the Data Preprocessing process, which includes: Word Normalization: Handling words with repeated characters, such as converting "¬ʻ¢¢¢¢¢¢¢¢" to "¬ʻ¢" and "∫ǫ@@@@@@@@" to "∫ǫ@"Word Segmentation: Splitting sentences into individual words. Remove Stop Words: Dealing with common words that carry little meaning or are irrelevant in the context of the sentence. Following this, models are built, and the model's performance is measured sequentially. This is similar to the process of developing the Bully Type and Bully Level models.

C. Models Algorithms

In this experiment, we employed two well-known algorithms for text classification, Support Vector Machine (SVM), as a comparison to Naïve Bayes. We generated models using the Political Domain Corpus. The process of model learning is as follows.

We use NLTK [9] as a tool for model creation and model evaluation. We created 3 models, the Bully Model, the Bully Type Model and the Bully Level Model for classified Bully/NoBully sentences, Bully Type and Level sentences respectively.

1) Naïve Bayes

We still used Naïve Bayes Classification which is one of the supervised learning methods to train our model. This algorithm will learn from data with class labels to classify problem instances as shown in formula (1).

For each k possible outcomes or class C_k, $p(C_k|X)$ is the probability of the class being problem instances with property x, $p(x|C_k)$ is the probability of property x being in class instances, $p(C_k)$ is the problem instances class, and p(x) is the probability of property x.

$$p(C_k|x) = \frac{p(C_k)*p(x\,|C_k)}{p(x)} \quad (1)$$

2) Support Vector Machine

Given a training dataset of n points, (x_1, y_1), …, (x_n, y_n) where y_i is the class which relates to x_i

Maximum-margin hyperplane and margins for an SVM trained with samples from two classes. Samples on the margin are called the support vectors, as formula (2) – (4)

$$w^T x - b = 0 \quad (2)$$

Where *w* is normal vector to the hyperplane, for prediction we define as formula (3)

$$w^T x - b \geq 1 \; if \; y_i = 1 \quad (3)$$

$$\text{or } w^T x - b \leq -1 \; if \; y_i = -1 \quad (4)$$

SVM not only reduces its time and cost but also highly accurate for classification[6].

D. Performance matrix

To evaluate the models, we used 10 - fold cross validation [10][11][12]. By slicing data into 10 subgroups. In each round, one of the 10 will be selected as a test set and the remaining k-1 will be used as a train set. Then repeat this process until all 10 are selected as a test set.

The confusion matrix is used to be a model of performance measurement as shown in formula 5-8. Where tp (true positive) is a test result that correctly indicates the presence of a condition, tn (true negative) is a test result that correctly indicates the absence of a condition, fp (false positive) is a test result which wrongly indicates that a particular condition, and fn (false negative) is a test result which wrongly indicates that a particular condition.

$$precision = \frac{tp}{tp+fp} \quad (5)$$

Precision represents how well a model can predict the correct answer out of all data samples.

$$recall = \frac{tp}{tp+fn} \quad (6)$$

Recall represents how well a model can extract data for classifying from all data samples.

$$f - measure = 2 \times \frac{precision \times recall}{precision + recall} \quad (7)$$

The F – measure, also known as F1 score, represents how well a model has a correct answer selection ratio to all selected outcomes.

$$Accuracy = \frac{tp+tn}{tp+tn+fp+fn} \quad (8)$$

Accuracy represents how accurate a model can predict results from all samples.

IV. Results and Discussion

The creation of predictive models using Machine Learning algorithms involves testing and experimentation to measure the efficiency and accuracy of these models. We have chosen to use Naïve Bayes and Support Vector Machine for performance evaluation, and we present the results with metrics, including Accuracy (%), Recall (%), Precision (%), and F-measure (%). We have increased the number of sentences in previous work on the corpus for each type, as shown in table iii, while the models' performance are shown in table iv and table v.

TABLE III NUMBER OF DATA CORPUS

Corpus type	# of sentences
Bully	5,000
No-bully	5,000
Exclusion	5,000
Gossip	5,000
Harassment	5,000
Hate speech	5,000
Violence	5,000
Level 1	5,000
Level 2	5,000
Level 3	5,000

TABLE IV BULLY MODEL PERFORMANCE

Corpus Type	Naïve Bayes				SVM			
	P	R	F	A	P	R	F	A
Bully	0.92	0.92	0.92	0.91	0.99	1.00	1.00	0.99
No-bully	0.85	0.84	0.87		0.99	0.99	0.99	

**P: Precision, R: Recall, F: F-Measure, A: Accuracy*

From results in table iv, it is found that domain specific and increasing the amount of data significantly improves the model's performance. As shown in table iv, both Naïve Bayes and SVM yield relatively high results, with an Accuracy score of 90% and an F-measure score for the 'Bully' class that is high for both algorithms, at 0.92 and 1.00, respectively. However, for the 'NoBully' class, it is observed that SVM can predict with greater accuracy, scoring above 0.12.

TABLE V BULLY TYPE & BULLY LEVELS MODEL PERFORMANCE

Corpus Type	Naïve Bay				SVM			
	P	R	F	A	P	R	F	A
Exclusion	0.94	0.59	0.73	0.83	0.96	0.99	0.98	0.96
Gossip	0.83	1.00	0.91		0.96	0.94	0.95	
Harassment	0.79	0.76	0.77		0.96	0.95	0.95	
Hate speech	0.57	0.24	0.33		0.98	0.99	0.98	
Violence	1.00	0.29	0.44		0.96	0.94	0.95	
Level 1	0.83	0.86	0.84	0.87	0.97	0.97	0.97	0.98
Level 2	0.82	0.97	0.89		0.97	0.99	0.98	
Level 3	1.00	0.80	0.89		1.00	0.99	0.99	

**P: Precision, R: Recall, F: F-Measure, A: Accuracy*

In table v, The Bully Type Model performance clearly indicates that SVM can significantly outperform Naïve Bayes in every class, while the 'Hate speech' and 'Violence' classes

exhibit lower performance with similar results like 0.30 and 0.40.

In the field of Cyberbullying, it has been found that, apart from the types, the severity of Cyberbullying can potentially lead to criminal activities or serious issues that necessitate monitoring. Therefore, this research attempts to classify and create predictive models for Cyberbullying severity levels (1, 2 and 3). When considering table v, the SVM algorithm still maintains higher performance than Naïve Bayes, as seen in the case of the Bully Type Model,' which achieves a performance score of 0.98. The results from table IV and V show findings as follows.

a) Models should be generated from specific domain corpus. In the field of text classification, data training is considered a critical factor. Selecting specific, topic-focused data can reflect similarities or close relationships among different types, resulting in models that can make more accurate predictions. This work provides an example in the context of 'Bully' related to politics, which is prevalent on Thai social media platforms. However, bullying occurs in other domains as well, such as entertainment, education, and more. Therefore, to be suitable for the data, it is necessary to create models in various domains, along with the need for regular model updates.

b) Thai NLP still challenged. It is well-known that Thai NLP tasks are inherently complex and constantly evolving due to the use of slang and the ever-changing landscape of social platforms. This includes the use of the term 'Bullying.' As a result, it is necessary to continually update the models.

c) Balanced classes is an important process for learning models. From previous work, it was observed that class imbalance affects the model learning process. Therefore, in this research, data was balanced by increasing the number of samples through random sampling to achieve an equal class distribution before model training. The results indicate that when the class distribution is more balanced, the model is less prone to bias, leading to more accurate predictions.

d) The seperation of the Bully Model from the Bully Type and the Bully Level Model. The reason we separate the Bully Model from the Bully Type Model and the Bully Level Model is because if no-bully data is trained together with bully-type data (or bully-level data), there will be fewer no-bully data than bully data. For example, in case we look at bully-type data as a whole bully data. If no-bully data is trained together with bully-type data, where each type has the same number of 5,000, it will make no-bully data had 5,000 (16.67%) and bully data had 25,000 (83.33%). On the other hand, if no-bully data were increased to the same amount as bully-type data, it would become clear that there would be more no-bullies than bully-types for each type.

e) The loss between model. From the framework presented in figure i and the management of the imbalance data in issue c) and d), although it is possible to manage the imbalance data between classes, there are still limitations. Some data may be lost when the bully-model incorrectly predicts that it is a no-bully (when in fact it is a bully) or vice versa. In the worst case, the Bully Type Model and the Bully Level Model may receive an incorrect answer as input data from the Bully Model.

f) Lager data corpus. Currently, the development of a Bully Analysis Application for real world use has been successfully completed, but we still plan to create pre-trained data to facilitate recursion, thereby expanding the size of the corpus. This will enable the utilization of AI algorithms capable of working with larger corpora (like more than 1M) of data to train when there is a larger amount of data (n), using SVM may not be effective enough. Therefore, it is inevitable to look for other algorithms that are more suitable for training on large amounts of data, such as Deep Learning etc.

V. Conclusion and Future works

The objective of this research is to enhance the model's performance in classifying Thai cyberbullying within specific topics. This aims to raise awareness and prevent cybercrime related to cyberbullying. We have continued to use data from widely-used social media platforms in Thailand, such as Facebook, X (formerly Twitter), and YouTube, to create a dataset, as in previous work. We observed that using domain-specific datasets resulted in improved model performance. In this research, we are testing a model developed through the learning of data specific to the political domain. Therefore, the predictions will be limited to the political domain only. The Thai language often develops new words, phrases, or sentences related to bullying in a specific domain. These may have different meanings and uses. For example, the word 'ÃÆË (s̄lim). In the context of politics, it may refer to a group of people who support or have certain political ideologies and is often associated with criticism or exclusion. If used in the context of food, it refers to a specific sweet dish with the same name. Therefore, the model needs to be updated continuously to ensure its accuracy, leading to the effective application of the predicted data for monitoring and discovering individuals at risk of committing crimes in each domain.

In the future, we plan to develop additional models across various domains to further enhance predictive performance. Subsequently, we will apply social network analysis to detect and prevent cyberbullying-related crimes in Thailand.

Acknowledgment

We would like to express our gratitude to all the individuals who contributed to this research. This study represents the second phase of the project, which involved a collaboration with Minerva Consultant Co., Ltd. under the initiative for the innovative development of Safe and Creative Media using Artificial Intelligence. Additionally, it received support from the Thai Media Fund in 2022.

References

[1] Smith, P. K., J. Mahdavi, M. Carvalho, S. Fisher, N. Russell, and N. Tippett (2008), “Cyberbullying: its nature and impact in secondary school pupils”, Journal of Child Psychology & Psychiatry, 49, pp. 376-385.

[2] A. Euajarusphan, (2021). Cyberbullying and Thai Generation Z Youths in Bangkok, Thailand. International Journal of Crime, Law and Social Issues, 8(2), 43-55.

[3] S. Thipcharoen, T.Wongsa W. Bumrungsri, P. Tooptompong, and R. Sukhahuta.(2022). Cyberbullying Type Classification Across Thai Social Media Platforms using Machine Learning Analysis., 5thInternational Conference on Culture Technology (ICCT2022). November 4-7 ChiangMai.

[4] A. Akhter, Uzzal K. Acharjee, Md Masbaul A. Polash. (2019). Cyber Bullying Detection and Classification using Multinomial Naïve Bayes and Fuzzy Logic . I.J. Mathematical Sciences and Computing, 2019, 4, 1-12.

[5] N. Jenkarn and M. Ketcham (2020). Thai-textual Cyberbullying Detection using Support Vector Machines.Science Technology and Innovation(STIJ). Vol. 1 No. 1 January–February 2020. pp24-34.

[6] Palanivinayagam, C. Ziad El-Bayeh, and R.Damaševic˘ius.(2023). Twenty Years of Machine-Learning-Based Text Classification: A Systematic Review *.Algorithms* 2023, *16*, 236.

[7] W. Phatthiyaphaibun, K. Chaovavanich, C. Polpanumas, A. Suriyawongkul, L. Lowphansirikul and P. Chormai, "PyThaiNLP: Thai Natural Language Processing in Python," Zenodo, 2016.

[8] Ratchakrit Arreerard, Stephen Mander, Scott Pia. (2020). Survey on Thai NLP Language Resources and Tools. Proceedings of the 13th Conference on Language Resources and Evaluation (LREC 2022), pages 6495–6505Marseille, 20-25 June 2022

[9] S. Bird, E. Klein and E. Loper, Natural language processing with Python: analyzing text with the natural language toolkit, O'Reilly Media, Inc., 2009.

[10] Y. Sasaki, "The truth of the F-measure," Teach tutor mater, vol. 1, no. 5, pp. 1-5, 2007.

[11] O. Caelen, "A Bayesian interpretation of the confusion matrix," Annals of Mathematics and Artificial Intelligence, vol. 81, no. 3, pp. 429--450, 2017.

[12] Vanwinckelen, Gitte (2 October 2019). On Estimating Model Accuracy with Repeated Cross-Validation. lirias.kuleuven. pp. 39–44. ISBN9789461970442

Anomaly Detection in Video through Semi-Supervised Learning with Graph Convolutional Networks

Harsha B M
Centre for Digital Innovation,
CHRIST (Deemed to be University), Bengaluru, India

Dr. Guydeuk Yeon*
Centre for Digital Innovation,
CHRIST (Deemed to be University), Bengaluru, India
*Corresponding author: yeon@christuniversity.in

***Abstract*—This paper proposes an innovative approach to video anomaly detection leveraging a semi-supervised learning paradigm. The methodology integrates Graph Convolutional Networks (GCN) to extract human skeleton structures from video frames, enabling the capture of spatial relationships among key joints. Subsequently, these structures are utilized in a Deep Neural Network (DNN) to derive positive and negative features crucial for robust anomaly detection. To enhance the learning process, positive bags are curated from normal video segments, and Normalizing Flows are employed to generate pseudo-positive videos, augmenting the model's comprehension of diverse normal patterns. The final anomaly detection stage incorporates Extreme Value Theory (EDL) for accurate classification of anomalous segments. The amalgamation of GCN, DNN, Normalizing Flows, and EDL establishes a resilient semi-supervised framework, showcasing notable accuracy in identifying anomalies within video data. This research contributes to the advancement of video surveillance and anomaly detection methodologies, with potential applications across domains such as security, healthcare, and industrial monitoring.**

Keywords— Semi supervised learning, GCN, DNN, Anomaly Detection, Video surveillance

I. INTRODUCTION (*HEADING 1*)

In today's digital landscape, the ubiquity of video data has presented both opportunities and challenges across numerous sectors. One of the critical challenges is the reliable detection of anomalies within these vast and complex video datasets [Ren, Jing, et al., "Deep video anomaly detection: Opportunities and challenges," 2021 international conference on data mining workshops (ICDMW). IEEE, 2021]. Anomalies represent deviations from expected or usual patterns and can range from subtle irregularities to potentially significant events. In surveillance applications such as attacks, robbery, and crowd monitoring, the ability to accurately identify anomalies amidst the abundance of normal activities is crucial [Zhu, Yuansheng, Wentao Bao, and Qi Yu, "Towards Open Set Video Anomaly Detection," European Conference on Computer Vision. Cham: Springer Nature Switzerland, 2022]. Traditional approaches often struggle to discern these anomalies effectively due to the intricacies and complexities inherent in video data.

To tackle this challenge, our research introduces an approach that leverages the synergy of reliable methods. Specifically, the integration of Graph Convolutional Networks (GCN) [Zhang, Si, et al., "Graph convolutional networks: a comprehensive review," Computational Social Networks 6.1 (2019): 1-23] and Deep Neural Networks (DNN) as it forms the backbone of our methodology. These methods collaboratively work to extract and interpret the underlying human skeletal structures within video frames. The significance of focusing on these structures lies in their ability to encapsulate vital spatial relationships among key joints. By utilizing GCN, our framework adeptly captures and analyzes these relationships, enabling a deeper understanding of the intricate movements and interactions within video sequences.

The human skeletal structure analysis is pivotal as it enables the framework to discern anomalies by detecting deviations from established spatial patterns. By comprehensively understanding the spatial relationships among key joints, the framework becomes adept at recognizing abnormalities or irregularities in movements that deviate from normal behavioral patterns. This capability forms the foundation for accurate anomaly detection within video data.

The strategic utilization of Normalizing Flows plays a pivotal role in distinguishing Normal videos from Anomaly videos. These flows facilitate the creation of positive bags derived from segments of normal video behavior. Synthesizing pseudo-positive videos from these bags allows the framework to undergo an augmentation process, enriching its understanding of diverse normal behavioral patterns. This augmentation significantly enhances the model's ability to differentiate between normal behaviors and anomalies, thereby strengthening the overall anomaly detection capability. Once again, the DNN model plays a vital role in segregating Anomaly videos from Normal ones by training it with pseudo-positive videos for segregation.

The amalgamation of GCN, DNN, and Normalizing Flows within our methodology yields highly promising results. Empirical testing demonstrates notable accuracy in identifying anomalies within complex video datasets. This success positions our methodology as a transformative solution with diverse applications across sectors such as public security and industrial

monitoring. The ability to accurately identify anomalies amid normal activities presents a significant advancement in video surveillance and anomaly detection methodologies.

II. Literature Review

A. Selecting a Template (Heading 2)

Video anomaly detection has emerged as a critical area of research owing to its applications in various domains such as surveillance, healthcare monitoring, and industrial safety. Over the years, researchers have explored diverse methodologies and technologies to improve the accuracy and efficiency of anomaly detection within video data. Until now, there hasn't been a thorough study focusing on specialized deep structures for spotting rare abnormal features in videos. Our study doesn't cover other related areas.

Traditional approaches to anomaly detection often relied on handcrafted features and supervised learning algorithms. For instance, the studies by Sabokrou, Mohammad, et al. in 2018 and Hasan, Mahmudul, et al. in 2016 applied motion-based features and background subtraction methods to detect anomalies within crowded scenes, employing predefined features and labeled training data. However, while these methods initially provided valuable insights, their application faced limitations in addressing the complexities of dynamic environments. In our case, obtaining ground truth images of skeletons has enhanced our approach, allowing for more comprehensive and adaptable anomaly detection.

In recent years, the advent of deep learning techniques has revolutionized video anomaly detection. Unsupervised methods by Zaheer, M. Zaigham, et al. in 2022; Wang, Jun, et al. in 2019; and Li, Hao, Alin Achim, and D. Bull in 2012, as well as semi-supervised learning paradigms by Sarker, Mohammad Ibrahim, et al. in 2021; Khaire, Pushpajit, and Praveen Kumar in 2022; and Kiran, B. Ravi, Dilip Mathew Thomas, and Ranjith Parakkal in 2018, have gained traction for their ability to adapt to diverse and evolving scenarios. The utilization of Convolutional Neural Networks (CNNs) for feature extraction and spatiotemporal modeling has been a cornerstone in many studies. For instance, Ravanbakhsh, Mehdi, et al. in 2017 introduced an unsupervised framework utilizing CNNs for anomaly detection by learning video representations in an unsupervised manner.

The incorporation of graph-based models in video anomaly detection has also garnered attention. Graph Convolutional Networks (GCNs) have been utilized to capture spatial dependencies and relationships within video frames. Notably, the work by Yan, Zheng, et al. in 2020 demonstrated the effectiveness of GCNs in modeling human interactions and detecting abnormal behaviors in crowded scenes.

Semi-supervised learning has emerged as a promising avenue to address the challenges of limited labeled data in anomaly detection tasks. The utilization of pseudo-labeling, as explored in the work by Zhang, Si, et al. in 2019, showcases the potential of leveraging unlabeled data to augment the learning process, enhancing model generalization and performance.

However, despite advancements in individual methodologies, the integration of multiple techniques for robust video anomaly detection remains an ongoing challenge. Few studies have explored the integration of graph-based models, semi-supervised learning, and theoretical frameworks to bolster anomaly detection accuracy within video data.

Our research seeks to improve the work by Zhu, Yuansheng, Wentao Bao, and Qi Yu in 2022 by proposing an innovative approach that integrates Graph Convolutional Networks (GCN) with Deep Neural Networks (DNN), utilizing semi-supervised learning principles. We leverage Normalizing Flows to enrich the model's understanding of normal behavioral patterns and incorporate Extreme Value Theory (EDL) for precise anomaly classification. This integrative approach aims to elevate the accuracy and robustness of video anomaly detection, addressing the limitations observed in existing methodologies.

III. Methodology

Our methodology commences by processing the UCF Crime and XD Violence datasets. Allocating 40% of each dataset for the extraction of skeletal sequences from video frames. The subsequent representation of these video sequences involves the precise capture of 2D or 3D coordinates, systematically delineating the nuanced movements across successive frames. Within this process, 40% of the UCF Crime dataset and 40% of the XD Violence dataset are specifically utilized to generate the skeletal sequences, ensuring a comprehensive yet focused sampling approach. These extracted sequences are then translated into node sets, effectively encapsulating individual joints and culminating in feature vectors that incorporate joint coordinates alongside confidence estimations. The development of the undirected spatial-temporal graph is crucial for modeling temporal relationships. From here, the formulation and execution of the Spatial-Temporal Graph Convolutional Neural Network (ST-GCN) intricately follow, leveraging the comprehensive skeletal representations extracted from the designated portions of the UCF Crime and XD Violence datasets. This methodical approach ensures and harnesses pertinent segments of these datasets to underpin the foundational aspects of our proposed methodology.

Within our anomaly detection framework, the creation of positive and negative bags, crucial for refining the model's discrimination capabilities, involves a sophisticated integration of triplet loss and logarithmic loss mechanisms within the Deep Neural Network (DNN) model. The positive bags, comprising normal behavior samples within abnormal videos, are meticulously formed using the concept of triplet loss. This loss function operates by selecting triplets, each consisting of an anchor sample, a positive sample containing normal behavior, and a negative sample containing anomalous behavior. The fundamental objective is to minimize the distance between the anchor and positive samples while simultaneously maximizing the distance between the anchor and negative samples within the embedding space. Mathematically, this is expressed as the triplet loss function, aiming to ensure that the model learns to separate normal behavior samples from anomalous ones effectively.

Furthermore, to augment the learning process and enhance prediction accuracy, logarithmic loss is integrated into the training of the DNN model. This loss function, also known as cross-entropy loss, operates by penalizing incorrect predictions made by the model, especially when the model exhibits strong confidence in incorrect classifications. The logarithmic loss

function's objective is to refine the model's prediction accuracy by penalizing deviations from the ground truth labels across the dataset. This integration into the DNN model's training process helps in refining the model's ability to distinguish between normal and anomalous behavior by guiding the learning process with a focus on minimizing prediction errors. By integrating both triplet loss for positive and negative bag feature creation and logarithmic loss within the DNN model's training, our approach holistically optimizes the model's embedding space, enhances discrimination between normal and anomalous behaviors, and refines prediction accuracy.

The integration of Normalizing Flows, specifically Inverse Autoregressive Flows (IAF), enhances the anomaly detection framework by leveraging the generation of pseudo-positive features from normal videos. Normalizing Flows, a class of generative models, enable the transformation of data distribution, particularly useful in creating complex data distributions from simpler ones. In this context, IAF, a type of Normalizing Flow, facilitates the modeling of the underlying distribution of normal video features, allowing the generation of pseudo-positive features.

The process involves training the IAF on normal video feature distributions, enabling the generation of pseudo-positive features. These features encapsulate the essence of normal behavior, aiding in the augmentation of the model's understanding of diverse normal behavioral patterns. The transformation performed by IAF essentially enriches the model's comprehension of normal video feature distributions, contributing to a more comprehensive representation of typical behaviors observed in the normal video segments.

After the incorporation of Normalizing Flow and the generation of pseudo-positive features, the Deep Neural Network (DNN) is further utilized for classification within the anomaly detection framework. The DNN, serving as the core classifier, leverages the enriched features obtained from the combination of various methodologies. It takes the augmented feature representations, including those derived from the Normalizing Flows and the pseudo-positive features, as inputs for the classification task. The primary objective of this classification stage is to discern and categorize between normal behaviors and anomalies within the video datasets.

The enriched feature representations obtained through the incorporation of Normalizing Flows capture a diverse spectrum of normal behavioral patterns, whereas the pseudo-positive features specifically encapsulate learned representations of typical behavior derived from normal video segments. The DNN, through its sophisticated architecture and trained parameters, performs the task of learning and discerning these nuanced representations.

Upon completion of the training phase, the DNN model is capable of accurately classifying new instances, determining whether they exhibit normal or anomalous behavior. The classification decision is based on the learned representations and patterns extracted from the amalgamation of Normalizing Flows' enriched features and the pseudo-positive features generated from normal videos.

Our comprehensive methodology for anomaly detection within video datasets integrates an array of models to achieve robust discrimination between normal and anomalous behaviors. Initiated with skeletal sequences, our approach embraces the utilization of triplet loss and logarithmic loss within a Deep Neural Network (DNN), forming positive and negative bags essential for refining the model's understanding. Incorporating IAF, it enriches the model's comprehension by generating pseudo-positive features from normal videos, the DNN is further employed for classification, effectively discerning between normal and anomalous instances within the video datasets. Through the amalgamation of methodologies encompassing Normalizing Flows, and DNN classification, our methodology stands poised to elevate anomaly detection, providing a robust framework capable of accurately differentiating between normal behaviors and anomalies within video data. And the same has been shown in Fig 1.

IV. Results

In our evaluation, we employed the UCF Crime and XD Violence datasets, skeletal sequence extraction and results are shown in table 1 for each datasets. The training process encompassed a deep neural network (DNN) architecture trained with triplet and logarithmic loss functions, augmented with Normalizing Flows, specifically Inverse Autoregressive Flows (IAF), to generate pseudo-positive features from normal videos. Quantitative analysis revealed our methodology's robust anomaly detection performance, showcasing its accuracy, for both datasets and Fig.2 and Fig.3 represents the same. Comparative assessments against baseline models underscored the superiority of our approach. The results are shown in Table 1. Visual representations of enriched feature spaces elucidated the model's enhanced comprehension of diverse normal behavioral patterns.

TABLE I. THE TRAINING RESULTS OF THE DATASETS

Dataset	Accuracy	loss
XD Violence	93.2	0.3188
UCF Crime	91.4	0.6901

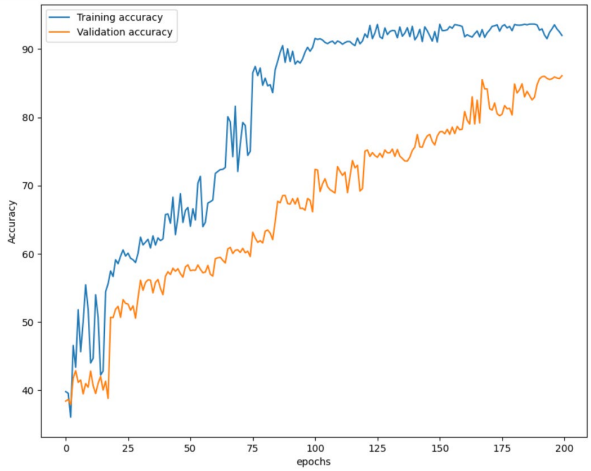


Fig. 2. Training graph of XD Violence dataset

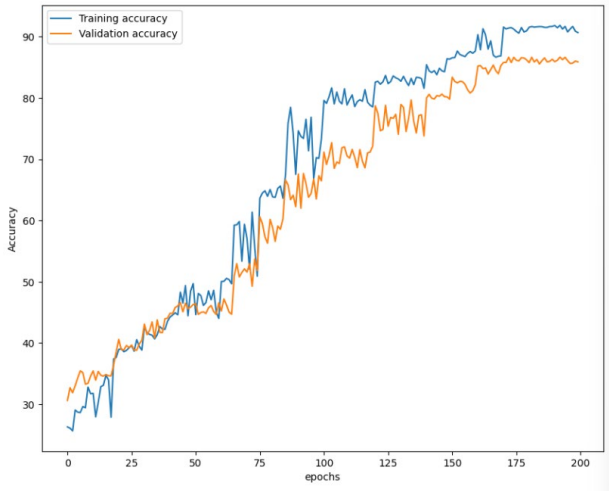


Fig 3. Training graph of UCF Crime dataset

V. Conclusion

Our exploration into anomaly detection using the semi-supervised method with video datasets has culminated in the development of a potent and adaptive framework. Leveraging a deep neural network (DNN), with IAF in our approach, showcases remarkable accuracy and precision in distinguishing between normal and anomalous behaviors. The results affirm the framework's robustness, validated through rigorous evaluations on diverse datasets, allocation strategies, and comparison against baseline models. The findings underscore the framework's capacity to effectively discern anomalies amidst complex video data. Moving forward, this methodology offers a promising trajectory for enhanced anomaly detection, paving the way for applications across surveillance, healthcare monitoring, and beyond, with potential avenues for further refinement and broader implementation.

References

[1] Li, Hao, Alin Achim, and D. Bull. "Unsupervised video anomaly detection using feature clustering." IET signal processing 6.5 (2012): 521-533.

[2] Hasan, Mahmudul, et al. "Learning motion patterns in crowded scenes for anomaly detection." Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops. 2016.

[3] Laine, Samuli, and Timo Aila. "Temporal ensembling for semi-supervised learning." arXiv preprint arXiv:1610.02242, 2016.

[4] Kiran, B. Ravi, Dilip Mathew Thomas, and Ranjith Parakkal. "An overview of deep learning based methods for unsupervised and semi-supervised anomaly detection in videos." Journal of Imaging 4.2 (2018): 36.

[5] Sabokrou, Mohammad, et al. "Adversarially learned one-class classifier for novelty detection." Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops. 2018.

[6] Zhang, Si, et al. "Graph convolutional networks: a comprehensive review." Computational Social Networks 6.1 (2019): 1-23.

[7] Wang, Jun, et al. "Abnormal event detection with semi-supervised sparse topic model." Neural Computing and Applications 31 (2019): 1607-1617.

[8] Ravanbakhsh, Mehdi, et al. "Abnormal event detection in videos using generative adversarial nets." arXiv preprint arXiv:1707.09750, 2017.

[9] Yan, Zheng, et al. "Spatial-temporal graph convolutional networks for anomaly detection in videos." Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2020.

[10] Ren, Jing, et al. "Deep video anomaly detection: Opportunities and challenges." 2021 international conference on data mining workshops (ICDMW). IEEE, 2021.

[11] Wang, Xuanzhao, et al. "Robust unsupervised video anomaly detection by multipath frame prediction." IEEE transactions on neural networks and learning systems 33.6 (2021): 2301-2312.

[12] Sarker, Mohammad Ibrahim, et al. "Semi-supervised anomaly detection in video-surveillance scenes in the wild." Sensors 21.12 (2021): 3993.

[13] Khaire, Pushpajit, and Praveen Kumar. "A semi-supervised deep learning based video anomaly detection framework using RGB-D for surveillance of real-world critical environments." Forensic Science International: Digital Investigation 40 (2022): 301346.

[14] Zhu, Yuansheng, Wentao Bao, and Qi Yu. "Towards Open Set Video Anomaly Detection." European Conference on Computer Vision. Cham: Springer Nature Switzerland, 2022.

[15] Zaheer, M. Zaigham, et al. "Generative cooperative learning for unsupervised video anomaly detection." Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2022.

Binary Image Shape Classification by LeNet-5

1st Mustafa Eren Yildirim
Department of Electronics and Communications Engineering
American University of Malta
Bormla, Malta
eren.yildirim@aum.edu.mt

2nd Yucel Batu Salman
Department of Software Engineering
Bahcesehir University
Istanbul, Turkey
batu.salman@bau.edu.tr

***Abstract*—Classification of the objects via their shapes is an interesting yet challenging topic. This task has been done in both 2D and 3D. Apart from the data type, also different approaches have been tried by using manually made feature extraction followed by a machine learning algorithm or directly by employing convolutional neural networks. This papers aims to classify objects from binary images. We used the MPEG-7 shape image dataset which is publicly available. We used LeNet-5 for this task and compared its performance with some other networks.**

Keywords—convolutional neural networks, binary shape, classification, LeNet-5

I. Introduction

Shape classification is one of the most popular and focused topics of image processing. This task is chosen to be performed on binary images instead of images with color and texture on. The reason of this is that the human eye is sensitive and capable of recognizing shapes rather than the color and texture.

Shape based object classification task can be achieved in two ways. The first one is the conventional method in which features are extracted manually and classification is done by one of the machine learning algorithms such as support vector machine (SVM) [1], dynamic time warping (DTW) [2], k-nearest neighbors (KNN) [3], random tree (RF) and forest (RF) [4]. In these methods, different types of features can be employed. One of these studies uses the combination of features based on contour and skeleton, according to the correspondence between and associating skeletal information with a shape contour [5]. Another one on the other side uses a shape representation called Bag of Contour Fragments (BCF) was developed, inspired by classical Bag of Words (BoW) model. In BCF, a shape is decomposed into contour fragments, each of which then is described using a shape descriptor, and is encoded into a shape code. Finally, a compact shape representation is built by pooling all of the shape codes in the shape [6]. Angular pattern, binary angular pattern, and a multiscale integration of both for shape retrieval are proposed for the efficient use of real-world shape retrieval applications [7, 8].

On the other hand, deep neural network (DNN) and convolutional neural network (CNN) are preferred to this job. In [9], four different CNN models are built for each channel, instead of using a single CNN. They receive the four input channels, and these four independent CNNs were trained in a sequence, and the weights of a trained CNN was used as starting point to train the other CNNs that processed the remaining channels. In [10], a new CNN model is proposed for the task of binary shape classification. Two prebuilt CNN architectures along with the proposed one are trained to recognize category of binary shapes.

In this paper, we analyze the classification performances of some of the most commonly used CNN models on a public dataset. Moreover, we optimize a model by tine tuning and perform it on the same dataset. The rest of the paper is organized as follows; in section II the used CNN model will be briefly explained, the experiments are given in Section III. The last but not least conclusion is in Section IV.

II. Convolutional Neural Networks

In the last decade, deep learning has improved fast and shown tramendous ability in object detection and recognition tasks. Thus, this study focuses on the use of convolutional neural network for shape classification tasks. In order to understand the deep learning, it is necessary to start from the neural network. Deep learning is a way of technology composed of a group of powerful neural networks [11]. Various convolutional neural network architectures along with their modified versions have been introduced since 2006. In this section we will explain one of the most fundamental models of CNN in the literature which is LeNet-5 since we use it for our task.

LeNet5 was the first CNN model and used for recognizing handwritten characters [12]. It is a gradient based model. The original LeNet-5 architecture block diagram is given in Fig. 1.

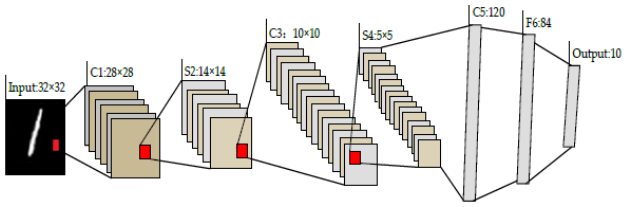


Fig. 1 Block diagram of LeNet-5 architecture.

It consists of convolution layers in which the feature extraction is realized. In this layer, features are extracted by kernels which can be 3x3, 5x5 and so on. This kernel is moved on the pixels to scan the entire image. The level of the information increases in each convolution layer. The first convolution layer extracts very basic features such as edges, color. However, the further layers are extracting features like the combination of prior features. Thus, in every layer of convolution, we learn deeper and more specific features of that image.

Each convolution layer is then followed by pooling layers. The pooling layer is used to select and eliminate features for

dimension reduction. The size of the kernel used in this layer affects the new dimension and the amount of information preserved. While there are three main types of pooling as average pooling, maximum pooling and random pooling, LeNet-5 uses average pooling.

The last convolution layer is connected to the fully connected (FC) layer. The neurons in this layer are using an activation layer and connected to the neurons of the previous layer. In this layer, the information coming from previous layers are merged. As a result of this, meaningful and distinguishing information is obtained.

The last layer is the output layer with size equal to the number of classes in the dataset. It is also named as softmax layer. Each neuron in this layer has a probability value of the input image to belong to neuron (class).

III. Experiments and Results

In this section, we explain the used dataset, performance metrics, and results of the models. The data preparations are made in Python and classification is made by DeepLearning4J package of Weka tool. All parts of this study run on a computer with Intel(R) Core(TM)i5-10400 2.90 GHz CPU and 8 GB RAM under Windows 10 operating system.

A. Dataset

MPEG-7 shape image dataset consists of 1,400 binary images of 70 different classes as shown in Fig. 2. There are 20 images per class with variations such as scale, translation, rotation, and deformation. This dataset is generated and usually used for image retrieval [13]. However there are several studies which are using it for classification as well.

Fig. 2 Mpeg-7 image dataset samples.

B. Tests and Results

Before applying the model on the dataset, we randomly split the dataset for train and test by 70% and 30% respectively. We did not make any preprocessing nor any type of data augmentation since the dataset contains variations of each shape. The size of the input images is 28x28 pixels.

The hyper-parameters of the used LeNet-5 model are set as follows. In all experiments, networks are trained by stochastic gradient descent (SGD) with 0.9 momentum. The learning rate is 0.001. The weight decay parameter is 0.0005. Initial weights are selected by Xavier initialization. We also use dropout with 0.5. The mini-batch size is set to 100 as well as the maximum number of epochs. We run the tests multiple times to find the optimum epoch number. The accuracy versus epoch relation is illustrated in Fig. 3. The highest classification rate is obtained with 8 epochs. Thus, it is chosen for benchmark.

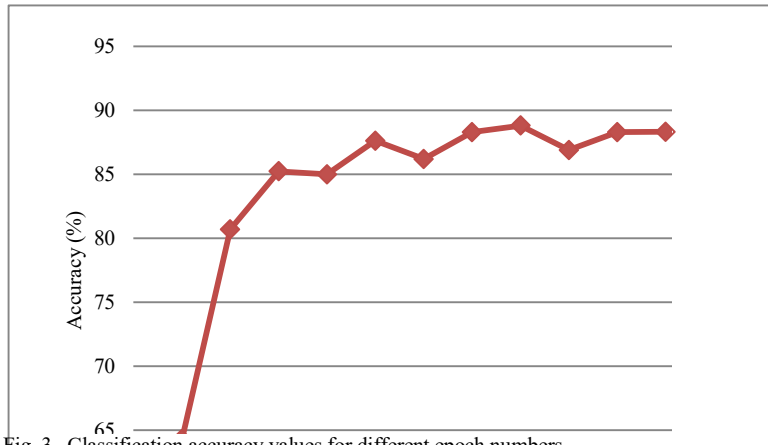


Fig. 3 Classification accuracy values for different epoch numbers.

In advance to benchmarking with other studies and methods, we optimized our network by validating it on the training set. We set the configurations which supplied 100% accuracy on the training set as our default settings.

TABLE I. The Classification Results on MPEG-7 Dataset

Method	Accuracy (%)
MNIST-CNN [10]	86.43
LeNet [10]	87.14
Our LeNet-5	88.81
BSCNN [10]	89.64
SCN [11]	90.99

We compared the performance of our model with other DL-based methods. The results belong to the papers in the literature which used the same dataset for classification. We did not include the results of retrieval.

Binary-shape CNN (BSCNN) and MNIST-CNN models achieved 86.43% and 89.64% respectively. The conventional LeNet obtained 87.17% which is slightly less than our model. The SCN architecture obtained the highest among all with 90.99%. However, the authors used data augmentation prior to implementation. Instead of 1,400 images, they used 50,400 images which lead to a better performance.

IV. Conclusion

In this study, we aim to handle the shape classification task on binary images by using a fine-tuned CNN architecture. We use the LeNet-5 as our base model. The proposed model is then validated and tested on a publicly available dataset which contains 1,400 images of 20 different objects. Our model did not outperform all of the DL-based studies on the dataset however it showed state-of-art performance. As our next step, we will update our model on the augmented version of the same and different benchmark datasets.

REFERENCES

[1] M. Keivani, J. Mazloum, E. Sedaghatfar, M.B. Tavakoli, "Automated analysis of leaf shape, texture, and color features for plant classification," Traitement du Signal, vol. 37, No. 1, pp. 17–28, 2020.

[2] H. Li et al., "Adaptively constrained dynamic time warping for time series classification and clustering," Information Sciences, vol. 534, pp. 97–116, 2020.

[3] S. R. Shinde and S. D. Thepade, "Gender classification with KNN by extraction of Haar wavelet features from canny shape fingerprints," International Conference on Information Processing (ICIP), Pune, India, 2015, pp. 702–707.

[4] D. Damopoulos, T.D. Lerch, F. Schmaranzer, *et al.,* "Segmentation of the proximal femur in radial MR scans using a random forest classifier and deformable model registration," Int. J. CARS, vol.14, pp. 545–561, 2019.

[5] W. Shen, Y. Jiang, W. Gao, D. Zeng, X. Wang, "Shape recognition by bag of skeleton-associated contour parts," Pattern Recognition Letters, vol. 83, Part 3, pp. 321–329, 2016.

[6] X. Wang, B. Feng, X. Bai, W. Liu, L. J. Latecki, "Bag of contour fragments for robust shape classification," Pattern Recognition, vol. 47, Issue 6, pp. 2116–2125, 2014.

[7] M. E. Yildirim, O. F. Ince, B. S. Yucel, and I. F. Ince, "Shape retrieval using angle-wise contour variance," Journal of Electrical Engineering, vol. 72, no. 2, pp. 99–105, 2021.

[8] H. Ling, X. Yang, and L. J. Latecki, "Balancing deformability and discriminability for shape matching," Proc. ECCV, Crete, Greece, 2010, pp. 411–424, 2010.

[9] L.A. Alexandre, "3D object recognition using convolutional neural networks with transfer learning between input channels," In: Intelligent Autonomous Systems, vol. 13, pp. 889–898, 2016.

[10] A. H. Atabay, "Binary shape classification using Convolutional Neural Networks" IIOAB Journal, vol. 7, pp. 332–336, 2016.

[11] C. Zhang, Y. Zheng, B. Guo, C. Li, and N. Liao, "SCN: A Novel Shape Classification Algorithm Based on Convolutional Neural Network," Symmetry, vol. 13, no. 3, p. 499, March 2021.

[12] Y. LeCun, L. Bottou, Y. Bengio, P. Haffner, "Gradient-based learning applied to document recognition," Proc. IEEE, vol. 86, pp. 2278–2324, 1998.

[13] M.E. Yildirim, "Quadrant-based contour features for accelerated shape retrieval system," J. Electr. Eng., vol. 73, pp. 197–202, 2022.

Empowering Education: AI-Driven Customized Learning for Inclusive Education of Differently-Abled Students in India

Anjali Gupta
Department of Statistics & Data Science
CHRIST (Deemed to be University),
Bengaluru - 560029, Karnataka, India

Dr Guydeuk Yeon*
Centre for Digital Innovation
CHRIST (Deemed to be University),
Bengaluru - 560029, Karnataka, India
*Corresponding author: yeon@christuniversity.in

***Abstract*— As lifestyles evolve, educational systems must adapt to meet the changing needs of learners. Physical disabilities can pose significant barriers to accessing educational resources and tools. Many individuals with physical disabilities face limitations in using technology due to the lack of accessible features and tools. As per 2011 census there are 2.68 crore differently abled persons in the country. Artificial Intelligence can play a significant role in addressing these challenges and helping individuals with disabilities become more independent, including in their education. This paper centers on the utilization of Artificial Intelligence (AI) to facilitate customized learning for differently-abled students within the Indian educational landscape. The focus lies in delivering recommendations for the effective implementation of this approach, aiming to overcome barriers and create an inclusive educational environment.**

The objective is to offer actionable suggestions to enhance accessibility, adaptability, and individualized learning for students facing various disabilities in the Indian educational system.

Keywords*— *Artificial Intelligence; Differently-Abled Students; Customized Learning; India.

I. INTRODUCTION (*HEADING 1*)

The needs of children with disabilities have long been disregarded by India's educational system. The schools have discovered that they lack the resources necessary to accommodate students with disabilities, whether it be in the form of inclusive classroom environments or adequate infrastructure. The Indian government has launched a number of programmes over the years to address the difficulties that people with disabilities face, especially in the area of education. The 2016 Law on the Rights of Persons with Disabilities, which superseded the 1995 Act and established a thorough framework for defending the rights and entitlements of people with disabilities, including access to education, is one important piece of legislation. The federal government's Inclusive Education for Disabled at Secondary Stage (IEDSS) programme provides states and union territories with financial support to improve the education of secondary school students with disabilities. Furthermore, the National Scholarships platform was created by the government to expedite the application process for students with disabilities wishing to apply for financial aid and scholarships. To further promote education, the government has implemented programs promoting the use of assistive technology, such as screen reading and braille displays. These policies collectively aim to facilitate the academic and professional advancement of individuals with disabilities in India.

II. INITIATIVES FOR DISABLED INDIVIDUALS IN INDIA

A. Budget for Disabled Persons

Although there are many programs aimed at facilitating the education of those with disabilities, the funding allotted for these programs requires rethinking. In FY 2023–24, the Department of Empowerment and Welfare of Persons with Disabilities (DEPwD) of the Ministry of Social Justice and Empowerment (MSJE) got INR 1,225 crore, or 0.027 percent of the total budget. The total amount spent in FY 2023–2024 rose by almost 7.5% over the prior year, however, the DEPwD expenditure only climbed by 1%. An improvement but more needs to be done.

In India, financial support for government policies and programmes that cater to the needs of the disabled is essential. It guarantees the use of inclusive policies that assist the community of disabled people. These funds are necessary to give those with disabilities access to jobs, healthcare, education, and infrastructure.

Budgetary assistance also contributes to the advancement of assistive technology research and development, which has the potential to greatly enhance the quality of life for those with disabilities. The government can foster a more inclusive society and enable individuals with disabilities to live independently and with dignity by allocating funds for policies and programmes that cater to their needs.

To ensure better programs for persons with disabilities, a strong budgetary allocation is required. This can be a game turner for a huge populated country like India. A large population implies an equally big number of differently-abled people and children as per the laws proportion. According to some estimates, out of the total population of the country in 2021, 2.21% were differently-abled.

B. *Differently-abled in Education*

It is important to give disabled children access to education so they can acquire the skills they need to live independent, happy lives. Through education, they can fulfil their potential, make a constructive contribution to society, and become financially independent. However, due to a shortage of qualified teachers, accessible curriculum, and accessible infrastructure, disabled children frequently confront substantial obstacles in their pursuit of a high-quality education. Schools and other educational facilities need to be outfitted with the right equipment, such as ramps, lifts, and wheelchair accessibility, in order to guarantee inclusive education for children with disabilities. In order to meet the distinct learning requirements of children with disabilities, teachers must also complete specialised training that covers the use of assistive technology and alternative teaching strategies.

The curriculum must also be designed in a way that it can accommodate the learning needs of disabled children, with an aim to develop their communication, cognitive, and social skills. Inclusive education is a good way to promote social inclusion by fostering friendships and reducing the stigma associated with disability. Providing equal educational opportunities for disabled children requires a collaborative effort between the educational institutions, government, and the community.

The government must provide adequate resources required to support inclusive education. Educational institutions and the community must come together as a team to create a supportive and inclusive environment for disabled children. Government policies for disabled in India are a must. By prioritizing inclusive education, we can create a more equitable society and empower disabled children to reach their full potential.

C. *Initiatives Taken*

Over the past few years, India has implemented a number of initiatives to establish an inclusive learning environment for specially abled children. All children, including those with disabilities, are guaranteed access to education through the Sarva Shiksha Abhiyan program. Children with disabilities must receive an inclusive education in regular schools, according to the Right to Education Act.

Through the provision of assistive technology and special education teachers, the Samagra Shiksha Abhiyan programme seeks to raise the standard of education for kids with disabilities. Ramps, elevators, and accessible restrooms are just a few of the accessibility features that the Accessible India Campaign advocates for in educational institutions. Additionally, to guarantee differently-abled students' access to higher education, the Ministry of Social Justice and Empowerment offers scholarships to them.

III. Enhancing AI and Special Need Education

The field of artificial intelligence (AI) has seen a number of studies. Typically, it is described as "the mechanical simulation system of gathering knowledge and information and processing universely intelligence: (gathering and interpreting) and distributing it to the qualified in the form of actionable intelligence" (Morrison, 2017). It is made up of information agents, which can be virtual like software or physical like robots or other devices. Morrison asserts that as AI advances, it will be more crucial than ever to comprehend the kinds of tools that people with disabilities would need in their technological toolbox. The benefits of AI have been acknowledged in education; however, the research fraternity has started exploring its benefits for people with special needs in education (M. A. Gernsbacher, 2016). AI and Special need Education can collaborate together to enable development of individuals suffering from disabilities. Students with learning, oratory, visual and mobility impairment can seek the benefits with the help of Artificial Intelligence in education. The research study has also indicated AI as one of the assistive technology for PwDs. As per an article by (Lynch, 2018), AI has provided around-the-clock care using Robotics for people with disabilities. AI has helped the people to use mobile applications without even clicking on it. For example, “Siri” in mobiles has enabled people to access everything in the mobile without even actually clicking taking the effort to access that. Another such example is Google “Alexa” that has enabled people to demand any information without typing on the search bar. Both these applications work on the concept of AI (speech recognition). Such applications can provide assistive services to people if used in a productive way for the needy.

A. *Specific Learning/ Language Disabilities*

In every institution each and every student requires some or the other assistance at their level in reading, writing, pronouncing and understanding. The level of learning and language disabilities are required to be identified and diagnosed at the tender age of a child (S. Drigas, 2012). Although there are manual techniques to diagnose learning and language disabilities, however, they can be tackled by machine learning algorithms as expert systems that can be fed to diagnose such disabilities (J. Prentzas , 2013). Various computers based tools as educational technologies has been development to manage the children with dyslexia; these tools include Phonological Awareness Educational Software‟ (PHAES) it facilitates the reader with Phonological awareness. They come as software that can be easily navigated and can be understood with the help of simple graphics without cognitive overload said by Gernsbacher. Robotic assistance is been implemented for managing the children suffering from Autism Spectrum Disorder (ASD) so that they can learn social skills understand the emotions (Y. D. Kim, 2010). For instance, a British Primary school used Humanoid Robots for children suffering from ASD to understand Human Emotional State said by Y. D. Kim . A game called LIFEisGAME is developed for kids with ASD which helps them to understand facial expressions, also an e-learning tool called Smart Tutoring Model made for children suffering from learning disabilities such that they could understand in an easy and better way. (Prentzas , 2013), (Kim Y-D,2010) research results have indicated that robots can be used as an assistive tool for people with disabilities, another such indication is that robot-assisted observation systems are used for children with autism in kindergarten for observing their development. Artificial Neural Networks and Fuzzy Logic can be used for diagnosis and accessing the children suffering from ASD, the pseudo algorithm is applied to access the autism disorder (S. Drigas, 2012). Equadex, an AI tool from Microsoft, is intended primarily for kids with autism. It is a mobile-navigated environment that facilitates easy communication and

comprehension through visual presentations (A. Allebee, 2017). AI is also used in diagnosis. For example, a web-based deep learning system is used to access children with autism and learning disabilities and sends reports to parents and the medical officer.

B. Hearing Impairment

Microsoft comes with Microsoft Translator as an AI assistive device for deaf learners. The device comes with a headset worn by the speaker that translates speech signals into captions that are visible to the learner with hearing impairment. The translator provides the translation in 60 different languages, giving an ease for deaf learners to understand (Roach, Sam McNeill, 2018). (Zeng, Fan-Gang, 2017) Artificial intelligence (AI) and deep learning technologies, like virtual and augmented reality, are used in over-the-counter (OTC) hearing aids to enhance hearing even in individuals with normal hearing. With features like sound scene analysis, sound protection and warning, and even real-time language translation, an intelligent hearing aid will provide a customised and interactive hearing experience.

C. Visual Impairment and Physical Impairment

It takes a while for children with visual impairment to adjust in a mainstream classroom, but it is not something that is impossible. These students require different setup which can not only ease their learning process but can also make them feel included. (Mboshi, 2018) suggests classroom teachers to adopt unique teaching ways for effective and inclusive classroom setting. Visually impairment may include totally blind or low-vision impairment, for any kind of visual impairment, visual devices (such as Braille, change in text size, color etc.) available to give them an access to equal education. Learner's with physical impairment requires support in availing educational services which helps them to maintain their autonomy during the educational course.

IV. Conclusion

The aforementioned study investigated how artificial intelligence (AI) is influencing people from all walks of life, particularly those of children with special needs. Its main effects are seen in the field of education, where organizations, educators, and parents work together to advance inclusive learning through assistive technologies that can eliminate the need for manual labour and advance learning without bias. The study also recommended using inclusive pedagogy, which incorporates all students without distinction. It also recommends creating pedagogies that foster creativity and aid in establishing a secure atmosphere for kids in which they can freely converse and share ideas while respecting the unique qualities of each person. I understand with proper collaboration of government, community and AI technology can help make this country a better learning ground for the differently abled kids. Fixing the loopholes that lies in the different initiative can make our vision of giving inclusive learning environment to these kids successful. This is a long process but with constant effort, dedication and management we will soon be able to achieve this.

References

The template will number citations consecutively within brackets [1]. The sentence punctuation follows the bracket [2]. Refer simply to the reference number, as in [3]—do not use "Ref. [3]" or "reference [3]" except at the beginning of a sentence: "Reference [3] was the first ..."

Number footnotes separately in superscripts. Place the actual footnote at the bottom of the column in which it was cited. Do not put footnotes in the abstract or reference list. Use letters for table footnotes.

Unless there are six authors or more give all authors' names; do not use "et al.". Papers that have not been published, even if they have been submitted for publication, should be cited as "unpublished" [4]. Papers that have been accepted for publication should be cited as "in press" [5]. Capitalize only the first word in a paper title, except for proper nouns and element symbols.

For papers published in translation journals, please give the English citation first, followed by the original foreign-language citation [6].

[1] WHO. (2011). World Report on Disability. [Online]. Available: https://www.who.int/disabilities/world_report/2011/report.pdf.

[2] M. Laabidi, M. Jemni, L. J. B. Ayed, H. B. Brahim, and A. B. Jemaa, "Learning technologies for people with disabilities," Research Laboratory of Technologies of Information and Communication & Electrical Engineering LaTICE, National Higher School of Engineering of Tunis, 2013.

[3] UNESCO Global Report: Opening New Avenues for Empowerment: ICTs to Access Information and Knowledge for Persons with Disabilities, United Nations Educational, Scientific and Cultural Organization (UNESCO), 2013.

[4] Microsoft annual report. (2017). Letter to Shareholders. [Online]. Available: https://www.microsoft.com/investor/reports/ar17/index.html#

[5] S. Drigas and R.-E. Ioannidou, "Artificial intelligence in special education: A decade review," International Journal of Engineering Education, vol. 28, no. 6, pp. 1366–1372, 2012.

[6] M. A. Gernsbacher, A. R. Raimond, M. T. Balinghasay, and J. S. Boston, "Special need is an ineffective euphemism," Cognitive Research: Principles and Implications, vol. 1, no. 1, p. 29, 2016.

[7] R. J. Berger, Introducing Disability Studies, Boulder: Lynne Reiner Publishers, 2013.

[8] D. S. Grewal, "A critical conceptual analysis of definitions of AI as applicable to computer," IOSR Journal of Computer Engineering, vol. 16, issue 2, pp. 9-13, 2014.

[9] C. Morrison, E. Cutrell, and A. Dhareshwar, "Imagining AI applications with people with visual disabilities using tactile ideation," in Proc. ASSETS '17, 2017.

[10] Lynch. (2018). How AI is improving assistive technology. [Online]. Available:https://www.thetechedvocate.org/how-artificial-intelligence-is-improv ing-assistive-technology/

[11] J. Prentzas. (2013). AI methods in early childhood education. [Online]. Available: https://www.researchgate.net/publication/287644942_Artificial_Intell igence_Methods_in_Early_Childhood_Education

[12] Y. D. Kim, J. W. Hong, W. S. Kang, S. S. Baek, H. S. Lee, and J. An, "Design of robot assisted observation system for therapy and education of children with autism," in Proc. ICSR 2010, 2010.

[13] A. Allebee. (2017). Equadex, Microsoft cognitive services. [Online]. Available: https://customers.microsoft.com/en-us/story/equadex-partner-professi onal-services-cognitive-services

[14] J. Roach. (2018). AI technology helps students who are deaf learn. [Online]. Available: https://blogs.microsoft.com/ai/ai-powered-captioning/

[15] Sam McNeil. (2018). AI technology assisting deaf students. [Online]. Available: https://educationblog.microsoft.com/2018/04/ai-technology-assisting- deaf-students/

[16] F. G. Zeng, "A new landscape for hearing aids," The Hearing Journal, vol. 70, issue 12, p. 6, 2017.

[17] N. S. Mboshi, "Teaching learners with visual impairment in an inclusive education setting: The cameroon perspective," International Journal of Education and Research, vol. 6, no. 2, February 2018

[18] UNICEF, "Teachers, inclusive, child-centred teaching and pedagogy," in Webinar 12 - Companion Technical Booklet, 2014.

Revolutionizing Rural Education: A Comprehensive Review of AI-Enabled Remote Education Initiatives in India

Anjali Upadhyay
Department of Statistics & Data Science
CHRIST (Deemed to be University),
Bengaluru - 560029, Karnataka, India

Dr. Guydeuk Yeon*
Centre for Digital Innovation
CHRIST (Deemed to be University),
Bengaluru - 560029, Karnataka, India
*Corresponding author: yeon@christuniversity.in

***Abstract*— In the realm of rural education in India, Artificial Intelligence (AI) takes center stage, unlocking transformative possibilities and reshaping the educational landscape. Its role goes beyond the buzz, delving into the very fabric of education in remote schools. In a setting where challenges often outnumber resources, AI steps in as a dynamic force, not to overshadow traditional methods but to enhance and customize them. This research explores the transformative role of AI in rural education across India, focusing on the substantial impact it can have on accessibility and quality. Against the backdrop of the rural-urban educational divide, the study meticulously reviews AI-enabled remote education initiatives implemented by government, NGOs, and EdTech companies. The research not only identifies the strengths and weaknesses of existing initiatives but also proposes strategic improvements for optimizing their impact. The findings aspire to contribute significantly to the ongoing discourse on leveraging AI for inclusive and effective education in rural India, offering valuable insights for policymakers, educators, and stakeholders invested in the advancement of educational accessibility and quality.**

Keywords*— *Artificial Intelligence, Rural, Education, Initiatives, Review, Stakeholders

I. INTRODUCTION (*HEADING 1*)

In India's expanding rural education scene, where resources are sometimes insufficient to address difficulties, artificial intelligence (AI) is emerging as a potent tool that is revolutionizing conventional teaching and learning approaches. This research investigates AI-enabled remote learning programmes in India in detail, emphasizing how they have revolutionized accessibility and quality in rural schools. AI has a deeper impact on education in distant environments than just being a trendy new technology. Instead of displacing traditional approaches, it helps to dynamically improve and tailor them. In order to bridge the educational gap between rural and urban areas, this study carefully examines programmes run by governmental agencies, non-governmental organizations (NGOs), and creative EdTech businesses. These programmes act as shining examples of innovation, aiming to overcome the challenges that persist in the realm of rural education.

II. AI-ENABLED REMOTE EDUCATION INITIATIVES

A. ConveGenius

In India's ever-changing rural education scene, where resources are frequently insufficient to meet problems, ConveGenius stands out as an innovative leader, spearheading a change in the way children and young people receive high-quality education. With its state-of-the-art Personalised and Adaptive Learning (PAL) platform, ConveGenius, positioned at the nexus of EdTech and social enterprise, is not only addressing but also reinventing educational injustice. ConveGenius uses a multipronged strategy to close learning and skill gaps, with PAL Labs serving as the foundation for its transformative projects. These labs, which are referred to as data-driven, internet-free digital classrooms, mark a paradigm shift in the way distant learning is provided. PAL Labs are dynamic spaces that help students learn in a simple, secure, and enjoyable way while being supervised by a facilitator or instructor. Their unique approach to learning is called nudged learning, which combines manual, data-driven initiatives with a sequence of automatic technological triggers to produce an environment that is both self-paced and adaptable. The PAL platform adjusts its learning paths to meet individual needs and provides a personalised learning experience for each child. It is available in online, offline, and hybrid modes.

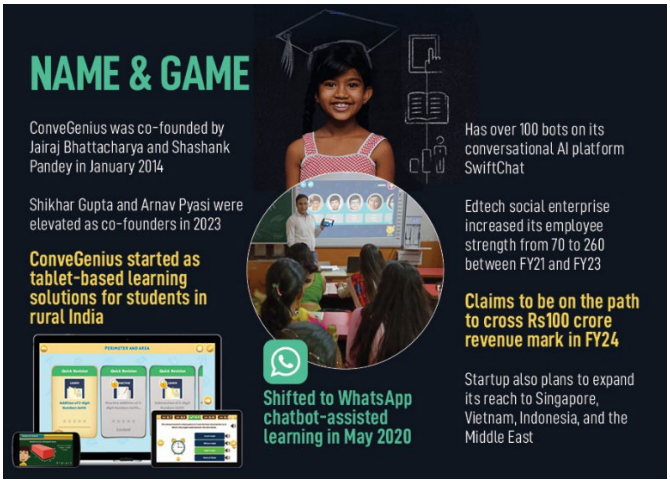


Fig. 1. Decoding ConveGenius (Source: Forbes India,2023)

Fig. 2. Structure showing PAL Lab's working (Rephrased from: https://convegenius.com/home)

Jairaj Bhattacharya and Shashank Pandey created ConveGenius in 2014 with the goal of giving free education to 100 million underprivileged Indian schoolchildren who live at the base of the socioeconomic pyramid. They developed a business-to-business income model after realising the necessity for sustainability and forming partnerships with government, affordable, private, corporate, and non-profit educational institutions. But the COVID-19 epidemic was the turning point that expedited ConveGenius's expansion and influence. Despite initial skepticism, they used WhatsApp to deliver education, adjusting to the socioeconomic realities of rural India. The startup shot to 15 million pupils in just one year, having first reached half a million in six years. The success highlights the value of adaptability and creativity by showing how unusual platforms like WhatsApp can democratise education on a never-before-seen scale. The founders' dedication to maintaining free education while investigating revenue streams demonstrates the possibility of aligning corporate objectives with social effect. Those looking to combine commercial sustainability with meaningful solutions in the education sector can learn a lot from ConveGenius's story. (Forbes India,2021)

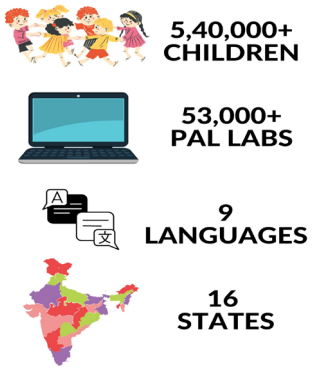


Fig. 3. Reach of ConveGenius (Data from: https://convegenius.com/home)

According to the co-founder of ConveGenius, the following are the strategic Interventions for optimizing their impact: (Report on Budget Private Schools in India, 2017)

1) Efficacy Measurement and Communication:

- Implement a robust efficacy measurement system and standardized benchmarks. Clearly communicate these outcomes to schools and parents, emphasizing transparent reporting for increased trust and confidence.

2) Strategic Partnerships for Comprehensive Solutions:

- Explore partnerships with reputable vendors, content aggregators, and school financing companies. Leverage existing content through collaborations to enhance the overall quality and comprehensiveness of educational offerings.

3) Data-Driven Education Models:

- Embrace a data revolution in education. Incorporate data and analytics into the business model to personalize learning experiences, track learner outcomes, and provide actionable insights for educators, parents, and students. Optimize technology for improved educational effectiveness.

B. Schoolnet

Schoolnet, a pioneering force in the field of education, is playing a crucial role in ensuring that India's poorest children receive quality learning opportunities. With a focus on government and affordable private schools, Schoolnet addresses the pressing challenges faced by the education system in impoverished regions. One of its standout initiatives is the development of KYAN, a compact and versatile device that serves as an all-in-one solution for classrooms in rural areas with limited internet connectivity. This 4kg plug-and-play device, rotating among different grades, transforms ordinary classrooms into smart classes, enhancing the learning experience for students. Unlike other edtech companies, Schoolnet's unique

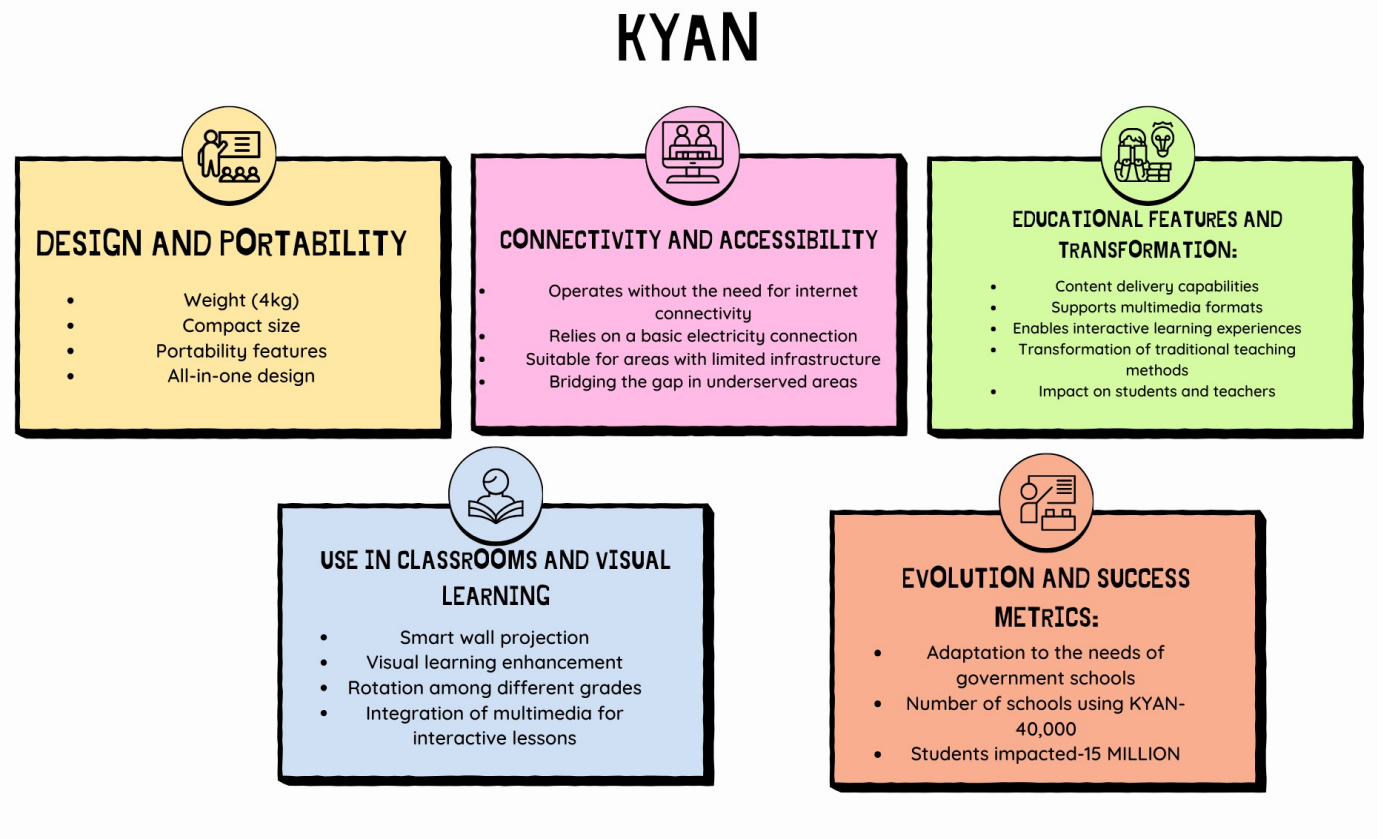


Fig. 4. KYAN Specifications (Rephrased from: https://www.schoolnetindia.com/offerings/digital-classrooms/k-yan.php)

approach involves working collaboratively with schools and teachers, positioning itself as an ally rather than a replacement for traditional education. The success of Schoolnet is evident in its positive impact, reaching over 40,000 schools, 15 million students, and one million teachers. By integrating technology seamlessly into the learning environment and emphasizing a holistic ecosystem approach, Schoolnet is contributing significantly to narrowing the education gap and ensuring that India's youth are equipped with essential foundational skills for a brighter future.

Strategic Improvement and Lessons for Others:

To enhance the effectiveness of the Schoolnet initiative and provide valuable insights for other stakeholders, the following strategic improvements can be proposed as mentioned by an expert from a renowned company. (Forbes India, 2023)

1) Infrastructure Development:

- Collaborate with government bodies and private entities to address the "digital access divide" by providing essential infrastructure, including projectors and computers, to rural schools.

2) Localized Content Creation:

- Invest in the creation of high-quality educational content in local languages, ensuring cultural relevance and enhancing accessibility for students in diverse linguistic backgrounds.

3) Comprehensive Teacher Training:

- Establish robust teacher training programs to enhance digital literacy among educators, empowering them to effectively utilize digital tools in the teaching-learning process.

4) Versatile Digital Tools:

- Continue expanding the capabilities of digital tools to serve various educational needs, including lesson planning, assessments, and administrative reporting, ensuring a comprehensive impact on the education ecosystem.

5) Advocacy and Partnerships:

Actively engage with central and state governments, school administrators, and teachers to build support and enthusiasm for the integration of technology in education. Foster partnerships with relevant stakeholders to amplify the initiative's reach.

C. WeJump

Established in 2021 amid the challenges of the pandemic, WeJump is a transformative force in education, specifically targeting underserved children. The organization's holistic approach revolves around empowering both educators and students on a global scale. Strategic teacher training programs encompass elements like planning, trainer training, course content development, and personalized 1:1 support, ensuring educators are equipped with the tools for effective teaching methods. Simultaneously, WeJump extends its impact directly to students, offering training in collaboration, financial support,

career guidance, and personalized 1:1 support, fostering a holistic educational experience. The organization's commitment to collaboration shines through initiatives such as online forums, in-person workshops, and joint projects, connecting schools globally and creating a supportive network. The innovative Dream Fund further exemplifies WeJump's dedication by providing financial support for creative educational projects, encouraging schools to explore novel approaches to teaching and learning. At the core of WeJump's success are the Master Trainers—experienced educators who play a crucial role in disseminating knowledge and supporting both teachers and students. More than an organization, WeJump is a global community committed to positive change, providing a platform for schools to connect, learn, and collectively enhance education for all.

Fig. 5. WeJump Structure (Source: https://www.wejump.org/)

Among these initiatives, WeJump's Treasure Hunt-A Python adventure flipbook, created using AI, stands out as an exemplary illustration of how AI can be harnessed to make learning Python engaging and enjoyable for children. As part of the broader review of initiatives, this unique approach exemplifies the flexibility and creativity AI brings to education.

Fig. 6. Flipbook-WEJUMP TREASURE HUNT (Source: https://www.wejump.org/)

III. CONCLUSION

As the paper details, AI has a significant impact on improving the quality and accessibility of education in rural areas, which is quite remarkable. ConveGenius, Schoolnet, and WeJump are a few examples of these efforts, which demonstrate creative methods of providing education in remote areas and provide a standard for other EdTech businesses and non-governmental organizations with related goals. These initiatives redefine educational fairness by providing individualized, flexible learning opportunities, going beyond simple innovation. Interestingly, they provide smart classroom solutions that are made to function well in places with spotty internet access. Effective collaboration among stakeholders at all levels, including as students, educators, legislators, and local communities, is critical to the success of these technology initiatives. In order to guarantee that education in rural areas is truly available to all children, there needs to be a communal effort. It is impossible to overestimate the significant contribution that the government can make in providing the infrastructure required for better execution of these programmes. Developing children's digital literacy also appears to be essential to ensuring that these technologies reach their full potential. To improve reach and efficacy in varied rural communities, local languages must be incorporated into the initiative's content. In essence, the transformative potential of AI in rural education unfolds fully when coupled with robust collaboration and support from all levels of stakeholders.

REFERENCES

[1] https://www.forbesindia.com/article/take-one-big-story-of-the-day/decoding-convegeniuss-common-sense-edtech-playbook/86691/1

[2] https://www.forbesindia.com/article/startups/convegenius-provides-free-education-to-rural-kids-can-it-make-money/68511/1

[3] Report on Budget Private Schools in India 2017, Centre for Civil Society

[4] https://www.forbesindia.com/article/take-one-big-story-of-the-day/how-schoolnet-is-ensuring-indias-poorest-children-dont-fall-further-behind/85293/1

The Effect of Institutional Ownership on the Relationship between CEO tenure and Firm Performance

Duangnapa Sukhahuta
Faculty of Business Administration,
Maejo University, Chiangmai, Thailand
duangnapa@mju.ac.com

Abstract— This research aims (1) to study the relationship between the CEO tenure and the firm performance and (2) to study the moderator influence on the relationship between CEO tenure and firm performance. The statistics used in the study include (1) descriptive statistics and (2) multiple regression analysis. The study findings reveal that CEO tenure has significantly positive association with the performance at 0.01 level. This is consistent with the Stewardship Theory, which explains that managers act as stewards to preserve the benefits of the organization more than their own benefits. Regarding the moderated variable testing, by studying variables that interact with the independent variables affecting the dependent variable, this research examines the interaction between the influence of CEO tenure and institutional ownership. The results of the study indicate that the interaction between the influence of CEO tenure and the institutional ownership has a negative impact on firm performance at the statistically significant level of 0.05, indicating that the increased presence of institutional investors leads to increased scrutiny of the company. However, when CEO tenure hold their positions for an extended period, their self-interest may have a negative impact on the firm performance.

Keywords—component, formatting, style, styling, insert (*key words****)***

I. INTRODUCTION

Chief Executive Officer (CEO), the highest position in the organization, has the responsibility of setting the direction, vision, and goals of the organization, serving as a guide for both short and long-term operations. Additionally, the CEO acts as a strategist in developing strategies to achieve the company's objectives, including resource allocation and investment decision-making. They analyze and manage risks that may affect the organization, necessitating the establishment of good relationships with stakeholders, such as investors, partners, and employees. This role is highly crucial for the company's success.

Within the organization, the CEO plays a vital role in overseeing and controlling operations to ensure alignment with objectives, regulations, policies, business strategies, operational plans, financial targets, and resolutions passed by the board of directors or shareholders' meetings. Furthermore, the CEO has the authority to determine the organizational structure, management methods, recruitment, training, hiring, and termination of employees, as well as setting wages, salaries, benefits, bonuses, and employee welfare. They are also responsible for signing legal agreements and various documents to facilitate the efficient and effective operations of the company.

However, the prolonged tenure of a CEO can impact the firm performance, both positively and negatively. Increased experience and improved abilities resulting from long-term tenure can positively influence the firm performance. Conversely, extended CEO tenure can influence the board's independence negatively, leading to their interference in operations. This weakens the oversight structure, indicating a reduced level of scrutiny. Furthermore, the CEO, as a representative of the shareholders, is not the owner of the business, according to the Agency Theory (Jensen and Meckling, 1976). CEOs tend to prioritize their own benefits over the company's, leading to conflicts when they mismanage operations to favor their interests. This aversion to high-risk investments and preference for guaranteed returns can create conflicts, as they fear the implications for their position.

II. LITURATURE REVIEW AND HYPOTHESIS DEVELOPMENT

A. Agency Theory

One of the well-known financial theories that has been extensively applied in corporate finance is the agency theory. Jensen and Meckling (1976) define the agency relationship as a contract between two parties where one is a principal (shareholder) and the other is an agent (manager) who represents the principal in transactions with a third party. Agency relationships occur when the principals hire the agents to perform some services on the principal's behalf. Principals commonly delegate decision-making authority to the agents.

B. Stewardship Theory

Stewardship theory is a theory that managers, left on their own, will act as responsible stewards of the assets and resources they control (Donaldson and Davis, 1991). Stewardship theorists assume that given a choice between self-serving

behavior and pro-organizational behavior, a steward will place higher value on cooperation than defection.

C. Corporate Governance

Shareholders not directly involved in management lack information about the concealed actions of executives, leading to less-than-deserved returns. This indicates the company's low corporate governance. The Securities and Exchange Commission (SEC) defines good corporate governance as efficient, transparent, verifiable, and considering the interests of all parties. Institutional ownership is deemed significant for the company's operations. The SEC categorizes investors into four groups: (1) institutional investors, (2) special large-scale investors, (3) large-scale investors, and (4) small-scale investors. Groups (2) and (3) must meet specific financial criteria, such as having a net worth of not less than 50 million baht, an annual income of not less than 4 million baht, or direct investments in securities or futures contracts of not less than 10 million baht, to be capable of assuming risk and accessing complex investment products.

D. Review of Related Literatures

According to research by Narisa Yaowalak and Silaporn Silpan, 2015, the Stock Exchange of Thailand promotes the involvement of retail investors and supports the increased role of institutional investors. Institutional investors are major shareholders, which can influence corporate governance and use their power to oversee the management's performance, potentially adding value to the business (Shleifer & Vishny, 1986). Previous studies concluded that good corporate governance could reduce the role of agency costs.

Therefore, this study aims to examine the direct relationship between CEO tenure and company performance, as well as study the indirect relationship using institutional ownership as a moderator variable.

Proposed conceptual model

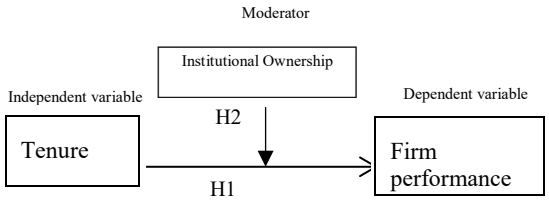


Figure 2.1 Conceptual Framework of the Relationship between Corporate governance and firm performance: the Effect of CEO Entrenchment

Objectives

1. *To test the relationship between the CEO tenure and the firm performance.*
2. *To test the moderator influence on the relationship between CEO tenure and firm performance*

Hypotheses

a) *H1: CEO tenure is related to the firm performance.*

b) *H2 Institutional ownership moderates the relationship between CEO tenure and the firm performance*

III. Research Method

This research is limited to the study of specific companies registered in the Stock Exchange of Thailand (SET) from 2017 to 2021. The selected group for this study does not include companies within financial institutions registered in the stock market. The research will employ a purposive sampling method. All companies selected from the population must not be undergoing liquidation or business recovery. Furthermore, the chosen companies must disclose complete data in their annual reports to the public, pertaining to the variables of interest and within the designated time frame.

Data collection for this research is based on secondary data, with primary officer data being collected from the SET Market Analysis and Reporting Tool (SETSMART) and the companies' annual reports. Operational and financial performance data can be gathered from the financial statements of the companies (Form 56-1) and the SET Market Analysis and Reporting Tool (SETSMART).

This research involves quantitative data analysis using statistical techniques. The cross-sectional data collected from the secondary data will be used to test the relationships among variables using the Ordinary Least Squares method. The analysis will consist of descriptive statistics, while the testing of the hypotheses will be conducted using multiple regression analysis. The testing of the moderator effect, which examines the impact of institutional ownership on the relationship between CEO tenure and firm performance, will involve the analysis of interaction terms from the multiple regression analysis.

This research aims to investigate companies registered on the Stock Exchange of Thailand (SET) from 2017 to 2021. The selected group for this research excludes financial institutions listed on the stock market. The study employs purposive sampling to select a sample of companies, ensuring they are not undergoing asset foreclosure or business recovery. Moreover, the chosen companies for analysis must have disclosed complete data in their annual reports for the specified period. The research employs a sample size of 1640 companies, utilizing secondary data collected from the SET Market Analysis and Reporting Tool (SETSMART) database.

The research employs descriptive statistics and tests the relationships using Ordinary Least Square (OLS) regression analysis. The statistical analysis is conducted using the Statistical Package for the Social Science (SPSS) software.

This study utilizes financial and operational data gathered from Form 56-1 and annual reports of companies between the years 2017 and 2021. The primary goal is to provide a quantitative

analysis of the relationships among variables, contributing to the body of knowledge in the field of management.

The research design includes a detailed methodology encompassing the scope, sampling process, data collection, analysis techniques, and ethical considerations. The outcome is expected to contribute valuable insights into the factors influencing the performance of companies listed on the Thai stock market during the specified period.

$$ROE_{it} = \beta_0 + \beta_1 CEOTenure_{it} + \beta_2 MBV_{it} + \beta_3 SIZE_{it} + \beta_4 LEV_{it} + \beta_5 INDDummy_{it} + \beta_6 YearDummy_{it} + \varepsilon_{it} \quad (1)$$

$$ROE_{it} = \beta_0 + \beta_1 CEOTenure_{it} + \beta_2 IO_{it} + \beta_3 CEOTenure_{it} * IO_{it} + \beta_4 MBV_{it} + \beta_5 SIZE_{it} + \beta_6 LEV_{it} + \beta_7 INDDummy_{it} + \beta_8 YearDummy_{it} + \varepsilon_{it} \quad (2)$$

Indendent variables

Tenure $_{it}$ = *Duration of the Chief Executive Officer's tenure*

Dependent variables

ROE$_{it}$ = *The measure of a company's net income divided by its shareholders' equity*

Moderator

IO = *Measured by the percentage of ordinary shares held by institutional investors*

IO * Tenure$_{it}$ = *Percentage of Institutional Investors' Shareholding * Duration of the CEO tenure*

Control variables

MBV$_{it}$ = *Market value to book value*

SIZE$_{it}$ = *Company Size*

LEV$_{it}$ = *Debt to equity ratio*

Variable measurement

Tenure$_{it}$ = Duration of the CEO tenure Measured by the length of time the Chief Executive Officer has been in office. (Guo, et al., 2015; and Meijer, 2017)

ROE$_{it}$ = Investment Performance It is used to measure the performance of accounting business.
Investment performance is calculated from net profit / shareholders' equity.

IO$_{it}$ = institutional ownership is measured by the percentage of ordinary shares held by institutional investors, represented by a value of 1 level greater than the median. Replace the value with 0.

MBV$_{it}$ = Market value to book value
Market value to book value is calculated from Market price/Book value.

Size$_{it}$ = Size of business
The size of an entity is measured by the logarithm of an asset as measured by the market capitalization of the company.

LEVit = Debt to equity ratio
Debt-to-equity ratio is calculated as Total Debt/Total Equity (Giroux, 2008).

INDDummy it = Industry is a dummy variable

YearDummy it = TIME is a dummy variable

IV. RESEARCH RESULTS

Table 1: Shows Descriptive Statistic.

Variable	Amount	Minimum value	Maximum value	Average	Standard deviation
Tenure	1640	.01	45.89	10.7136	9.34093
ROE	1640	-95.67	31.97	6.5558	11.79724
MBV	1640	.12	47.44	1.9102	2.38371
Size	1640	26.89	40.26	31.8815	2.41223
Lev	1640	.02	19.13	1.0011	1.18032
IO	1640	.01	84.57	16.0751	17.92673

According to the results of the descriptive statistics of the variables, the ROE variable has a mean of 6.5558, a maximum value of 31.97 and a minimum value of -95.67. CEO tenure has a mean of 10.7136, a high value of 45.89, and a minimum value of 0.01. Institutional ownership (IO) has a mean value of 16.0751, a high value of 84.57 and a minimum value of 0.01. Market Value to Book Value (MBV) has a mean value of 1.9102, a maximum value of 47.44, and a minimum value of 0.12. Size is measured using the logarithm of an asset measured by market capitalization. It has a mean of 31.8815, a high of 40.26, and a minimum of 26.89. The debt to total assets (LEV) ratio has a mean value of 1.0011, a maximum value of 19.13, and a minimum value of 0.02.

Table 2 shows the correlation coefficient of variable

	Tenure	ROE	MBV	Size	Lev	IO
Tenure	1	.053*	.053*	-.049*	.018	-.077**
ROE	.053*	1	.163**	.395**	-.220**	.181**
MBV	.053*	.163**	1	.334**	.137**	.067**
Size	-.049*	.395**	.334**	1	.122**	.459**
Lev	.018	-.220**	.137**	.122**	1	-.001
IO	-.077**	.181**	.067**	.459**	-.001	1

** represents significant statistical values that 0.01
* represents significant statistical values that 0.05

The result from Table 2 shows the relationship of correlation coefficients between variables. It was found that

the level of the relationship was not very high. The correlation coefficient of all variables is less than 0.8, indicating that initially the relationship between variables was not have a multicollinearity problem (Stevens, 1992).

The results showed a negative correlation between the CEO tenure and institutional ownership at a significant level of 0.01 and with Size at a significant level of 0.05.

Table 3 shows the results of the multiple regression analysis of correlations. Model 1 is to test the hypothesis 1 (H1): The CEO tenure is related to the firm performance. The results showed that the CEO tenure has a positive correlation with the operating results (ROE) at a significant level of 0.01, indicating that the longer the Chief Executive Officer is in office, the better the firm performance. For the control variables, it was found that the market value to book value (MBV), the size of the company was positively correlated with the performance (ROE) at a significant level of 0.01, the debt to total assets ratio (LEV) was negatively correlated with the performance (ROE) at a significant level of 0.01, and the Durbin-Watson value was 1.995, indicating that there was no autocorrelation problem.

Table 3 shows the Multiple Regression Analysis

Test the model 1

$$ROE_{it} = \beta_0 + \beta_1 CEOTenure_{it} + \beta_2 MBV_{it} + \beta_3 SIZE_{it} + \beta_4 LEV_{it} + \beta_5 INDDummy_{it} + \beta_6 YearDummy_{it} + \varepsilon_{it}$$

Variables	Coefficient	t-value	VIF
Intercept		.601	
Tenure	0.075**	3.461	1.008
MBV	0.054**	2.334	1.165
Size	0.40*	17.631	1.158
LEV	-0.276**	-12.547	1.036
Industry Dummies	Yes		
Year Dummies	Yes		
Adjust R^2	0.237		
Durbin-Watson	1.995		
N	1640		

** *Indicates a significant statistic value of 0.01*
* *Indicates a significant statistic value of 0.05*

A. Table 4 shows the Multiple Regression Analysis.

of the model. Model 2

$$ROE_{it} = \beta_0 + \beta_1 CEOTenure_{it} + \beta_2 IO_{it} + \beta_3 CEOTenure_{it} * IO_{it} + \beta_4 MBV_{it} + \beta_5 SIZE_{it} + \beta_6 LEV_{it} + \beta_7 INDDummy_{it} + \beta_8 YearDummy_{it} + \varepsilon_{it}$$

Variables	Coefficient	t-value	VIF
Intercept		.543	
Tenure	0.114**	3.856	1.883
MBV	0.054*	2.319	1.172
Size	0.409**	15.596	1.479
LEV	-0.275**	-12.507	1.041
IO	.040	1.200	1.932
IOTenure	-0.070*	-1.952	1.954
Industry Dummies	Yes		
Year Dummies	Yes		
Adjust R2	0.238		
Durbin-Watson	1.991		

** *Indicates a significant statistic value of 0.01*
* *Indicates a significant statistic value of 0.05*

Table 4 shows the results of the multiple regression analysis of correlations. This table was to test the hypothesis 2 (H2): The level of institutional ownership affects the relationship between CEO tenure and the firm performance. The results of the hypothetical test showed that CEO tenure has a relationship with firm performance (ROE) at a significant level of 0.01. It indicates that the longer the Chief Executive Officer holds the position, the more he will use his experience to have a positive impact on the firm performance. For control variables, it was found that market value to book value (MBV) and size of the company were positively correlated with the firm performance at a significant level of 0.0. 1 The debt to total asset ratio (LEV) was negatively correlated with the firm performance at a significant level of 0.01. The Durbin-Watson value is 1.991, thus it indicates no autocorrelation problem. When the Adjusted R^2 value is considered, it is found that the independent variable can describe the company's risk indicator (Risk) of 23.8%.

V. Conclusion

The objectives of this research were 1) to study the relationship between CEO tenure and the firm performance. The results of Table 3 research showed that the CEO tenure was positively correlated with the performance at a significant level of 0.01 with a coefficient of 0.075. When the business is successful, it will make the management socially acceptable. The second

Identify applicable funding agency here. If none, delete this text box.

objective is to study the impact of institutional ownership on the relationship between the CEO tenure and the firm performance. Table 4 shows that the interaction between institutional ownership and CEO tenure has a negative influence on the firm performance at a significant level of 0.05. The results can be concluded that institutional ownership acts as a moderator variable. It is a variable that results in reversing the relationship between the primary variable (the period of the CEO's tenure) and the dependent variable (the firm performance). Therefore, it shows that when the Chief Executive Officer is in office for a long time, he may try to avoid scrutiny and consider his own interests. Worsening operations Suggestions for next research the test examined the impact of institutional ownership on the relationship between the CEO tenure and the firm performance. The results show a positive correlation in which the firm performance tested using ROE is significant, indicating that the Company's corporate governance may allow the Chief Executive Officer to exercise his or her authority. Therefore, corporate governance should be developed to be more consistent and comprehensive, which will improve the Company's operating results.

REFERENCES

[1] Donaldson, L. and Davis, J., Stewardship theory or agency theory. Australian Journal of Management, Vol.16, 1991, pp.49-64.

[2] Jensen, M. C., and Meckling, W. H., Theory of the firm: Managerial behavior, agency costs and ownership structure. *Journal of financial Economics*, 3(4), 1976, pp. 305-360.

[3] Narisa Yaowalak and Silaporn Silpan, Insider ownership and institutional investor ownership on firm performance and stock return,. *Journal of Business Administration The Association of Private Higher Education Institutions of Thailand*, *4*(2),2015, pp.18–31. Retrieved from https://so02.tci-thaijo.org/index.php/apheitvu/article/view/95213

[4] Shleifer, A. and Vishny, R. W. Large shareholders and corporate control. The Journal of Political Economy. 1986, pp.461-488.

A Study on the Optimization of LED Light Source and Nutrient solution in Plant Factory Model Using Hydroponic Cultivation Technology

1st Hyunjin Chun
Arts college
Nanjing University of Aeronautics and Astronautics
China, Nanjing
tough4324@naver.com

2nd Jiahao Xu
Arts college
Nanjing University of Aeronautics and Astronautics
China, Nanjing
jiahao8868@126.com

3rd Xiaonan Zhan
Faculty of Arts
Jiangsu University
China, Nanjing
2903548972@qq.com

***Abstract*—Recently, environmental pollution has caused many problems in agricultural production. To solve this problem, an agricultural production method called a plant factory is used. Plant factories will be combined with various new technologies to lead the future agricultural revolution. In particular, the problem of labor shortage can be solved through automation of plant factories in relation to the Fourth Industrial Revolution. Hydroponic cultivation is widely used in current plant factories. When using hydroponics in plant factories, the most important factors for plant growth are light and nutrients. Therefore, this paper studied light sources and nutrients to develop an optimization model for plant factories. This study presented a plan for the optimal light source and nutrient solution for plant factories. The results of this study are as follows. There is a difference in plant growth depending on the LED light intensity. In plant growth, blue light and red light affect plant growth. Based on these results, LED light suitable for plant growth will shorten the period of plant growth. There is a difference in plant growth depending on the ions of the nutrient solution. These results show that the rate of plant growth increases as the EC of the nutrient solution increases. However, the EC of the nutrient solution of 3.0 dS·m-1 appears to be poor in the quality of the plants of the EC of the other nutrient solution. Therefore, in consideration of plant quality and plant growth in plant factories, 2.0dS·m-1~2.5dS·m-1 is the most appropriate EC. The results of this study can be theoretical basic data to actually install a plant factory.**

Keywords—Plant Factory, Hydroponic Cultivation, Artificial intelligence, LED Light, Nutrient solution

I. INTRODUCTION

A. Research Background and Purpose

In the era of the 4th Industrial Revolution, automation takes place in various fields. In particular, as it becomes an aging society, the need for automation increases in the agricultural sector, which requires a lot of labor. Through this automation, the problem of labor shortage in the agricultural sector is solved. Recently, there is a problem with agricultural productivity due to climate change. A representative system for automation in the agricultural sector is a plant factory through artificial intelligence and hydroponic technology. Plant factories, in particular, provide optimal conditions for cultivating crops by automatically controlling temperature, humidity, and light. And the appropriate environment is recognized as an important technology for plant factories. Therefore, research is needed on the optimization of the plant factory model using hydroponic cultivation technology. Therefore, this paper presents optimized light sources and nutrients used in plant factories. An optimized plant factory will greatly contribute to the stable production of plants. And the results of this study can be theoretical basic data when establishing a plant factory.

B. Literature Review

With the recent development of the Internet of Things, plant factory is widely used in the agricultural sector. In particular, plant factory is recognized worldwide as a complex of cutting-edge technologies using various new technologies. Therefore, many studies are conducted on plant factory in various universities and research institutes. In addition, the governments of each country are also increasingly interested in plant factory, supporting various research funds. Therefore, the research related to plant factory is as follows. Jung, Hyeon Ji et al. (2021) confirmed that data transmission and reception between Windows and Android, other operating system environments in smart farm [1]. Kwon Oh Hoon et al. (2021) confirmed that the automation function of smart farms can create an environment suitable for plants to grow [2]. Lee, sangwon et al. (2020) analyzed the the use of renewable energy and reduce the amount of fossil fuel used in agricultural activities in smart farms [3]. Kim, Han don et al. (2021) solved problems in the supply of food ingredients in the agricultural field by changing the existing growing environment [4]. Kim Chang-hum (2018) finded out the problem between service provider, customer, and product in smart house farm service design process [5]. Lim Gye Jae. (2015) researched the foundation of the future construction and operation of large plant factories [6]. Kim Ha-Neul et al. (2014) examined through an experiment the heating effects of a greenhouse by using energy discharged from air source heat pumps and greenhouses through ventilation or surplus solar energy, solar collectors with their storage of heat, and their

energy [7]. Choi Hok Il et al. (2019) proposes “A Smart Farm Environment Optimization and Yield Prediction Platform based on IoT and Deep Learning” which gathers bio-sensor data from farms, diagnoses the diseases of growing crops, and predicts the year's harvest [8]. Atsumasa Yoshida et al. (2019) installed sensors to observe the environmental factors in a plant factory and evaluated lettuce growth [9]. Chan-Il Uoon et al. (2017) selected suitable kohlrabi cultivar for hydroponics in a closed-type plant factory system [10]. Researchers in developed countries study a lot about plant factory. Hydroponic cultivation is widely used in plant factory. When using hydroponics in plant factory, the most important factors for plant growth are light and nutrients. However, in most studies related to plant factory, research on light sources and nutrients is practically difficult to use in plant factory. Therefore, this paper studied light sources and nutrients for models optimized for plant factory. This study aims to present a plan for the optimal light source and nutrient solution for plant factory models using hydroponic cultivation technology. The results of this study can be theoretical basic data to actually install a plant factory.

II. A Theoretical Study

A. Plant Factory

Recently, there are many problems with outdoor plant culture due to climate change and environmental pollution. To solve these problems, plant factory that can grow plants regardless of external environments such as climate or season are activated. A plant factory is a method of producing plants by applying a factory-type production system. plant factory first developed in research facilities, and plant factory developed when artificial light sources began to be used. Plant factory is a different concept from existing smart farm. Smart Farm is a plant production system that combines ICT technology with the agricultural sector as a higher concept of plant factory. However, plant factory produces plants through artificial light sources. The advantages of a plant factory are as follows. First, plant factory can increase plant production because it artificially adjusts the external environment. Second, because plant factory grows plants in their internal environment, it can grow plants in deserts or polar regions. For this reason, research on plant factory is also increasing. And the disadvantages of plant factory are as follows. First, plant factory uses a lot of energy to cultivate plants. For this reason, plant factory uses more energy. As a result, plant factory can increase environmental pollution. Second, plant factory costs a lot of construction and operation costs. So, plant factory can lose a lot of money (Figure 1).

Figure 1. Plant Factory

B. Light Source in a Plant Factory

A light source is needed for plant growth. The best source of light for plant growth is sunlight. However, artificial light sources have recently used for plant growth due to climate change and environmental pollution. Special techniques are needed to use artificial light sources for plant growth. In particular, artificial light sources are widely used in plant factory. Plant factory is separated into solar-powered plant factory and fully controlled plant factory according to the disconnection from the external environment. Solar-powered plant factory is mainly operated using solar power. On the contrary, a fully controlled plant factory cultivates plants in a space completely isolated from the natural environment. The role of artificial light sources is very important in fully controlled plant factory. Light sources used in plant factory include fluorescent lamps, high-pressure sodium lamps, and LED. Among these light sources, LED is most commonly used in plant factory. With the development of LED in the 1990s, the technology of plant factory developed rapidly. Before that, the efficiency of artificial light sources was very low. However, LED can save electricity bills compared to other light sources. And LED light sources can emit white, red, blue, green, etc. In addition, LED light sources emit less heat than light emission. So, it is very suitable for plant cultivation. For this reason, LED is recognized as a necessary light source for high-efficiency plant growth. Therefore, technology to control LED wavelengths according to plant growth is currently studied in many countries.

C. Nutrient solution in Plant Factory

Plant factory generally uses hydroponics. So, plant factory absorbs nutrients and oxygen from plants through nutrient solution. Plants need a variety of elements to grow. These elements are diverse, including carbon, hydrogen, phosphorus, manganese, and oxygen. Therefore, it is an important factor in plant growth in the amount of nutrient solution. If the concentration of nutrient solution is high, plants have excessive nutritional symptoms. When a plant develops excessive nutritional symptoms, the plant's leaves decay. In addition, if the concentration of nutrient solution is low, plants become malnourished, causing growth problems. It is the electrical conductivity (EC) that measures the value of the nutrient solution. If the concentration of nutrient solution increases, the electrical conductivity increases. And if the concentration of the nutrient solution decreases, the electrical conductivity decreases. As such, the concentration of nutrient solution is a very important factor in plant factory. Therefore, research is needed on how to maintain the proper concentration of nutrient solution.

D. Fully controlled plant factory using artificial intelligence

Plant factory generally uses ICT technology to provide an optimal environment for plant growth (Figure 2). In particular, artificial intelligence technology is combined with plant factory and uses as an automated system for growth conditions. It is also expected that plant factory will make new changes in agriculture by utilizing IOT and big data technology. The information through big data can provide a lot of information for plant growth to operators who operate plant factory. This information can improve the yield and harvest quality of plant factory. And using artificial intelligence, a fully controlled plant factory can

automatically control all the environments in which plants are needed. Various sensors are located in the fully controlled plant factory to check plant conditions [11]. Temperature sensors, humidity sensors, and illumination sensors are located in these sensors. After these sensors monitor the plant growth environment, they send information to computers and artificial intelligence systems. This information is analyzed by an artificial intelligence system to create an optimal environment. It is connected to a smartphone through Bluetooth based on optimal plant environment information created by computers and artificial intelligence systems. And managers can control temperature, humidity, and illumination in real time through smartphones even from a long distance [11].

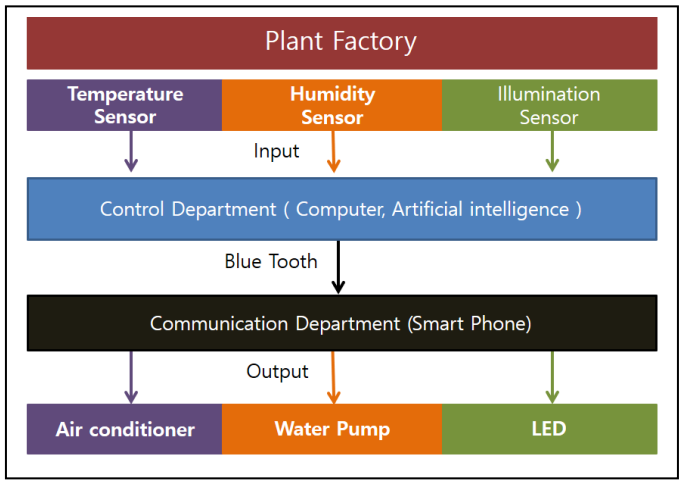


Figure 2. Plant Factory Control System

III. EXPERIMENTAL DESIGN AND METHOD

In this study, optimized light sources and nutrient solution for plant factory models is studied. First, LED is used as artificial light sources used in the experiment. Depending on the light quality of the LED, it affects the growth of the plant. Therefore, the purpose is to study the optimal illuminance for plant growth. In addition, plant factory provides nutrients through hydroponic cultivation without using soil. Therefore, in this study, the growth rate of plants is analyzed according to the concentration of nutrient solution. To carry out this experiment, a plant factory made of containers was targeted. The growth rate of plants was analyzed through changes in LED and nutrient solution. All growth factors other than LED and nutrient solution were the same. For the experiment of this study, the hardware configuration of the plant factory is as follows (Figure 2). The temperature inside the plant factory is 22 o C, and the humidity inside the plant factory is 60%.

A. LED

In this experiment, LED used blue light and red light, which are mainly used for plant growth [11]. Four cultivars are installed for the LED experiment. These four cultivation groups are A, B, C, and D. First, in this experiment, the seeds are grown in the darkroom under the same conditions without using an LED light source until they reach small sprouts [11]. After that, an LED light source consisting of 80% blue light and 20% red

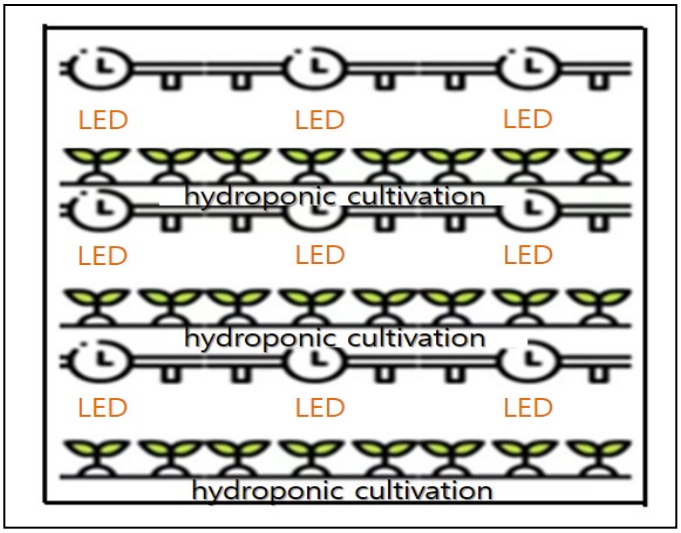


Figure 3. Plant Factory Hardware Composition

light is used in A cultivation group. In addition, LED light source consisting of 60% blue light and 40% red light is used in B cultivation group. In addition, LED light source consisting of 40% blue light and 60% red light is used in the C cultivation group. In addition, LED light sources consisting of 20% blue light and 80% red light is used in the D cultivation group. The length of the LED bar installed in each cultivation group is 50 cm. The LED was turned on from 9 am to 9 pm, and the darkroom was maintained for the rest of the time.

TABLE I. LED Installation Configuration

Cultivation group	Installation Ratio
A cultivation group	4: 1 (Blue : Red) Mix LED
B cultivation group	3: 2 (Blue : Red) Mix LED
C cultivation group	2: 3 (Blue : Red) Mix LED
D cultivation group	1: 4 (Blue : Red) Mix LED

B. Nutrient solution

Nutrient solution used in this experiment affects plant growth depending on the nutrient concentration. In general, the concentration of nutrient solution is cultivated in the range of 1.5-3.0. And if the concentration is too high, the leaves of the plant decay, and if the concentration is too low, the growth rate slows. Therefore, nutrient solution must be maintained at an appropriate concentration for plant growth [11]. In this experiment, a total of four cultivation groups are installed. These four cultivation groups are a, b, c, and d. The concentration of the nutrient solution in the a cultivation group was 1.5 dS·m-1. In addition, the concentration of the nutrient solution in the b cultivation group was 2.0 dS·m-1. In addition, the concentration of the nutrient solution in the c cultivation group was 2.5 dS·m-1. In addition, the concentration of the nutrient solution in the d cultivation group was 3.0 dS·m-1. Something else that can affect plant growth are all equal. Table 2. Changes in the concentration of nutrient solution

TABLE II. LED Installation Configuration

Cultivation group	Concentration of nutrient solution (EC)
a cultivation group	1.5 dS · m−1
b cultivation group	2.0 dS · m−1
c cultivation group	2.5 dS · m−1
d cultivation group	3.0 dS · m−1

IV. Experimental Design and Method Experimental Results

A. Comparison of plant growth according to LED illumination

This experiment compared plant growth according to LED illumination. For the LED experiment, a total of four cultivation groups were installed in this experiment. These four cultivation groups were A, B, C, and D. The results of the experiment analyzed the growth of plants every two days after germination of the seeds. On the 5th day of the A, B, C, and D cultivation, the size of the plant was all the same at 25mm. In A-cultivation group, the size of the plant was 33 mm on the 7th and 40 mm on the 9th. And the size of the plant in the A- cultivation group was 46 mm on the 11th, 53 mm on the 13th, and 59 mm on the 15th. In the B-cultivation group, the size of the plant was 32 mm on the 7th and 38 mm on the 9th. And in the B-cultivation group, the size of the plant was 43 mm on the 11th, 48 mm on the 13th, and 53 mm on the 15th. In the C-cultivation group, the size of the plant was 29 mm on the 7th and the size of the plant was 34 mm on the 9th. And in the C-cultivation group, the size of the plant was 38 mm on the 11th, 42 mm on the 13th, and 45 mm on the 15th. In the D-cultivation group, the size of the plant was 28 mm on the 7th and the size of the plant was 31 mm on the 9th. And in the D-cultivation group, the size of the plant was 34 mm on the 11th, 37 mm on the 13th, and 39 mm on the 15th. The growth rate of plants was the fastest in A-cultivation group, followed by B cultivation group and C-cultivation group. Plant growth was the slowest in the D-cultivation group.

TABLE III. Changes in plant growth according to LED illumination

Cultivation group	5days	7 days	9 days	11 days	13 days	15 days
A Cultivation group	25mm	33 mm	40 mm	46 mm	53 mm	59 mm
B Cultivation group	25 mm	32 mm	38 mm	43 mm	48 mm	53 mm
C Cultivation group	25 mm	29 mm	34 mm	38 mm	42 mm	45 mm
D Cultivation group	25 mm	28 mm	31 mm	34 mm	37 mm	39 mm

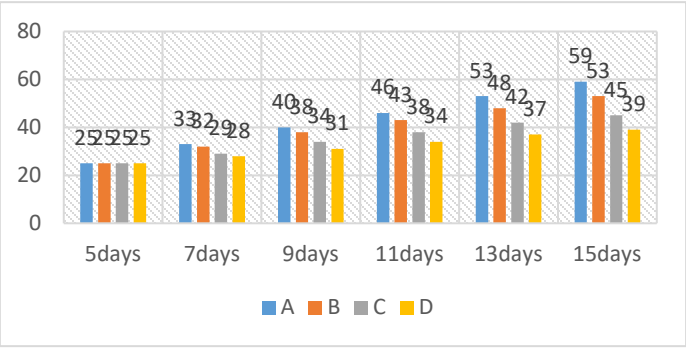


Fig. 4. Plant growth according to LED illumination

B. Comparison of plant growth according to the ions of nutrient solution

This experiment compared plant growth according to the ions of the nutrient solution. For the nutrient solution experiment, a total of four cultivation groups were installed in this experiment. These four cultivation groups were a, b, c, and d. The results of the experiment analyzed the growth of plants every two days after germination of the seeds. On the 5th day of the a, b, c, and d cultivation group, the size of the plant was all the same at 25mm. In the a-cultivation group, the size of the plant was 28 mm on the 7th and the size of the plant was 31 mm on the 9 mm on the 9th. And the size of the plant in the a-cultivation group was 34 mm on the 11th, 37 mm on the 13th, and 40 mm on the 15th. In the b-cultivation group, the plant size was 29 mm on the 7th and the plant size was 34 mm on the 9th. And in the b-cultivation group, the size of the plant was 38 mm on the 11th, 43 mm on the 13th, and 49 mm on the 15th. In the c-cultivation group, the plant size was 30 mm on 7 days and the plant size was 35 mm on 9 days. And in the c-cultivation group, the size of the plant was 40 mm on the 11th, 44 mm on the 13th, and 51 mm on the 15th. In the d-cultivation group, the plant size was 31 mm on the 7th and 36 mm on the 9th. And in the d-cultivation group, the size of the plant was 41 mm on the 11th, 46 mm on the 13th, and 52 mm on the 15th. The growth rate of plants was the slowest in a-cultivation group, followed by b-cultivation group and b-cultivation group. The rate of plant growth was the fastest in the d-cultivation group.

TABLE IV. Changes in plant growth according to the ions of nutrient solution

Nutrient solution	5 days	7 days	9 days	11 days	13 days	15 days
a-nutrient solution	25mm	28 mm	31 mm	34 mm	37 mm	40 mm
b-nutrient solution	25 mm	29mm	34 mm	38 mm	43 mm	49 mm
c-nutrient solution	25 mm	30mm	35 mm	40 mm	44 mm	51 mm
d-nutrient solution	25 mm	31mm	36mm	41mm	46 mm	52 mm

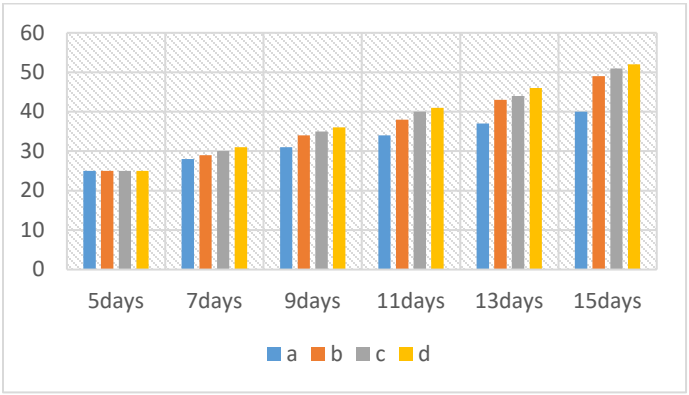


Fig. 5. Plant growth according to EC in Nutrient Solution

V. DISCUSSION

A. Plant growth according to LED illumination

Size of plant growth varies depending on the LED illumination. The 4:1 mixed LED of blue light and red light grew about 20mm faster after 15 days than the 1:4 mixed LED of blue light and red light. In addition, the 3:2 mixed LED of blue light and red light and the 2:4 mixed LED of blue light and red light showed similar plant growth sizes. In plant growth, blue light and red light affect plant growth. The results of this study showed that blue light promotes plant growth rather than red light. Therefore, if the exact LED light intensity is determined based on these results, the period of plant growth can be shortened. And in the future, the optimal light intensity can be determined by studying the LED mixing of red and blue light. In addition, the optimal distance between LEDs and plants should be analyzed to determine the most appropriate separation distance between plant and LED when installing LED in a plant factory.

B. Plant growth according to EC of nutrient solution

There is a difference in the size of plant growth depending on the ions of the nutrient solution. EC of 3.0 dS·m-1 nutrient solution and EC of 1.5 dS·m-1 nutrient solution grew about 12mm fast. In addition, EC of 2.0 dS·m-1 nutrient solution and EC of 2.5 dS·m-1 nutrient solution showed similar plant growth sizes. These results show that the plant's growth rate increases as the EC of the nutrient solution increases. However, the EC of a nutrient solution of 3.0 dS·m-1 is inferior to that of other nutrients. Therefore, 2.0 to 2.5 dS·m-1 is most appropriate in consideration of plant quality and plant growth when producing plants as hydroponic cultivation in a plant factory. In the future, research should be conducted with other factors affecting hydroponic cultivation and the most reasonable amount of nutrient solution should be presented [12].

VI. CONCLUSION

With the Fourth Industrial Revolution, it is becoming automation in all fields. Through this automation, the problem of labor shortage in the agricultural sector can be solved. A representative system for automation in the agricultural sector is a plant factory through artificial intelligence and hydroponic technology. Therefore, this paper presents optimized light sources and nutrient solution used in plant factory. The results of this study are as follows. Blue light promotes plant growth rather than red light. Therefore, if the exact LED illumination is determined based on these results, the period of plant growth can be shortened. In addition, as the EC of the nutrient solution increases, the speed of plant growth increases. However, the EC of a nutrient solution of 3.0 dS·m-1 is inferior to that of other nutrients. Therefore, when producing plants in a plant factory, a nutrient solution of 2.0 to 2.5 dS·m-1 is most appropriate in consideration of plant quality and plant growth. The optimized plant factory is expected to contribute greatly to the stable production of plants. And the results of this study can be theoretical basic data when establishing a plant factory.

REFERENCES

[1] Jeong Hyeon Ji and Wan-Bum Lee. 2021 A study on the implementation of smart farm environment control system using unity and photon. Journal of the Semiconductor & Display Technology. Vol 20 No. 1 pp 104-107.

[2] Kwon Oh-Hoon, Wook Park Yong, Im He-Beom and Min Dong-Sun. 2021 A study on the smart farm characteristics using multiple sensors. Journal of the KIECS. Vol 16 No. 4 pp 719-724.

[3] Lee sangwon and Cho Young-sang. 2020 Economic feasibility analysis of the renewable energy based business model in the agricultural sector and policy implications - focusing on the “Smart Farms” using renewable energy -. INNOVATION STUDIES. Vol 15 No. 1 pp 1-28.

[4] Kim Handon1, Lee Jeonga, Choi Seun, Jang Hyounseung and Kim Jimin. 2021 Development of building system for achieving an optimal growth environment in a vertical smart farm. Korea Journal of Construction Engineering and Management. Vol 22 No. 4 pp 3-10.

[5] Kim Chang-hum. 2018 A study on the revitalization of smart farm using service design –focused on the plant cultivation system for urban agriculture - master's thesis at technology and energy, Korea Polytechnic University.

[6] Lim Gye Jae. 2015 A design of IT-Convergence plant factory system using the renewable energy. Information Processing in Agriculture. Vol 40 No. 4 pp 769-779.

[7] Tae Kim Hyeon, Young bong Min and Yong Cheol Yoon. 2014 Thermal storage and heating effect by using heat pump and SurplusSolar energy in greenhouse. Journal of Agriculture & Life Science 48(6) pp.411-423.

[8] Hokil Choi, Ahn Heuihak, La-Na Jung and Lee ByungKwan. 2019 A smart farm environment optimization and yield prediction platform based on IoT and deep learning. Journal of Korea Institute of Information, Electronics, and Communication Technology. Vol 12 No. 6 pp 672-680.

Transformative Learning:
An Exploration of AI-Driven Tools in Workplace Education

Saloni Arora
Department of Psychology, CHRIST University, India

Dr. Guydeuk Yeon*
Centre for Digital Innovation, CHRIST University, India
**Corresponding author: yeon@christuniversity.in*

Abstract
As the dawn of the fifth industrial revolution is upon us, innovative tools and systems, including but not limited to artificial intelligence (AI), quantum computing, the Internet of Things (IoT), and cloud computing, and their structured usage, have become themes of immense topicality. These conspicuous technologies have been forging their trajectories into every domain connected to life on our planet. Education and learning, combining to form one such domain, have witnessed an abundance of technology-driven interventions in the past decade. Furthermore, education and learning have become critical components of every stage of life including work. As such, technologically-enhanced tools, more specifically, AI-based applications with the capabilities of providing customized training and learning content to employees at their convenience, have provided relief to human resources professionals across the globe. This paper aims to conduct a review of the available literature pertaining to the existing AI-driven tools and applications that are being used to provide customized educational content in the workplace. Moreover, this paper hopes to suggest potential research directions to researchers working in the field of AI-driven education and also provide guidance to employers worldwide regarding the implementation of AI-based educational tools at their respective places of work.

Keywords: *Artificial Intelligence, Learning, Education, Human Resources, Workplace.*

Introduction

Hurtling through a wave of continuous technological advances, the world is currently witnessing the unfolding of the fifth industrial revolution (5IR) (Pereira, 2021) while being on the cusp of a sixth one (Nardo, 2021). The impact of the fifth industrial revolution is marked by the rise in the topicality of such complex concepts, instruments, and tools as artificial intelligence (AI), machine learning (ML), quantum computing, cloud computing, Internet of Things (IoT), and their systematic and sustainable usage (Pang, 2023; Noble, 2022). Finding its understructure in the potentiality of congenial human-machine relationships, the broad focus of the 5IR is on the well-being of all stakeholders involved, including corporations, employees, customers, consumers, along with society at large (Noble, 2022).

Unlike the fourth industrial revolution that predominantly focused on competition and eliminating the human element from many industrial processes to increase efficiency, the fifth industrial revolution prioritizes a collaborative effort between humans and technology. The vision of 5IR is to encourage a synergistic interaction between machines and humans where the strengths of both parties may be exploited while the weaknesses are simultaneously compensated. For human

resources professionals and departments worldwide, the implications have been profound, with elements of digitalization having seeped through all the domains that comprise human resource development and management (Da Silva, 2022). Every day new forms of technology are emerging that are modifying the ways tasks are performed at work, with the fifth industrial revolution defining the manner in which people grow, learn, interact, and perform.

The fifth industrial revolution has impacted all existing industries and sectors globally. In the education and learning domain, too, the impact of 5IR has been far-reaching. As per Pang et al. (2023), the inclusion of technologies such as virtual and augmented reality in the curriculum, along with the introduction of personalized learning modules, would greatly benefit learners and academia. In recent years, a growing body of literature has focused on artificial intelligence, one of the key elements of the current, ongoing technological advancement (Ibarra, 2018).

While the literature on the application of AI for learning and education, especially at the workplace, is scarce as yet, a few studies have formed the ground by providing nascent insights into the possibility of AI-driven tools and technologies providing learning and educational opportunities for people at work (Rukadiakr, 2023; Tapalova, 2022). Moreover, the content delivered by AI-based technologies is highly learner-centric, personalized, and can be accessed on-demand.

Quite a few corporations have already started incorporating and delivering AI-based learning content to their employees, including Air Methods, Baidu, and International Business Machines [IBM], to name a few (Upadhyay, 2018). While not a huge number of companies currently provide AI-driven training and education to their employees, a number of corporations have emerged that are rendering AI-driven learning content to their customers by identifying their pain points and tackling them through personalized modules that are delivered on-demand (Bersin, 2023). The success of such firms as LinkedIn Learning, SeekOut, Degreed, and Eightfold, to name a few, attests to the significance of on-demand, personalized learning content in the current world where unpredictability and skill redundancy have combined to form the zeitgeist of the century.

Research Questions:

The primary aim of the current study is to explore the emerging themes in the application of artificial intelligence to render personalized, on-demand educational content to employees at work and its impact on employee outcomes. As such, the research question has been formulated as stated below:

What are the potential implications of the application of AI-based tools and technologies to deliver personalized, on-demand, and potentially, microlearning content to employees at the workplace?

A critical analysis of the literature has been carried out to answer the following emerging sub-questions:

1. What is the current state of the application of AI at the workplace when employees are concerned?
2. Why are AI-based tools required to deliver learning content at the workplace?
3. How are AI-driven technologies being utilized to render personalized training to employees?

Methodology:

This study involves a comprehensive examination of scholarly articles, encompassing a critical analysis and comparison of perspectives on the research questions. Relevant articles were sourced through access to prominent scientific search engines, including Scopus (recognized as a premier global database for scientific articles), Google Scholar (available at

www.scholar.google.com), and Scirus (accessible at www.scirus.com). The scientific citation index was employed to identify high-impact articles from peer-reviewed and influential journals, specifically focusing on AI and workplace learning research. This rigorous approach ensured the inclusion of authoritative and well-regarded sources in the exploration of the chosen subject matter.

Current State of AI-based Technologies at Work:

Artificial Intelligence is rapidly gaining popularity in today's technologically advancing world (Ibarra, 2018). The ability of AI to learn from data it collects and provide results based on that data is gathering interest from corporations.

Organizations are keen to adopt AI and ML-based software to determine if these technologies could help reduce operating costs, amplify profits, and boost organizational efficiency (Russell & Norvig, 2016). Corporations all over the globe are experiencing a digital evolution. However, there is growing apprehensiveness among employees about this digital transformation, and a major portion of the employees is averse to this change. This aversion can be attributed to the uncertainty regarding job security, suspicions regarding the capabilities of AI-powered software, and concerns regarding data privacy (Levy, 2018; Müller & Bostrom, 2016). As such, it has been accepted that this digital revolution will be a slow process as employees adapt to this change.

The benefits of adopting AI-based applications are undeniable. AI has found to not only reduce redundancy by automating certain tasks but has also been found to improve customer satisfaction rates (Bolton, 2018). In the hospitality industry, for example, AI is being used to assist customers with self-service and bill payment (Chung, 2018; Ivanov, 2017; Larivière, 2017). Large multinational organizations such as Hilton Hotels have been employing AI to provide necessary information to customers (Solomon, 2016).

Focusing on the current literature, however, a number of research studies show that employees fear losing their jobs because of the incoming of AI-based technology, and this finding holds true across industries and sectors such as healthcare, banking IT, hospitality, etc. (Abdullah, 2020; Ardon, 2020;). Moreover, these fears are not unfounded. A research study by McKinsey reports that 5% of jobs are expected to become obsolete due to AI (Manyika, 2017). Another study by Oxford University has estimated that nearly 47% of the jobs are expected to be taken over by AI-based software and humanoids by 2033 (Ramaswamy, 2017). For the finance industry, the odds are even worse with nearly 50% of the jobs expected to become automated (He, 2018). There is some respite yet, however, as it is expected that AI would only take over monotonous, low-level occupations, whereas jobs that require a high level of expertise would be augmented by AI and ML (Smith, 2017).

Nonetheless, the attitudes that employees possess regarding AI are going to be a defining factor in its adoption (Vasiljeva, 2021). Although negative attitudes toward AI are prevalent, positive attitudes exist as well (Lichtenthaler, 2020). Research has shown that employees are more accepting of AI-based applications if they perceive it to be beneficial or if they perceive it to bring about a boost in their performance. Amazon's Alexa and Apple's Siri have people much enthused about the prospect of AI.

The attitudes of employees can be looked through the lens of the technology acceptance model (Davis, 1989). The model proposes that a number of factors such as the perceived usability of the application influence attitudes towards it (Lichtenthaler, 2020). If the employees feel that the application is easy to learn and use, they are less likely to have negative attitudes toward it. However, poor experience with the application can exacerbate existing negative attitudes.

Another key factor is autonomy. If the employees perceive that they have free will while adopting new technologies, there is a higher likelihood of them having a positive attitude toward it.

There is also a sense of mistrust regarding external technology, and a growing field of research is dedicated to it (Lichtenthaler, 2006; Katz, 1982). A major portion of the employees also prefers to interact with a human instead of a machine, with the focus being on empathy and emotional intelligence, as demonstrated by the desire to converse with a human instead of a machine on the phone (Gartenberg, 2018). However, these attitudes are quick to flip as people would prefer to have a machine talk on their behalf while receiving a call.

Requirement of AI for Rendering Workplace Learning:

Unpredictability is the zeitgeist of the twenty-first century. The markets and the business world are becoming increasingly dominated by the elements of the VUCA world (Walia, 2021). VUCA, an acronym, stands for volatility, uncertainty, complexity, and ambiguity. The VUCA elements require all aspects of an organization to be highly adaptable, swift, and flexible (Gandhi, 2017).

The Human Resources personnel have a chief role in assisting organizations in adapting to the current VUCA world (Kinsinger, 2012). They need to be involved in a lot of self-directed learning, initiate the acquisition of new skills for themselves as well as others in the organization. However, traditional methods of classroom training that are currently prevalent in corporations are not suitable for keeping with the constant changes that the corporate world is subject to (Upadhyay, 2018).

Considering the fact that classroom training requires setting aside time by both the earner and the instructor, is unable to account for the individual differences of all learners, and may not be up-to-date with the ongoing trends, it cannot provide the competitive advantage a company requires to win the talent war that is continuously being waged in the corporate world. As such, AI-based technologies which have the capacity to identify the skills that a particular employee or a set of employees need, deliver personalized, on-demand, highly engaging, and microlearning content is much more effective, time and cost-minimizing than traditional classroom training methods (Upadhyay, 2018; OttoLearn, n.d.).

Further, psychological theories have provided bases for the applicability of artificial intelligence in workplace learning. For instance, Metacognitive theory posits that learners learn better when they are aware of how they learn, what concepts are inherently easy for them to grasp and what areas need special attention (Jaleel, 2016; Posner, 2017). Further, engaging in metacognition-based learning involves brainstorming and specifying strategies to enhance learning. As AI is capable of analyzing a large amount of data in a very short span of time, it can determine the areas where the learner needs to put in more effort, highlight these areas to the learner, and customize the learning material in such a way that the learner's weak areas become the focal point (Jaleel, 2016; Posner, 2017; Upadhyay, 2018).

The Utilization of AI for Delivering Personalized Training Content at the Workplace:

A number of corporations in a myriad of industries are incorporating AI-based applications into their training and development programs, irrespective of whether these programs are employee-centric or user-centric (Bersin, 2023; Davenport, 2001; Wang, 2009; Zhu, 1997). These firms range from big, multinational organizations to startups in the initial stages of development. When it comes to user-centric programs, a number of ed-tech firms are providing AI-based learning content (Bersin,

2023). For instance, LinkedIn Learning has taken a central position by utilizing AI to identify the learners' pain points, determine the methods that are expected to be most effective, and then finally implement neural networks to provide the precise micro-credentialing that is required (Bersin, 2023). Other companies that are providing a similar sort of learning content include Cornerstone, Docebo, Gloat, Eightfold, SeekOut, and Degreed.

SeekOut, for instance, allows the upper management or learning agents to get a microscopic view of the specific skill set of the employees, group them together based on their skills, and then analyze the groups to understand the training requirements (Bersin, 2023). SeekOut has been a pioneer in establishing team-based assessments. Similarly, Uplimit, an ed-tech startup providing technical courses, has used AI to build virtual training aides to facilitate learning for its users.

Companies with employee-centric training programs include IBM, Baidu, and Air Methods, to name a few (Upadhyay, 2018). Out of these companies, IBM has been the most successful in the utilization of AI-based applications for employee education as covered by the Massachusetts Institute of Technology [MIT] in a recent study (Qin, 2020). By investigating the learning systems present in IBM from 2014 to 2019, the research scientists at MIT found how the learning methods followed at IBM affected the employees' job performance and career prospects.

The learning systems at IBM are accentuated by a multitude of technologies, including cloud computing, artificial intelligence, and augmented reality (Qin, 2020). Specifically, AI chatbots are used to answer any questions or doubts users may have about the learning material, provide specialized recommendations to employees on what courses to take next, and use AI to segregate and tag the learning content available on IBM's platform Your Learning.

Another previously impossible achievement is the real-time diagnosis of issues emerging on the platform and the implementation of their solution (Qin, 2020). IBM uses AI to identify problems by analyzing thousands of comments in real time, and the solution is implemented immediately.

Discussion and Conclusions

The world is amidst the fifth industrial revolution (5IR), marked by the integration of advanced technologies such as artificial intelligence (AI), machine learning (ML), and the Internet of Things (IoT). Unlike its predecessor, the 5IR prioritizes collaborative efforts between humans and technology, aiming for a symbiotic relationship rather than eliminating the human element for efficiency.

According to the available literature, AI-based technologies emerge as more effective, offering personalized, on-demand, and microlearning content, especially in comparison with traditional classroom training methods. Corporations, ranging from multinational giants to startups, are incorporating AI-based applications into training and development programs. Ed-tech firms like LinkedIn Learning, Cornerstone, Docebo, and SeekOut use AI to identify learners' needs, provide effective methods, and implement micro-credentialing. Companies with employee-centric programs, such as IBM, Baidu, and Air Methods, leverage AI chatbots, cloud computing, and augmented reality for continuous improvement in learning systems.

The fifth industrial revolution and the advent of AI are reshaping industries, HR practices, and learning methodologies. The challenges posed by VUCA elements require organizations to embrace AI-driven solutions for efficient, personalized, and adaptive learning in a rapidly evolving world.

Future Scope

The future scope of research in the realm of artificial intelligence (AI)-based personalized workplace learning is extensive and holds significant potential for addressing critical aspects of employee development and organizational dynamics. First and foremost, there is a need to delve into the impact of AI-driven learning on job satisfaction and retention rates, exploring whether the personalized and on-demand nature of such training contributes to a more content and engaged workforce.

Ethical considerations in the implementation of AI, encompassing privacy, bias, and transparency, warrant thorough investigation to ensure the responsible integration of these technologies into workplace learning. Examining the adaptability of employees to AI-based training and its correlation with skill development is crucial for understanding the long-term effects on workforce capabilities. Furthermore, research should explore how the adoption of AI in workplace learning influences overall organizational performance and whether it translates into a competitive advantage for companies in the market.

References:

[1] Abdullah, R., & Fakieh, B. (2020). Health care employees' perceptions of the use of artificial intelligence applications: survey study. Journal of medical Internet research, 22(5), e17620.

[2] Bersin, J. (2023, May 30). EdTech Is Going Crazy For AI. joshbersin.com. https://joshbersin.com/2023/04/edtech-is-going-crazy-for-ai/

[3] Ardon, O., & Schmidt, R. L. (2020). Clinical laboratory employees' attitudes toward artificial intelligence. Laboratory Medicine, 51(6), 649-654.

[4] Bolton, C., Machova, V., Kovacova, M., & Valaskova, K. (2018). The power of human-machine collaboration: Artificial Intelligence, business automation, and the smart economy. Economics, Management, and Financial Markets, 13(4), 51–56. doi:10.22381/EMFM13420184

[5] Chung, M., Ko, E., Joung, H., & Kim, S. J. (2020). Chatbot e-service and customer satisfaction regarding luxury brands. Journal of Business Research, 117, 587-595.

[6] Da Silva, L. B. P., Soltovski, R., Pontes, J., Treinta, F. T., Leitão, P., Mosconi, E., ... & Yoshino, R. T. (2022). Human resources management 4.0: Literature review and trends. Computers & Industrial Engineering, 168, 108111.

[7] Davenport, T. H., Harris, J. G., De Long, D. W., & Jacobson, A. L. (2001). Data to knowledge to results: building an analytic capability. California management review, 43(2), 117-138.

[8] Davis, F. D. (1989). Technology acceptance model: TAM. Al-Suqri, MN, Al-Aufi, AS: Information Seeking Behavior and Technology Adoption, 205-219.

[9] Gandhi, L. (2017). Human resource challenges in VUCA and SMAC business environment. ASBM Journal of Management, 10(1), 1.

[10] Gartenberg, C. (2018, October 9). Google Duplex will start rolling out on Pixel devices next month. The Verge. https://www.theverge.com/2018/10/9/17955866/google-duplex-release-date-pixel-devices-next-month-november-assistant-calls

[11] He, D., & Guo, V. (4). ways AI will impact the financial job market. In World Economic Forum. Retrieved from https://www. weforum. org/agenda/2018/09/4-ways-ai-artificial-intelligence-impact-financial-job-market.

[12] Ibarra, D., Ganzarain, J., & Igartua, J. I. (2018). Business model innovation through Industry 4.0: A review. Procedia manufacturing, 22, 4-10.

[13] Ivanov, S. H., & Webster, C. (2017). Adoption of robots, artificial intelligence and service automation by travel, tourism and hospitality companies–a cost-benefit analysis. Artificial Intelligence and Service Automation by

Travel, Tourism and Hospitality Companies–A Cost-Benefit Analysis.
[14] Jaleel, S. (2016). A Study on the Metacognitive Awareness of Secondary School Students. Universal Journal of Educational Research, 4(1), 165-172.
[15] Kinsinger, P., & Walch, K. (2012). Living and leading in a VUCA world. Thunderbird University, 9.
[16] Katz, R., & Allen, T. J. (1982). Investigating the Not Invented Here (NIH) syndrome: A look at the performance, tenure, and communication patterns of 50 R & D Project Groups. R&d Management, 12(1), 7-20.
[17] Larivière, B., Bowen, D., Andreassen, T. W., Kunz, W., Sirianni, N. J., Voss, C., ... & De Keyser, A. (2017). "Service Encounter 2.0": An investigation into the roles of technology, employees and customers. Journal of business research, 79, 238-246.
[18] Levy, F. (2018). Computers and populism: artificial intelligence, jobs, and politics in the near term. Oxford Review of Economic Policy, 34(3), 393-417.
[19] Lichtenthaler, U. (2020). Extremes of acceptance: Employee attitudes toward artificial intelligence. Journal of business strategy, 41(5), 39-45.
[20] Lichtenthaler, U., & Ernst, H. (2006). Attitudes to externally organising knowledge management tasks: a review, reconsideration and extension of the NIH syndrome. R&D Management, 36(4), 367-386.
[21] Manyika, J., Lund, S., Chui, M., Bughin, J., Woetzel, J., Batra, P., ... & Sanghvi, S. (2017). Jobs lost, jobs gained: What the future of work will mean for jobs, skills, and wages.
[22]Müller, V. C., & Bostrom, N. (2016). Future progress in artificial intelligence: A survey of expert opinion. In Fundamental issues of artificial intelligence (pp. 555-572). Springer, Cham.
[23] Di Nardo, M., & Yu, H. (2021). Special issue "Industry 5.0: The prelude to the sixth industrial revolution". Applied System Innovation, 4(3), 45.
[24] Noble, S. M., Mende, M., Grewal, D., & Parasuraman, A. (2022). The Fifth Industrial Revolution: How harmonious human–machine collaboration is triggering a retail and service [r] evolution. Journal of Retailing, 98(2), 199-208.
[25] OttoLearn. (n.d.). Microlearning | Try OttoLearn's adaptive microlearning platform. https://www.ottolearn.com/microlearning
[26] Pang, T. Y., Lee, T. K., & Murshed, M. (2023). Towards a New Paradigm for Digital Health Training and Education in Australia: Exploring the Implication of the Fifth Industrial Revolution. Applied Sciences, 13(11), 6854.
[27] Ramaswamy, S. (2017). How companies are already using AI. Harvard Business Review. https://hbr.org/2017/04/how-companies-are-already-using-ai
[28] Rukadikar, A., & Khandelwal, K. (2023). Artificial intelligence integration in personalised learning for employee growth: a game-changing strategy. Strategic HR Review.
[29] Russell, S. J., & Norvig, P. (2016). Artificial intelligence: A modern approach. Malaysia: Pearson Education Limited.
[30] Qin, F., & Kochan, T. (2020). The Learning System at IBM: A Case Study.
[31] Smith, A., & Anderson, M. (2017). Automation in everyday life.
[32] Solomon, M. (2016). Technology invades hospitality industry: Hilton robot, Domino delivery droid, Ritz-Carlton mystique. Forbes.
[33] Tapalova, O., & Zhiyenbayeva, N. (2022). Artificial Intelligence in Education: AIEd for Personalised Learning Pathways. Electronic Journal of e-Learning, 20(5), 639-653.
[34] Upadhyay, A. K., & Khandelwal, K. (2019). Artificial intelligence-based training learning from application. Development and Learning in Organizations: An International Journal, 33(2), 20-23.
[35] Vasiljeva, T., Kreituss, I., & Lulle, I. (2021). Artificial Intelligence: The Attitude of the Public and Representatives of Various Industries.

Journal of Risk and Financial Management, 14(8), 339.
[36] Walia, A. (2021). Linking Business Strategy through HR in VUCA Times. Amity Management Review, 10.
[37] Wang, J., Gwebu, K., Shanker, M., & Troutt, M. D. (2009). An application of agent-based simulation to knowledge sharing. Decision Support Systems, 46(2), 532-541.
[38] Zhu, D., Prietula, M. J., & Hsu, W. L. (1997). When processes learn: Steps toward crafting an intelligent organization. Information Systems Research, 8(3), 302-317.

Simulating Sensor Data and Modeling Human Activity with Markov Chain for Energy Optimization in Apartments

Malik Muhammad Saad, Muhammad Ashar Tariq,
Seri Park, Dongkyun Kim*
School of Computer Science and Engineering, Kyungpook National University, Daegu, Republic of Korea
{maliksaad,tariqashar,psr0527,dongkyun}@knu.ac.kr

***Abstract*—This research investigates the application of virtual sensors in Node-RED and digital twin technology to simulate sensor data within apartment environments. The primary objective is to model human activity patterns using Markov Chains and analyze state transition probabilities to optimize energy consumption. The study leverages two types of virtual sensors, namely light monitoring and water tap detection, implemented through the open-source Node-RED platform. These sensors replicate changes in light intensity and analyze characteristic sound patterns associated with water usage, respectively. To create a dynamic representation of apartment settings, digital twin technology is integrated, allowing for emulation of diverse scenarios including occupancy levels, user behaviors, and different times of day. The digital twin model is linked to the virtual sensors, facilitating the generation of realistic sensor data based on the simulated environment. Data collection spans several weeks, capturing a wide range of scenarios and user interactions, forming the foundational dataset for subsequent Markov Chain modeling. The Markov Chain model is established based on observed transitions, where states represent varying configurations of human activity within the apartment. State transitions are determined by the simulated sensor data, accounting for factors such as light intensity changes and water usage patterns.**

***Index Terms*—Home Automation, Digital Twin, Node-Red, Human Activity**

I. Introduction

In an era marked by rapid technological advancement, the fusion of smart home automation and digital twin technology has emerged as a pivotal trend in the realm of residential energy optimization. This research, focused on simulating sensor data and modeling human activity patterns with Markov Chains for energy efficiency in apartments, delves into this cutting-edge intersection. By harnessing virtual sensors implemented through Node-RED and digital twin technology, this study aims to revolutionize the understanding of human activity within apartment environments, ultimately leading to unprecedented energy efficiency. In today's landscape, the integration of virtual sensors and digital twins into residential settings represents a significant stride towards enhanced energy conservation and resource utilization. Node-RED, an open-source flow-based programming platform, serves as the foundation for implementing virtual sensors [1]. These sensors, specializing in light monitoring and water tap detection, bring forth a dynamic means to emulate changes in room illumination and analyze water usage patterns. This technological synergy not only redefines our approach to data collection but also sets the stage for a comprehensive analysis of human behavior within the simulated apartment environment. Central to this research is the integration of digital twin technology, a concept gaining remarkable traction in the realm of smart homes. This technology encapsulates the creation of a dynamic, virtual representation of apartment environments, mimicking real-world scenarios encompassing different times of day, occupancy levels, and user behaviors. The seamless interplay between digital twins and virtual sensors facilitates the generation of authentic sensor data, mirroring the intricacies of human activity within the simulated environment. Through this innovative framework, we embark on a journey to unravel the intricate dance between human behavior and energy consumption, charting a course towards unparalleled optimization [2].

As data collection spans over several weeks, capturing an extensive array of scenarios and user interactions, the resulting dataset becomes the cornerstone for the subsequent development of a Markov Chain model. Markov Chains, revered for their capacity to model dynamic systems, serve as the bedrock for this research's analytical approach. States within the model encapsulate varying configurations of human activity, ranging from "lights on" or "lights off" to "water tap in use" or "water tap not in use." Transitions between these states are determined by the simulated sensor data, reflecting the nuanced interplay of light intensity changes and water usage patterns. By examining the observed transitions, we derive crucial state transition probabilities, which unveil the likelihood of different human activities occurring within the simulated apartment environment [3].

To this end, this work embarks on a pioneering journey at the intersection of virtual sensors, digital twin technology, and Markov Chain modeling. Through this integrated approach, we aim to revolutionize the understanding of human behavior within apartment environments, ultimately propelling us towards a future of unparalleled energy efficiency and resource optimization. This research not only underscores the transfor-

mative potential of technological convergence but also sets a precedent for the next frontier in residential sustainability and smart living. In the remainder of this paper section II presents the related work. Section III presents the system model. Section IV presents the Markov Chain model and results (i.e., state transition probabilities). Finally, the conclusion is drawn in section V and future work is also discussed.

II. Related work

With the increasing complexity of modern households and the desire to reduce energy consumption, there is a critical need to monitor, control, and optimize various home and industrial systems intelligently [4]. Digital twin technology enables the creation of virtual replicas of physical homes, allowing for real-time monitoring and analysis of human activities and energy usage. In this regard, authors in [5] explore cost-effective product lifecycles using Digital Twins highlighting IoT data management, offering case studies, and discussing scalability and future prospects. Another work presented in [6] explores the integration of Digital Twins and Blockchain, offering a concept for secure real-time asset monitoring. The authors propose a framework and design for merging digital twins with IoT-driven blockchain records, enabling secure, decentralized real-time supervision of physical assets and processes. Challenges and limitations in this integration are also discussed in addition to the future research directions. This study in [7] creates a Digital Twin framework with machine learning to predict smart home appliance functionality and enhance device lifespan, introducing a Deep Convolution Neural Network-based Logistic Regression Model with Digital Twins (DCNN-LR-DT) model for precise predictions.

Leveraging Markov modeling in smart homes aids in efficient human activity recognition through digital twins, reducing costs and energy consumption compared to sensor deployment. Machine learning models, essential for intelligent energy management, also benefit from Markov modeling integration, ensuring energy efficiency in smart homes [8], [9]. In this respect, authors in [10] introduce a Markov-chain-based model for user activity patterns, enabling energy forecasts in smart homes. Their prediction- and feedback-based proactive energy conservation (PF-PEC) algorithm ensures energy savings without compromising comfort. The proposed scheme is implemented within a fog-based IoT architecture making it a suitable solution for real-world scenario. In terms of energy consumption, [11] proposes a solution to manage sensor energy in smart homes. Employing recurrent neural network (RNN) and dynamic time warping (DTW), activities are predicted while conserving energy. Bidirectional long short-term memory (BLSTM) RNN forecasts future activities, and DTW selects a guard sensor for unusual activities, balancing energy and accuracy.

III. Methodology

Virtual sensors were implemented using Node-RED, an open-source flow-based programming platform. Two types of sensors were simulated: light monitoring and water tap detection, also shown in Fig. 1. For light monitoring, a virtual sensor node was created to generate random light intensity values, simulating changes in room illumination over time. These values can be described as follows:

$$L(t) = L_0 + \sum \Delta L(i) \quad (1)$$

where L_0 is the initial light intensity, $\Delta L(i)$ represents incremental changes in light intensity at each time step (i), and t represents time. For water tap detection, a virtual sensor node was designed to analyze audio input for characteristic sound patterns associated with water usage. To create a dynamic representation of apartment environments, digital twin technology will be integrated. This technology will allow for the emulation of various scenarios, such as different times of day, occupancy levels, and user behaviors. This technology will allow for the emulation of various scenarios, such as:

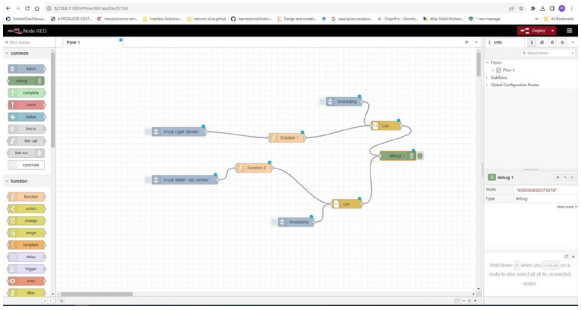


Fig. 1. Node-Red

- different times of day (T)
- occupancy levels (O), and
- user behaviors (B).

The digital twin model will be linked to the virtual sensors, enabling the generation of realistic sensor data (D) based on the simulated environment. This relationship can be expressed as follows:

$$D(T, O, B) = f(L(t), S) \quad (2)$$

Data collection will be conducted over a period of several weeks, capturing a wide range of scenarios and user interactions. In this work based on the virtual sensor and with varying light illumination conditions data is generated. This dataset served as the foundation for the subsequent Markov Chain modeling.

IV. Markov Chain Model

A Markov Chain is a mathematical model characterized by a set of states and transition probabilities between those states. In this research, states represent different configurations of human activity within the apartment environment. For instance, states include ”lights on” or ”lights off” and ”water tap in use” or ”water tap not in use”. Fig. 2 shows the states in the Markov Chain model and possible transitions.

The transition probabilities (P_{ij}) between different states (S_i, S_j) were estimated based on the observed data. For

example, if the light intensity increases beyond a certain threshold, a transition from ”lights off” to ”lights on” is recorded. Similarly, if the water tap detection sensor identifies the sound of running water, a transition from ”water tap not in use” to ”water tap in use” is registered. The Markov Chain model is constructed based on the observed transitions. Transition probabilities are calculated by dividing the number of observed transitions from one state to another by the total number of transitions from that state. These probabilities are crucial for understanding the likelihood of different human activities occurring within the apartment environment. The transition probabilities are defined in the next section.

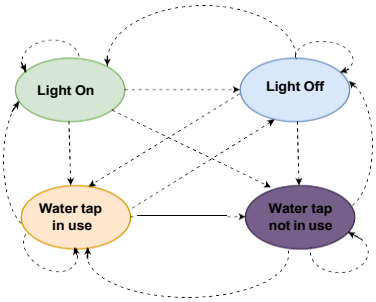


Fig. 2. Markov Chain States

A. Results

State transition probabilities quantify the likelihood of moving from one state to another in the Markov Chain model. These probabilities are calculated from the collected data and serve as a fundamental component in predicting and optimizing human behavior.For example, if there were 50 transitions from ”lights off” to ”lights on” out of a total of 100 transitions originating from ”lights off”, the state transition probability from ”lights off” to ”lights on” would be 0.5 or 50%. Similarly, if there were 30 transitions from ”water tap not in use” to ”water tap in use” out of 40 transitions originating from ”water tap not in use”, the state transition probability from ”water tap not in use” to ”water tap in use” would be 0.75 or 75%.

To accommodate environmental factors, Markov Chain state transition probabilities were analyzed for day and night. Table I shows the transition probabilities at the daytime and Table II shows the transition probabilities at the night time. Analyzing these state transition probabilities provides valuable insights into the patterns of human activity within the simulated apartment environment. These insights are essential for implementing energy optimization strategies, such as automated lighting controls and water conservation measures, based on predicted user behavior.

TABLE I
STATE TRANSITION PROBABILITIES AT DAY TIME

State	Next State	Transition Probability
Light On	Light On	0.2
Light On	Light Off	0.3
Light On	Water tap in use	0.3
Light On	Water tap not in use	0.2
Light Off	Light Off	0.3
Light Off	Light On	0.1
Light Off	Water tap in use	0.4
Light Off	Water tap not in use	0.2
Water tap in use	Water tap in use	0.2
Water tap in use	Light Off	0.3
Water tap in use	Light On	0.1
Water tap in use	Water tap not in use	0.4
Water tap not in use	Water tap not in use	0.3
Water tap not in use	Light On	0.1
Water tap not in use	Light Off	0.4
Water tap not in use	Water tap in use	0.2

TABLE II
STATE TRANSITION PROBABILITIES AT NIGHT TIME

State	Next State	Transition Probability
Light On	Light On	0.4
Light On	Light Off	0.3
Light On	Water tap in use	0.1
Light On	Water tap not in use	0.2
Light Off	Light Off	0.4
Light Off	Light On	0.1
Light Off	Water tap in use	0.1
Light Off	Water tap not in use	0.4
Water tap in use	Water tap in use	0.3
Water tap in use	Light Off	0.1
Water tap in use	Light On	0.3
Water tap in use	Water tap not in use	0.3
Water tap not in use	Water tap not in use	0.3
Water tap not in use	Light On	0.2
Water tap not in use	Light Off	0.3
Water tap not in use	Water tap in use	0.2

V. CONCLUSION AND FUTURE WORK

This research demonstrates the effectiveness of virtual sensors in conjunction with Node-RED for simulating sensor data in apartment environments. The integration of light monitoring and water tap detection sensors, implemented through Node-RED, provides a robust framework for replicating real-world scenarios. The resulting dataset derived, serves as a valuable foundation for constructing the Markov Chain model. In future, digital twin technology will be incorporated to further enhance the dynamic representation of apartment settings, allowing for the emulation of diverse occupancy levels and user behaviors. Moving forward, the focus will be on refining the integration of digital twin technology to achieve more accurate state transition probabilities, thus enhancing the optimization of energy consumption in real-world applications.

ACKNOWLEDGMENT

This work was supported by the Digital Innovation Hub project supervised by the Daegu Digital Innovation Promotion Agency(DIP) grant funded by the Korea government(MSIT and Daegu Metropolitan City) in 2023(No. DBSD1-04, Smart management system for preventing to lonely deaths of elderly people living alone based on automatic meter reading information and CCTV access information). In addition this study was

supported by the BK21 FOUR project (AI-driven Convergence Software Education Research Program)funded by the Ministry of Education, School of Computer Science and Engineering, Kyungpook National University, Korea (4199990214394).

REFERENCES

[1] S. Yoon, "Virtual sensing in intelligent buildings and digitalization," *Automation in Construction*, vol. 143, p. 104578, 2022.

[2] C.-C. Lee, J. Hussain, and Y. Chen, "The optimal behavior of renewable energy resources and government's energy consumption subsidy design from the perspective of green technology implementation," *Renewable Energy*, vol. 195, pp. 670–680, 2022.

[3] Z. Zhaoyun and L. Linjun, "Application status and prospects of digital twin technology in distribution grid," *Energy Reports*, vol. 8, pp. 14170–14182, 2022.

[4] D.-E. A. Mansour, M. Numair, A. S. Zalhaf, R. Ramadan, M. M. Darwish, Q. Huang, M. G. Hussien, and O. Abdel-Rahim, "Applications of iot and digital twin in electrical power systems: A comprehensive survey," *IET Generation, Transmission & Distribution*, 2023.

[5] A. Murthy, M. Irshad, S. M. Noman, X. Tang, B. Hu, S. Chen, and G. Khader, "Internet of things, a vision of digital twins and case studies," in *IoT and Spacecraft Informatics*, pp. 101–127, Elsevier, 2022.

[6] E. E.-D. Hemdan, W. El-Shafai, and A. Sayed, "Integrating digital twins with iot-based blockchain: Concept, architecture, challenges, and future scope," *Wireless Personal Communications*, pp. 1–24, 2023.

[7] V. Padmapriya and M. Srivenkatesh, "Digital twins for smart home gadget threat prediction using deep convolution neural network," *International Journal of Advanced Computer Science and Applications*, vol. 14, no. 2, 2023.

[8] J. Bai, "Markov model in home energy management system," in *Journal of Physics: Conference Series*, vol. 1871, p. 012043, IOP Publishing, 2021.

[9] A. Alzoubi, "Machine learning for intelligent energy consumption in smart homes," *International Journal of Computations, Information and Manufacturing (IJCIM)*, vol. 2, no. 1, 2022.

[10] M. Umair, M. A. Cheema, B. Afzal, and G. Shah, "Energy management of smart homes over fog-based iot architecture," *Sustainable Computing: Informatics and Systems*, vol. 39, p. 100898, 2023.

[11] M. Khan, J. Seo, and D. Kim, "Modeling of intelligent sensor duty cycling for smart home automation," *IEEE Transactions on Automation Science and Engineering*, vol. 19, no. 3, pp. 2412–2421, 2021.

Analyzing the Development of Intelligence in Non-player Characters in Digital Games

1st KaiXing Wang
Student, Dept. of Visual Contents
Dongseo University
Busan, Korea
wkx1213@gmail.com

2nd Kihong Kim
Professor, Visual Contents
Dongseo University
Busan, Korea
khkim@gdsu.dongseo.ac.kr

***Abstract*—With the continuous development of game technology, making players feel the authenticity of the game world has become an important goal of game design. In addition to graphics rendering technology, artificial intelligence technology plays a crucial role in achieving this goal. This article focuses on the impact of artificial intelligence technology on non-player characters (NPCs) in digital games and summarizes the significance and prospects of intelligent NPCs for games.**

The application of artificial intelligence technology makes NPCs become game objects capable of thinking and possessing intelligence in the game, thereby increasing the immersion of the game for players. This article describes the development of artificial intelligence technology for NPCs in digital games and summarizes suggestions for the intelligent development of NPCs.

First, this article introduces the definition of NPC, discusses the application of artificial intelligence technology in game development, and then analyzes the significance of NPC intelligence in games. Currently, in digital games, the intelligent performance of NPC is mainly concentrated in MMORPG games. Its main function is to guide players to complete the main tasks of the game, thereby improving the immersion of the game. However, the game itself is not only a way of entertainment but also has important historical communication, cultural inheritance, social education, and aesthetic functions. Therefore, future research should actively strive to expand the types of games that apply NPC intelligence so that they can play a more active role in spreading culture and increasing educational interest.

To sum up, this article briefly analyzes the intelligent development of non-player characters in digital games. Through the application of artificial intelligence technology, NPCs can think and be intelligent, thereby enhancing the authenticity and rationality of the game environment. With the continuous advancement of technology, the intelligence of NPC will be better developed in the future.

Keywords—artificial intelligence, non-player characters, intelli gent development

I. INTRODUCTION

With the continuous evolution of digital game technology, the intelligent development of non-player characters (NPCs) has become an issue of concern in the field of game design. In modern games, NPC is no longer a simple virtual character, but a virtual existence with thinking and wisdom, leading the game experience forward. The purpose of this article is to deeply explore the intelligent trend of NPCs in digital games, and analyze it from multiple perspectives such as definition, technology application, significance, and prospects. Through this comprehensive lens, we will gain a clearer understanding of the impact of AI technology on non-player characters and its profound impact on player immersion and gaming experience.

II. APPLICATION OF ARTIFICIAL INTELLIGENCE TECHNOLOGY IN DIGITAL GAMES

A. Definition of NPCs

This research focuses on exploring the development and importance of artificial intelligence technology to "NPCs" (Non-Playable Characters) in video games, virtual worlds, role-playing games, and other interactive media. The behavior of non-player characters (NPCs) is often crucial for the success of a game. We broadly define NPCs as visible components of the game that are under the control of the computer, and that either work with the human player or against him or her [1]. The emergence of the game term NPC did not originate from the context of video games but began to be used and popularized with the popularity of the tabletop role-playing game "Dungeons & Dragons" in the early 1870s. NPCs are virtual characters created and controlled by game developers or creators who usually interact with the characters played by the player in the game. They play a variety of roles and maybe quest givers, shop owners, hostile characters, friendly interactive objects, or simply background elements in the game world.

NPCs have a variety of functions, driving the game's storyline and providing quests and interactions that enhance the depth and variety of the game. Players can talk to NPCs, buy items, accept quests, unlock new areas, and more. NPCs in different games have their characteristics in appearance, personality, dialogue, and behavior to adapt to the theme and situation of the game.

However, traditional NPCs have become somewhat outdated and limited. They usually can only perform predetermined fixed

[1] A. Fink, J. Denzinger and J. Aycock, "Extracting NPC behavior from computer games using computer vision and machine learning techniques," 2007 IEEE Symposium on Computational Intelligence and Games, USA, 2007, pp. 24-31

This research was supported by BK21 Four Service Design driven Social Innovation Educational Research Team in Dongseo University.

actions and dialogues, and cannot provide a lively and interesting interactive experience. This limitation may lead to a decrease in game playability as players struggle to obtain a sense of freshness, which may affect the appeal of the game and limit its potential to attract more players. Therefore, traditional NPCs need to transform to meet the growing game market needs and player expectations.

Improvements in artificial intelligence technology play a key role in the intelligence and interactivity of NPCs. NPCs can better adapt to players' behaviors and situations, provide a more immersive and personalized gaming experience, break the limitations of traditional NPCs, and thus bring players a more challenging and attractive gaming experience. This is crucial to the future development of the gaming industry.

B. The development of artificial intelligence technology in the gaming field

In the domain of gaming, the continuous evolution of Artificial Intelligence (AI) technology has remained a subject of profound interest among scholars and game developers. Examining this from a technological historical perspective, early electronic games featured Non-Playable Characters (NPCs) typically constrained by scripted instructions, executing predefined paths, fixed behavioral patterns, and rudimentary responses. For instance, in the iconic arcade game "Pac-Man" (1980), game characters employed straightforward pursuit algorithms governed by predefined rules to track the player-controlled character.

As time progressed, advancements in game technology spurred the ongoing development of AI. Initially, NPCs were governed by the Finite State Machine (FSM) computational model. For instance, in "StarCraft" (1998), game units could transition their behaviors based on their current states (e.g., attacking, defending, patrolling); however, these states remained bound by predefined rules. This approach resulted in NPC behaviors that appeared relatively rigid and predictable.

Nonetheless, with the increasing complexity of games, FSMs began to demonstrate limitations. This prompted the introduction of Behavior Trees, which describe NPC behaviors through a hierarchical structure composed of nodes. Modern games like "Assassin's Creed" employ Behavior Trees encompassing diverse nodes such as "patrol," "pursue target," and "attack." These nodes can orchestrate complex behaviors based on various combinations of game context and player actions. The modular nature of Behavior Trees empowers developers to define and manage NPC behaviors with greater ease.

The latest trend involves integrating deep learning techniques into gaming, yielding AI endowed with neural network-driven capabilities. For instance, in the open-world game "Red Dead Redemption 2," NPCs exhibit more intelligent decision-making based on environmental cues, social interactions, and player actions. These NPCs demonstrate advanced emotional and cognitive responses, rendering the game world more vibrant. This developmental trend proves pivotal in enhancing the immersion and authenticity of gaming experiences.

In conclusion, AI technology in the realm of electronic gaming has continuously evolved, progressing from early script-driven NPC behavior to FSMs and, more recently, to modern Behavior Trees and deep learning. The introduction of these technologies has not only elevated the intelligence of game characters but has also enriched the interactivity and realism of game worlds.

III. THE EMERGENCE OF INTELLIGENT NPCS

A. Intelligent development of NPCs in digital games

Today, the development and use of ChatGPT, an important artificial intelligence with deep learning capabilities based on neural networks, in digital games has played a role in promoting the design of NPCs in online games. In January 2023, developers produced a game based on the game "Mount & Blade II: Banner Lord," this mod is for non-player characters in the game, using a custom story engine and ChatGPT. Without adding any given script, you only need to enter the background information of these NPCs and their identities in the game. When they interact with the players, they will generate answers in real-time based on any questions raised by the players. The NPCs in the game are powered by artificial intelligence, allowing them to adapt to the environment and display realistic behaviors. These behaviors make traditional NPCs unable to do so. NPCs will react according to the actions of the characters, and sometimes even remember the player's actions, affecting events and interactions in the game world. This greatly improves the playability of the game and the player's sense of immersion.

Fig. 1. Mount & Blade II: Banner Lord

Zhejiang Jinke Tom Culture Industry plans to iteratively launch game products such as "Talking Tom Cat" by integrating the IP "Talking Tom Cat" with ChatGPT technology. Under the influence of AI technology, Tom Cat can not only chat with players but also has its personality. For example, if a user asks it to help arrange the schedule, it will reply, "I have no time to care about you." The game company said that the technology of the functional prototype of the test product is not yet fully mature, and further optimization will be done after the closed beta test.

Fig. 2. Talking tom cat

On February 15, 2023, the mobile game "NiShuiHan" created by NetEase announced that it would install China's first game version of ChatGPT, allowing intelligent NPCs to freely generate conversations with players and independently give logical actions based on the conversation content. "NiShuiHan" is an MMORPG game with an ancient Chinese background. In the game, all dialogue texts, expressions, voices, and shots of intelligent NPCs are freely generated by AI, and the voices are temporary mechanical versions. Players interact with NPCs to change the relationship between players and NPCs, and even affect the relationship between NPCs. The technology used by NiShuiHan Intelligent NPC has the same origin as ChatGPT. It's just that ChatGPT has a wider learning scope and is more like a search engine; while the pre-training content of "NiShuiHan GPT" is mostly Chinese martial arts novels, history books, poems and songs, etc., and is constrained to the game background being the history of ancient China. This is to avoid situations where NPCs do not conform to the game background settings when communicating with players. This kind of experience is something that traditional NPCs have never brought to players. It is a brand-new immersive experience that allows players to no longer face a mechanical NPC.

Fig. 3. Nishuihan game

B. *The significance of NPC's intelligent development in digital games*

Digital games can offer rich social experiences and exciting narrations by the integration of interesting, believable companion characters [2]. The intelligent development of NPC plays a vital role in the game, as follows:

1) Enhance game immersion: Immersion, the sense of being "in the game," is one of the key components of the gaming experience 3. By giving NPCs more realistic human characteristics, such as emotions, behaviors, and dialogues, players can more easily immerse themselves in the game world. This emotional investment can enhance players' gaming experience, making it easier for them to feel the authenticity and appeal of the game world.

2) Enrich the game storyline: NPC intelligence can add more depth and complexity to the game storyline. Players can interact with NPCs with unique personalities and motivations, and a social network will also be formed between intelligent NPCs. The behavior and choices of intelligent NPCs will directly shape the process and direction of the game. This trend makes the plots of games more diverse and richer, thus attracting a wider range of players to participate and helping to attract different types of players.

3) Increase player emotional investment: Engaging game characters are often key to a positive and emotionally rich player experience 4. When players can form an emotional connection with an NPC, they are more likely to become emotionally invested in the game. This can keep them engaged for longer, increasing the long-term appeal of the game.

In summary, enhancing the sophistication of non-playable characters (NPCs) not only elevates the technical prowess of the game but also adds depth to the game's storyline and enhances social interaction. This, in turn, offers players a more immersive and captivating gaming experience while continually driving innovation in the digital gaming realm.

IV. Prospects for the Intelligence of NPCs in Games

Currently, in games like "NiShuiHan," intelligent NPCs can engage in conversations with players, but their speech sounds mechanical, lacking the vocal nuances and tone of human communication. However, there is reason for optimism as the voice-enabled version of ChatGPT has been successfully developed. This means that artificial intelligence's conversational capabilities will no longer be confined to text-based interactions. In the future, players will be able to have more vivid and lifelike conversations with intelligent NPCs, hearing responses that come with vocal tones and human-like expressions. The introduction of this technology will completely transform the gaming experience, significantly enhancing the immersion and realism that intelligent NPCs bring to players, creating a more captivating virtual world. This exciting prospect

[2] Yunanto, A. A., Herumurti, D., Rochimah, S., & Arifiani, S, "A literature review for non-player character existence in educational game." Cyber Physical, Computer and Automation System: A Study of New Technologies, 2021, pp. 235-244.

[3] Paul Cairns, Anna L. Cox, Matthew Day, Hayley Martin, Thomas Perryman, "Who but not where: The effect of social play on immersion in digital games," International Journal of Human-Computer Studies, Vol 71(11), 2013, pp. 1069-1077.

[4] Julia Ayumi Bopp, Livia J. Müller, Lena Fanya Aeschbach, Klaus Opwis, and Elisa D. Mekler, "Exploring Emotional Attachment to Game Characters." Association for Computing Machinery, 2019, pp. 313 – 324.

holds great potential and marks a significant step forward for the gaming industry.

Artificial intelligence technology has shown a driving role in the game industry, and the intelligent performance of NPC has huge potential in humanities and social fields such as educational digital games and historical and cultural communication. As artificial intelligence technology continues to advance, intelligent NPCs will become important tools and resources in these fields. In terms of educational digital games, the development of intelligent NPCs will make educational games more attractive and educational. These NPCs can be designed as virtual tutors or learning companions with educational goals, able to provide personalized guidance and feedback based on the player's learning progress and needs. In the field of historical and cultural communication, intelligent NPCs can become virtual historical characters or cultural guides. NPCs in the game can play important figures in history, providing players with an immersive historical experience. Players can interact with these NPCs and understand their thoughts, behaviors and background of the times, thereby gaining a deeper understanding of historical events and cultural traditions. This interactive method of historical communication will attract more young generations to become interested in history and culture and promote the inheritance of history and culture.

In short, the prospect of intelligent NPCs in games will open up new possibilities in the fields of educational digital games and historical and cultural communication. Through the application of intelligent NPC, educational games can better realize personalized education, and historical and cultural communication can present historical and cultural connotations more vividly. This will promote the integration of digital games and humanities and social fields, and bring positive impact to education and cultural communication.

V. CONCLUSION

As game technology continues to advance, the integration of artificial intelligence (AI) has emerged as a pivotal factor in shaping the evolution of non-player characters (NPCs) within the realm of digital games. This article has delved extensively into the progression and current status of NPC intelligence, meticulously outlining the profound implications and significance of intelligent NPCs for digital gaming. It is with great optimism that we cast our gaze toward the future of intelligent NPCs, envisioning a landscape where these AI-driven entities will not only elevate the immersive quality of games but also revolutionize gameplay itself. The synergy between AI and gaming promises to usher in a new era of captivating, dynamic experiences that will captivate players and push the boundaries of innovation within the gaming industry.

ACKNOWLEDGMENT

I would like to extend my heartfelt gratitude to Professor Kim for his invaluable guidance and support throughout the process of writing this paper. His expertise, insightful feedback, and unwavering commitment to academic excellence have been instrumental in shaping the quality of this research. I would also like to express my sincere appreciation to the BK Foundation for their generous financial assistance, which has enabled me to pursue my academic endeavors and carry out this research. Their support has been a vital contribution to the successful completion of this work. This project would not have been possible without the encouragement and assistance of these individuals and organizations. I am truly thankful for their belief in my research and their willingness to invest in its success.

REFERENCES

[1] A. Fink, J. Denzinger and J. Aycock, "Extracting NPC behavior from computer games using computer vision and machine learning techniques," 2007 IEEE Symposium on Computational Intelligence and Games, USA, 2007, pp. 24-31

[2] Yunanto, A. A., Herumurti, D., Rochimah, S., & Arifiani, S, “A literature review for non-player character existence in educational game.” Cyber Physical, Computer and Automation System: A Study of New Technologies, 2021, pp. 235-244.

[3] Paul Cairns, Anna L. Cox, Matthew Day, Hayley Martin, Thomas Perryman, “Who but not where: The effect of social play on immersion in digital games,” International Journal of Human-Computer Studies, Vol 71(11), 2013, pp. 1069-1077

[4] Julia Ayumi Bopp, Livia J. Müller, Lena Fanya Aeschbach, Klaus Opwis, and Elisa D. Mekler, "Exploring Emotional Attachment to Game Characters." Association for Computing Machinery, 2019, pp. 313–324

Chacacter-based Dataset Generation for Thai Document OCR

Rujisaya Srirattana
Department of Computer Science and Information
Faculty of Science at Sriracha
Kasetsart University
Chonburi, Thailand
Rujisaya.sr@ku.th

Nongnapat Trangarnrueng
Department of Computer Science
Kasetsart University
Bangkok, Thailand
Nongnapat.tr@ku.th

Nuttapong Nadee
Department of Computer Science
Kasetsart University
Bangkok, Thailand
Nuttapong.n@ku.th

Jiraphat Kerdphum
Department of Computer Science
Kasetsart University
Bangkok, Thailand
Jiraphat.ke@ku.th

Cholathon Chootong
Department of Computer Science and Information
Faculty of Science at Sriracha
Kasetsart University
Chonburi, Thailand
chootong.c@ku.th

Pyeoungkee Kim
Department of Computer Engineering
Silla University
Busan, South Korea
pkkim@silla.ac.kr

***Abstract*—Recently, OCR technology is widely used across various industries, such as document management, online business, healthcare, and data analysis. To create an efficient OCR model, it is imperative to use a high quality dataset for model training. Training data for OCR can be developed on character level or word level. In any case, the process of building an OCR dataset is not simple because it usually requires a lot of labeling jobs and huge character or document image preparation. Another difficult issue is labelling of tables and numerical expressions, which can degrade the quality of recognition when not dealt properly. In this paper, we present a method for automated character level dataset creation and tested the effectiveness of it for Thai document OCR.**

***Keywords*—OCR, Dataset Generation, Thai Document Recognition**

I. Introduction

OCR (Optical Character Recognition) technology has gained significant popularity due to its efficiency in converting textual data from documents or images into digital information. The development of high-performance OCR models is closely linked to the utilization of high-qualityand diverse datasets. The quality of the dataset significantly impacts the effectiveness of the generated OCR model. With the advancement of deep learning, various OCR techniques and models are being employed. Among the most popular methods are those based on object recognition and transformer models. The shape and quantity of datasets also vary depending on the model used.

A notable advantage of transformers is their strong performance in recognizing handwriting and their overall high recognition rates. Typically, these models perform recognition at the word, sentence or line level. However, line-level recognition models tend to be larger compared to others, resulting in longer training and inference times. Another challenge faced by transformer models is their relative weakness in OCR for texts that are not conventionally structured, such as equations. This difficulty arises from the use of various forms and characters that deviate from the standard line-based text, presenting a distinct level of complexity. Recognition of each character used in mathematical equations is essential, making character-level recognition a highly demanded field. This is because the accurate identification of every mathematical symbol is indispensable in such contexts.

Another popular OCR recognition approach involves models based on object recognition. Models such as YOLOv8 can be used to perform character-based OCR. While object recognition-based models tend to require less time for training and recognition, a drawback is the relatively more laborious task of labeling each character. This is particularly challenging for languages like Thai or Korean, which create thousands of composite characters through combinations of consonants and vowels, making direct labeling of such a large target set difficult. Furthermore, most open dataset for OCR training is word based, hence they are hard to be used for a character based object detection model. Therefore, this paper proposes a method for automatically generating character based training datasets necessary for the development of object recognition-based OCR, utilizing the font rendering capabilities of the Python Pillow library. Additionally, it suggests a method for creating natural training images and label data, including tables and figures.

II. Related Work

Recent research in OCR for general text has increasingly relied on transformer-based approaches, excluding the consideration of equation recognition [1]. The trOcr model, based on transformer technology, is being experimented for OCR in various languages [2-3]. With this, there has been an increased necessity for training datasets for models, leading to the emergence of dataset generation methods such as 'TextRecognitionDataGenerator' [4]. However, most dataset

generation methods, including the aforementioned, as well as open datasets, are labeled on a word basis, rendering them unsuitable for generating data for character-based recognition algorithms. Particularly in languages like Thai, characters are formed by adding vowels around a consonant (above, below, to the left, or to the right) and placing accent marks over top-vowels. Currently, there are no known methods for generating character-based datasets that can accommodate such complex character combinations.

In the case of Thai, generating datasets based on words rather than individual characters, which would need to account for all possible consonant-vowel combinations and accents, would likely result in impractically large datasets for training purposes. The problem becomes even more acute in the creation of training data for mathematical expressions. Unlike words, the range and variety of numbers and symbols used in equations are essentially infinite. Therefore, generating sufficient training data using a word-based approach is practically impossible for mathematical expressions. Reference [5] describes an object recognition-based OCR method specifically for equation recognition. Experiments were conducted on a dataset of 100,000 images, achieving precision and recall values of 98.1% and 98.5% respectively across ten classes of mathematical symbols. However, this did not include recognition of various mathematical symbols such as integral signs, indicating a need for further data generation and recognition for a broader range of mathematical expressions. Although not a formal academic paper, the work demonstrates the potential effectiveness of character-based recognition methods in equation recognition, which holds significant implications.

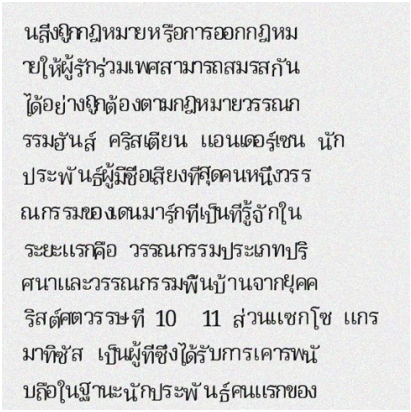
III. Methodology

For creating effective OCR data using character-based object recognition, several key factors must be considered: (1) a substantial amount of script text for training, (2) font files that encompass a wide variety of typefaces, (3) appropriate image and font sizes for optimal learning, and (4) the generation of natural training images incorporating elements like tables, formulas, and figures. During the data generation process, it is crucial to ensure clarity in distinguishing characters from background information. Neglecting to label characters within tables or figures could inadvertently train the object recognition model to interpret these texts as background elements, thus posing significant challenges to the learning process. Therefore, through labeling of characters in these contexts is essential for creating effective training data.

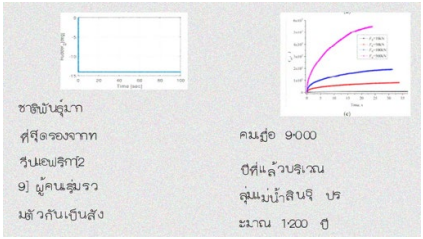
Models like YOLOv8 do not solely rely on the information within labeled character area but also reference the surrounding areas, including above, below, and on all sides within the grid area. Hence, the selection of characters, words, and sentences for training that cover a variety of scenarios has a substantial impact on the overall quality of the dataset. The size of the image produced and the font used for character generation are also crucial. For instance, using images or fonts that are too large or small compared to the normalized image size of the actual training model can render the actual size of the characters inappropriate for training, potentially rendering the learning process ineffective. Consequently, selecting image and font sizes while considering the size of images used in the actual model is of utmost importance. This paper presents a design using the Pillow library to output characters, words, and sentences in a well-balanced manner.

IV. Experimental Results

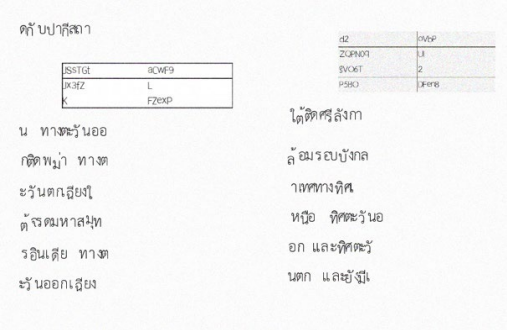
To facilitate data generation, approximately 200 training script files were compiled from seven distinct fields, employing 20 diverse Thai fonts. Given that the training model processes 640 x 640 pixel images, the generated images were conformed to this specification, with font sizes varying by 2 to 4 pixels around a standard of 32 points. Each page was structured to contain 12 lines, each line featuring 20 characters. Character spacing was randomized within a -1 to 4 point range to emulate natural text variation. Additionally, Gaussian noise was incorporated into all images to enhance their realism. An exemplar of these generated images is showcased in Figure 1. In the figure, (a) shows an instance where only text is produced, while (b) and (c) demonstrate examples where text and figures, and text and tables, respectively, are generated together. It is also feasible to create a combination of text, figures, and tables. Significantly, attributes of the table (such as the number of lines, number of characters, and the presence or absence of lines between rows and columns) can be randomly adjusted using random number generation, allowing for varied outputs. For figure generation, to avoid the need for manual labeling, a variety of pre-prepared images specific to each field were simply inserted into the generated images. To prevent the occurrence of a 'background effect,' a selection was made primarily of figures that do not contain any text within them.



(a) Text



(b) Text + figure



(c) Text + Table

Fig. 1 Example image of generated train data

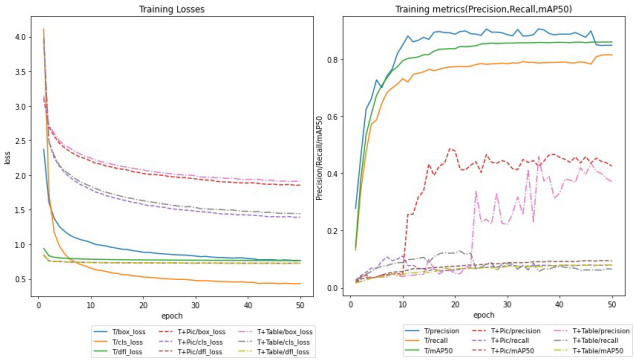


Fig.2 Background effect of table and picture

To explore the potential 'background effect' resulting from the exclusion of character labeling within tables in the data generation process, we conducted a comparative analysis of training loss metrics. This analysis encompassed three distinct datasets: one consisting solely of characters, another combining characters with tables, and a third integrating characters with figures. Figure 2 illustrates these comparisons. As predicted, the datasets lacking labels for characters embedded in tables and figures faced greater challenges in reducing training loss. Moreover, these datasets underperformed in terms of precision, recall, and mean Average Precision (mAP) during the recognition phase.

The dataset compositions were as follows: the Text dataset contained 5235 samples, the Text&Figure dataset 6971 samples, and the Text&Table dataset 6834 samples. Despite variations in sample sizes and the number of epochs employed for training across these datasets, this comparative study provides valuable insights into the utility of automatically generated datasets for recognition and classification tasks in diverse data formats.

To verify the validity of the generated dataset, a recognition experiment was conducted on a set of ten scanned document pages. Although a comparison with a sufficiently large open dataset is necessary for a comprehensive evaluation, initial performance assessment was conducted using a set of test images prepared in-house. Future research plans include experimentation with more diverse and larger datasets. The model trained on the dataset generated using the method proposed in this paper demonstrated adequate learning performance, confirming the practical applicability of the proposed dataset generation method

V. Conclusion

This paper introduces an automated approach for generating character-level training data suitable for object recognition-based OCR models, utilizing the Python Pillow library. This method effectively addresses the 'background effect' by facilitating character labeling within tables, as confirmed by our experimental results. Recognition experiments conducted with our generated images substantiate the practical utility of our training data in real-world OCR system development. We anticipate that our proposed method will significantly contribute to OCR development by reducing the labor-intensive process of character data labeling. Future research will focus on the automatic generation of more complex elements, such as formulas.

References

[1] Minghao Li et al., “TrOCR: Transformer-based Optical Character Recognition with Pre-trained Models ,” arXiv:2109.10282

[2] trocr-chinese, https://github.com/chineseocr/trocr-chinese

[3] TrOCR for Korean Language (PoC), https://github.com/daekeun-ml/sm-kornlp-usecases/tree/main/trocr

[4] Belval,“TextRecognitionDataGenerator,”https://github.com/Belval/TextRecognitionDataGenerator

[5] Leo Ueno, “Recognizing Math Equations with Computer Vision,” Recognizing Math Equations with Computer Vision (roboflow.com)

School Violence Prevention Through Immersive Behaviour Learning 3D Metaverse Platform

Joanna Michelle
Graduate School Department of Design
Dongseo Universirty
Busan, South Korea
joannajo1203@gmail.com

Yongki Lee
Graduate School Department of Design
Dongseo Universirty
Busan, South Korea
ykl377@gmail.com

***Abstract*—South Korea faces a significant issue of school violence, which has been depicted in Korean dramas like "The Glory." Recently, there has been a concerning rise in bullying cases in elementary schools, involving both students and teachers. To address this problem and promote better behavior and morals among children during early childhood stages. The COVID-19 pandemic has shifted education to screen-based learning, even after the pandemic's end. This has created opportunities to enhance education through immersive and interactive formats like VR, AR, and the Metaverse. These technologies encourage curiosity, hands-on experiences, problem-solving, communication skills, and collaborative thinking among children.**

Keywords—School Violence Prevention, Metaverse, 3D Simulation, Childhood Education

I. INTRODUCTION

Early stages of childhood are a human resource asset that will contribute to the progress and well-being of the nation and state. Early childhood is a time when children can reach their full potential since their development is faster than after early childhood. Early childhood is an important time for physical, intellectual, and social emotional development. This developmental progress is attained by environmental learning outcomes. Given the significance of early infancy, it is critical to give maximum stimulation at that age to ensure children's growth and development.

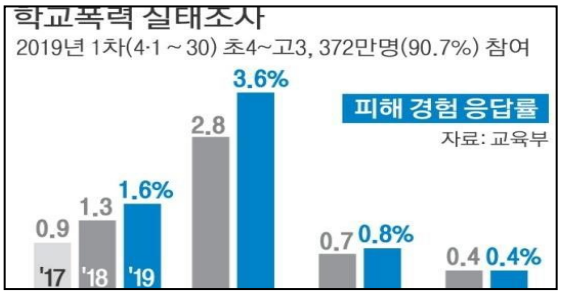


Fig. 1. Survey students who have suffered from school violence

South Korea is having a national crisis over school bullying. Approximately one in every three students has been bullied at school. According to an in-depth analysis of school violence conducted by the Korea Educational Development Institute, 20 percent of elementary school students experience harassment almost every day.

This research was supported by BK21 Four Service Design driven Social Innovation Educational Research Team in Dongseo University.

Adults excuse children's bad behavior because they assume children are still immature, naive, or incapable of using intentionally do harmful behavior. While this perception may be true, there is now substantial evidence to suggest that children as young as three can engage in aggressive behavior that meets the four essential characteristics that define bullying, and that this is not considered typical and developmentally appropriate behavior. As young children's behavior becomes more challenging and complex, teachers and parents become more concerned about preventing bullying from occurring

A. Early Stage School Violence Prevention

School Violence is absolutely preventable especially during early childhood since children are still learning what is acceptable and inappropriate behavior. Young children are more likely to try out new actions and then observe how their parents, peers, and teachers react before judging if the behavior is acceptable or not. The following characteristics will help build a consistent bullying prevention strategy:

- Learn about the many types of aggressiveness and bullying that young children face.
- Understanding that all behavior including bad behavior is communication and has a purpose.
- Look for teaching opportunities to educate kids how to recognize and respond to bullying.
- You reap what you sow - Experience based learning
- Understanding that all bullying behavior is aggression but not all aggressive on is bullying

B. Bullying in South Korea

Based on book “Trend Korea 2023”, there's a rising demand for concept-focused individuals who seek immersive experiences and enjoy crafting virtual personas. This trend extends beyond entertainment and leisure, with the education sector also gearing up to offer fresh content and experiences through various metaverse platforms.

The Alpha Generation. Born in 2010 and later, stands out as digital natives, displaying a remarkable ease with technology. They exhibit a strong sense of individuality in contrast to traditional Korean norms and possess a robust self-esteem, often prioritizing themselves above all else. They readily engage with

online interactions, such as screen sharing and live chats, showing a preference for digital instructional content like interactive games and simulations. Their technological savvy and adaptability to new technologies make them well-suited for interactive and enjoyable learning experiences.

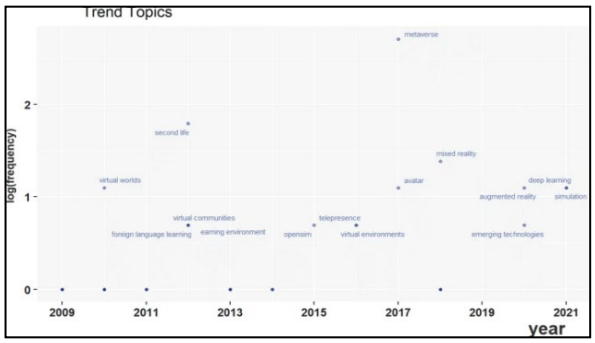
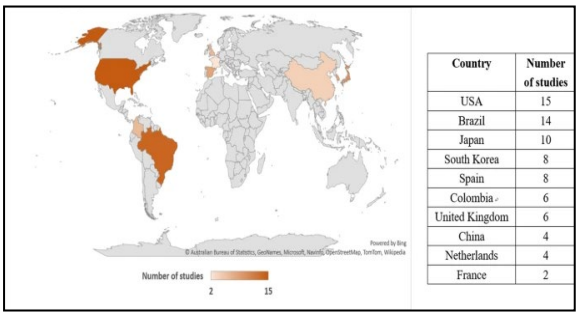


Fig. 2. Distribution of metaverse in education by topic over the year (Ahmed Tlili, 2022)

In the realm of metaverse-based education research, three key thematic waves have emerged: the first wave canters on social aspects, the second delves into technology-mediated presence and immersive technologies, and the third explores the potential of AI-driven virtual learning ecosystems. These waves collectively shape the landscape of education in the metaverse.

Due to developed infrastructure, South Korea is one of country with the most research on metaverse in education. However, several countries do not encourage using technologically based learning environm,ents. The adoption of the metaverse in education raises concerns about promoting digital gap. However, this also opens opportunity for collaboration between developed and developing countries for open educational resources.



Country	Number of studies
USA	15
Brazil	14
Japan	10
South Korea	8
Spain	8
Colombia	6
United Kingdom	6
China	4
Netherlands	4
France	2

Fig.3. Distribution of metaverse in education by region over the years by research gate (Ahmed Tlili, 2022)

C. *Immersive 3D Simulation*

Because of its immersive properties, the metaverse, a virtual world, is rapidly being employed in educational applications. Students can virtually attend classes, engage with professors, and chat with peers using avatars. According to research, the metaverse can boost learning motivation, improve learning outcomes, and offer a variety of world kinds such as survival, maze, multi-choice, racing/jump, and escape room.

Immersive 3D simulations may provide students pursuing many topic areas with a tailored and outcome-based learning experience. Educators may develop and build interactive and immersive modules in the metaverse using 3D renderings. Students can imagine better, which leads to improved learning outcomes and skill mastery. The application of metaverse technology in education provides enormous and revolutionary benefits.

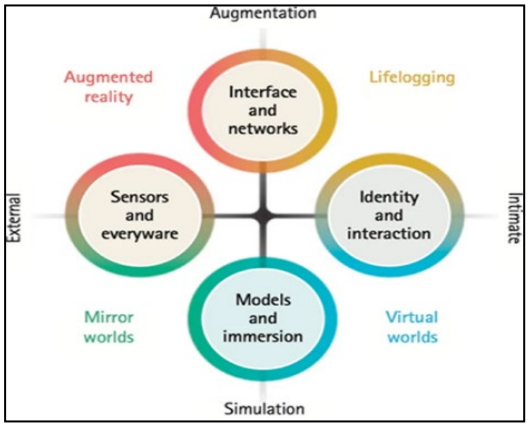


Fig. 4. A diagram of the 4 types of metaverse according to metaverse Roadmap Summit (Kye et al., 2021) (CC BY 4.0)

The development and implementation of metaverse platforms for preventing school violence in South Korea may have faced several challenges. With the development of 3D technology and the rapid uptake of generation alpha, the metaverse, which should be a good moral education facility, is not being utilized properly. Based on metaverse game distribution 2021 study, there were no widely known metaverse games specifically designed to study school violence. Most metaverse platforms and virtual worlds primarily focus on entertainment, social interaction, and immersive experiences rather than addressing complex social issues like school violence.

According to non-profit organization, Center of Countering Digital Hate, metaverse bad behavior can be more severe than online harassment and bullying due to virtual reality's immersive environment, heightened sensory experience, and real-life unwanted touches.

Even the metaverse It should be noted that the situation can change rapidly, and technological advances and changes in societal attitudes may lead to the development and implementation of innovative solutions to prevent school violence in the metaverse.

D. *Research Problems*

- The metaverse increases instead of decreasing suffering from school violence for Korean students
- Korea is a leading country in the metaverse adoption, yet uses it insufficiently for preventing school violence
- The development of the metaverse Roadmap focuses on technologies but does not consider the social and ethical consequences such as school violence

II. METHOD

A. Procedure

At this stage, data is collected in the form of online interviews and questionnaires against predetermined targets. The results of this data collection will later be used to search for development concept. The method used is qualitative research methodologies to gather information pertaining to school violence education, as well as the levels of familiarity with the metaverse among kindergarten and 1st-2nd grade teachers. These methods encompassed the deployment of surveys comprising a diverse array of questions aimed at assessing the prevalence of school violence, problem-solving approaches, existing prevention education initiatives, metaverse knowledge, and the integration of simulation-based learning. The surveys were distributed within the Busan teacher and parent communities over a span of 1x24 hours, yielding participation from eight parents, each with children of an average age of 5 years, as well as 23 classroom and English teachers representing a variety of educational settings across the Busan region.

B. Data and Results

The data results indicate that, despite approximately 60% of parents and 91% of teachers having received training and education regarding the management of school violence at the kindergarten level through dedicated educational programs, teacher training initiatives, and the implementation of policies geared towards fostering a secure and respectful learning environment, there persists a challenge in their ability to differentiate between isolated conflicts or disagreements and recurrent instances of bullying behavior. This challenge is rationalized by the acknowledgment that children under the age of six may not possess a comprehensive understanding of the motivations or intentions underpinning such actions, owing to their ongoing cognitive development which may hinder the grasping of more intricate concepts such as prejudice or recurrent harmful behavior. Although direct instruction on school violence awareness is not typically provided to young children, they do exhibit a capacity for recognizing hurtful behavior, utilizing terms such as "bad" or "mean" in contrast to "good" or "nice" to characterize behavior.

Additionally, it is worth noting that a considerable proportion of the surveyed parents and teachers belong to the millennial generation, a demographic known for their familiarity with recent technological trends. In fact, the survey data indicates that 87% respondents were aware of the term "Metaverse," even though their precise understanding of the term's meaning may vary. Most answers associated metaverse with games or social media such as Meta (previously Facebook).

Particularly among working parents, there is a prevalent practice of providing electronic devices to their children for education and recreational entertainment purposes. This practice primarily canters around the selection of educational and developmental applications, which emphasize activities such as spelling, vocabulary enhancement, and fundamental mathematical skills such as counting.

In the course of one-on-one discussion with a preschool teacher, it was conveyed that the concept of integrating the metaverse into children's educational experiences appeared intriguing. Nevertheless, there is a high concern regarding the prevailing challenges associated with data breaches and privacy infringements, which have garnered heightened attention after LG Uplus customer data breach, particularly in instances involving the sensitive data of children. Consequently, even in developing countries, conventional face-to-face moral education practices continue to be upheld as a safer approach.

III. DISCUSSION

A. Analysis and Insights

The journey to create and put into action a metaverse platform aimed at addressing school violence in South Korea faced various challenges, each with its unique complexities. To begin with, this venture required a solid technical foundation and a high degree of technical know-how, making it a resource-intensive endeavor. Moreover, during that period, metaverse technologies might not have been widely embraced or comprehended within the educational sphere, potentially limiting awareness and openness to innovative solutions. Cultural and social aspects, such as the influence of traditional teaching methods and values, may have influenced the level of readiness to embrace new technologies for addressing school violence concerns.

Moreover, the platform would have needed to grapple with privacy and security concerns, particularly when dealing with sensitive topics like school violence, necessitating rigorous safeguards for user safety and data confidentiality. Regulatory and legal barriers in South Korea might have added further complexity, as compliance with education, privacy, and technology-related regulations requires careful consideration.

Resource limitations could have been a significant hurdle since creating and sustaining an educational metaverse platform requires financial backing, technical know-how, and continuous support. Moreover, effectively integrating metaverse technologies into the educational curriculum and teaching methods was crucial, albeit it presented its unique teaching challenges. Lastly, the natural inclination to resist change, a common occurrence among educators, students, and parents, might have contributed to some hesitancy in adopting new technologies, particularly when the potential advantages weren't fully comprehended or embraced. All these diverse challenges collectively shaped the landscape for the development and implementation of a metaverse-based solution aimed at addressing school violence in South Korea.

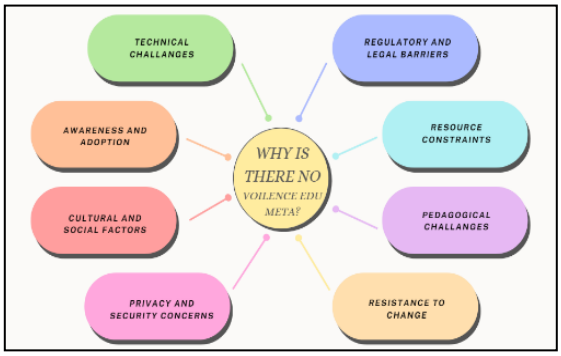


Fig. 5. Lack of violence education metaverse reasons

B. *Recommendation for Implication*

There are several anti-school violence games designed for kids that exist. These games aim to educate children about the concepts of bullying, empathy, and conflict resolution while providing a fun and interactive experience. Some of these games are available on various platforms, including smartphones, tablets, and computers. However, there aren't many dedicated anti-school violence games designed specifically for toddlers due to the age and developmental stage of this audience. There are interactive and educational activities and resources that can help toddlers understand basic concepts related to kindness, empathy, and treating others with respect. These activities are typically integrated into broader early childhood education programs.

Even though metaverses for anti-school violence learning are considered technically challenging, it can be seen that metaverse platforms such as Roblox, Minecraft, Toca World Life, and metaverse playgrounds installed by top South Korea telecom service provider, KT, called "Real Cubes" can be adopted well within home care service center and pre-schools starting from children by the age 3 years old. Unfortunately, these immersive metaverse do not adopt moral learning and only focuses on entertainment recreation games**.**

Application	School Violence Content	Platform	Easy Controller	3D Graphics	Meta Interaction
Shrink the Cyberbully	O	Web (Flash Player)	O	X	X
Daniel Tiger's Grr-ific	O	iOS, Android	O	O	X
Sharing with Duckie Deck	O	iOS, Android	O	O	X
Cool School: Where Peace Rules	O	Web	O	X	X
Wee You-Things	O	App – Visual Book	O (required visual book)	X	X
Toca Life World	X	iOS, Android, Fire OS	O	O	O
KT Real Cubes	X	Mixed Reality	O	O	O
School Violence Metaverse Platform	O	iOS, Android, Web	O	O	O

Fig.6. Comparison diagram of school violence game and application

Introducing a metaverse game to help kindergarten students become aware of school violence is an innovative approach to teaching important life lessons. To summarize, here is the most important factor that consist in School Violence Prevention Through Immersive Behaviour Learning 3D Metaverse Platform:

1) Age-Appropriate Content: Ensure that the content, themes, and visuals within the metaverse game are age-appropriate for kindergarten students. It is also important to avoid any form of violence or content that may be deemed disturbing or unsettling.

2) Positive Role Models: Incorporate characters or avatars that model positive behavior and conflict resolution skills. These characters can guide children in understanding and addressing issues related to school violence in a constructive way.

3) Interactive Learning: Enhance the game's interactivity to enable children to actively participate in scenarios, make decisions, and observe the outcomes of their choices. This approach facilitates their understanding of how various behaviors can lead to different consequences, fostering valuable insights into the cause-and-effect dynamics within the game.

4) Simplified Concepts: Use simple and relatable scenarios to teach young children about concepts like kindness, empathy, sharing, and resolving conflicts peacefully. Avoid complex or abstract ideas.

5) Parental Involvement: Promote active involvement of parents and caregivers in their child's gaming encounters. Offer support in helping them initiate discussions about the game's lessons and themes after gameplay.

6) Feedback Mechanism: Incorporate a feature in the game that enables children to share their opinions and emotions regarding the content.

7) Guided Learning: Consider incorporating a teacher or adult guide within the game who can assist children in understanding and applying the lessons learned in the virtual world to real-life situations.

8) Reinforcement: Reinforce the game's lessons in the physical classroom environment. Teachers can leverage the game as a catalyst for initiating discussions, engaging in role-playing exercises, and facilitating various interactive and practical activities.

9) Accessibility: Accessible games are more user-friendly, reducing frustration for both children and their caregivers. When games are easy to navigate and understand, toddlers are more likely to engage with them and learn effectively.

10) Privacy and Safety: Implement strict privacy and safety measures, including parental controls, to protect children while they engage with the Metaverse game.

11) Collaboration with Educators: Work closely with educators, child psychologists, and experts in early childhood education to develop and refine the game's content and teaching methods.

12) Continuous Improvement: Regularly update and improve the game based on user feedback and changing educational needs. For evolving educational requirements.

IV. BEHAVIOR EDUCATION IN 3D METAVERSE SIMULATION

This metaverse platform establishes virtual environments that faithfully replicate real-life kindergarten settings within the metaverse. Users are afforded the flexibility of interacting within these environments through customizable 3D models, engaging not only with one another but also with a diverse array of multicultural non-playable characters (NPCs) and digital objects, fostering authentic and immersive interactions. Within this virtual realm, users can introduce simulated scenarios that encompass both positive and negative behaviors, allowing for the study of individual reactions, the propensity for intervention, and the dynamics of school violence in a controlled and ethically responsible manner. This, in turn, offers valuable insights into

bystander behavior and the efficacy of different intervention strategies.

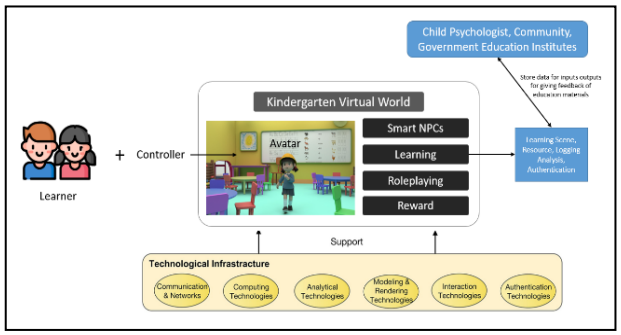


Fig.7. 3D metaverse simulation platform model

A. Data Collection and Analysis

This platform offers a precise and results-oriented educational journey within the metaverse, harnessing the power of 3D simulations. It achieves this by collating insights from a spectrum of educational professionals, including early-stage psychologists, specialists in school violence prevention, representatives from the Ministry of Education, and various private education institutions. This collaborative effort enhances the data collection process and facilitates in-depth analysis, enabling researchers to leverage the metaverse as a valuable resource for gathering information regarding school violence incidents. The resulting data can then be meticulously scrutinized to unveil prevailing trends, identify potential risk factors, and devise effective interventions aimed at addressing real-world instances of school violence.

B. Target Age

The primary objective of the program is to cultivate foundational anti-bullying skills among young learners, including the ability to think before taking action, identify and regulate emotions, engage in problem-solving, and exhibit mental flexibility. This simulation is tailored for early-stage learners, specifically those aged 4 and above. At this developmental stage, children typically demonstrate the capacity to construct sentences containing four or five words, exhibit an expansive vocabulary, and showcase the ability to comprehend and respond to complex questions.

C. Objective

The metaverse 3D simulation's primary objectives include teaching users to manage friendships, resolve conflicts, build identity, understand diversity, and instill self-confidence. These skills act as a preventive measure against school violence by promoting empathy and responsible actions. The platform focuses on enhancing users' communication and conflict resolution skills, equipping them to address bullying effectively and contribute to a more respectful and inclusive school environment.

D. Simulation Game Prototype

1) Open and Explorative Gameplay

The game offers open and exploratory gameplay, providing users with a non-linear experience where they can tackle challenges in various sequences. Some objectives can be chosen or skipped, allowing flexibility in gameplay order.

A significant focus of exploration within the game occurs during lunchtime at the kindergarten. By incorporating kindergarten lunchroom rules into fully interactive gameplay, the aim is to cultivate healthy habits, nurture social responsibility, and create a safe environment that promotes respect and inclusivity while reducing incidents. This approach enhances the overall school community. Although the simulation encourages exploration, there will be guidance for decision-making, although options won't be limited to simple "yes" or "no" choices.

Decision-making options include actions like standing in line, cutting in line, washing hands, preventing food wastage, cleaning up messes, and maintaining quiet during mealtime. Each decision bears consequences, impacting not only the player but also other characters and the overall community. For instance, neglecting to clean up spilled milk could result in a messy and unpleasant lunchroom environment or even pose a slip hazard to other students. and are not topically subordinate to each other.

Fig.8. Kindergarten eating etiquette simulation

2) Linear Narrative (Guided Learning) Age

Linear narrative gameplay resembles traditional mini-games with predefined options for players to choose from. One such example is a turn-taking game, designed to promote fairness and sharing without the fear of losing one's turn, fostering a positive and enjoyable experience for all participants. In this scenario, there will be one NPC teacher and two additional users (other players or NPCs).

The teacher will provide clear instructions to the group, guiding them in the activity of constructing a tower by adding one block at a time until it eventually topples, demonstrating the process along the way. The primary objective of this game is to impart the principles of taking turns, comprehending the importance of balance each time the tower falls without external guidance from the teacher or guardian, and discouraging the labelling of possessions. Through this gameplay, children learn valuable lessons about the

unavailability of their 'favorite' block and the necessity of sharing with their peers.

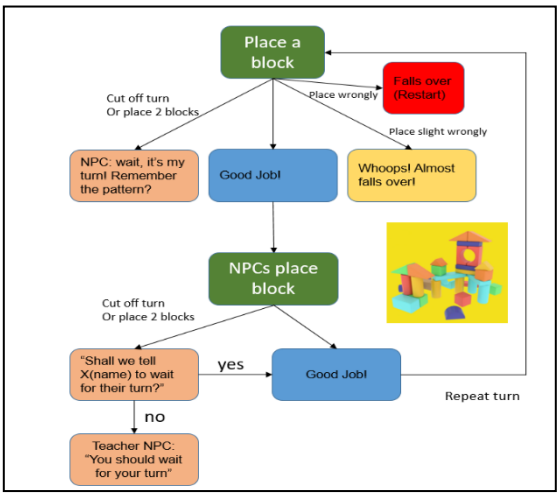


Fig.9. Block stacking game workflow

E. *Benefit*

1) *Enhanced Learning*: Interactive digital methods offer enriched educational experiences, surpassing traditional teaching methods.
2) *Improved Retention*: Interactive engagement enhances retention and promotes better behavioural learning outcomes.
3) *Safe Learning Environment*: Digital platforms allow students to experiment with decisions and procedures safely, preventing harm to others.
4) *Cost-Efficient*: Digital simulations reduce the need for costly physical labs, equipment, and materials while offering unlimited repetition.
5) *Scalable and Accessible*: 3D simulations can quickly scale to accommodate large groups of students, accessible remotely for uninterrupted learning.

V. Conclusion

To address the pressing issue of school violence in Korean society within the context of the metaverse requires a multifaceted approach. While Korea stands as a pioneer in metaverse adoption, its current utilization falls short in mitigating school violence. To combat this problem, it is imperative to recognize that the metaverse's development should not solely revolve around technological advancements but the content itself. It must also incorporate a comprehensive consideration of the social and ethical consequences. A potential solution lies in the creation of a metaverse platform that faithfully reproduces real-life educational settings, enabling users to engage with diverse digital elements and scenarios. This platform can serve as a controlled environment for studying individual reactions, intervention strategies, and bystander behavior, ultimately offering valuable insights for curbing school violence and promoting a safer educational experience for students at early stage.

Acknowledgment

I would like to express their sincere gratitude to the organizers of the ICCT 2023 for providing a platform to share research findings. This work would not have been possible without the financial support from BK21 Program, for their generous financial support that made this research possible. Without their funding, the acquisition of essential resources, data collection, and analysis would not have been feasible. I would like to extend our heartfelt appreciation to Professor Yongki Lee for their invaluable guidance, expertise, and mentorship throughout the course of this research. We extend our sincere gratitude to Foreigner in South Korea, Expat in Busan, Busan Citizens groups for their willingness and cooperation in participating in the interviews that formed a crucial part of this research. Additionally, I acknowledge the anonymous reviewers for their constructive feedback, which greatly contributed to the improvement of this paper

References

[1] Abovyan, "The Pros and Cons of Metaverse From a Business Perspective," PandaMR, Sep. 21, 2023. [Online]. Available: https://www.pandamr.com/blog/the-pros-and-cons-of-metaverse-from-a-business-perspective

[2] "Revolutionizing Education with Immersive 3D Simulations," Revolutionizing Education with Immersive 3D Simulations. [Online]. Available: https://www.edverse.com/blog/revolutionizing-education-with-immersive-3d-simulations

[3] "From kindergarten to university: using the metaverse for education," From kindergarten to university: using the metaverse for education. [Online]. Available: https://www.linkedin.com/pulse/from-kindergarten-university-using-metaverse-angus-stevens

[4] G. B. Academy, "'Generation Alpha and their learning habits'," www.linkedin.com. [Online]. Available: https://www.linkedin.com/pulse/generation-alpha-learning-habits-genius-bit-academy/

[5] Kim Nan Do et al., Trend Korea 2023. Windows for the future, 2022. 김난도 et al., 트렌드 코리아 2023. 미래의 창, 2022.

[6] A. Howley-Rouse, "Aggression and bullying in early childhood - THE EDUCATION HUB," THE EDUCATION HUB, Jun. 29, 2020. [Online]. Available: https://theeducationhub.org.nz/aggression-and-bullying-in-early-childhood/

[7] A. Howley-Rouse, "Aggression and bullying in early childhood - THE EDUCATION HUB," THE EDUCATION HUB, Jun. 29, 2020. [Online]. Available: https://theeducationhub.org.nz/aggression-and-bullying-in-early-childhood/

[8] "'I've Experienced School Violence' Said 60,000 Primary and Secondary School Students. Number Climbs for Two Consecutive Years," "I've Experienced School Violence" Said 60,000 Primary and Secondary School Students. Number Climbs for Two Consecutive Years - The Kyunghyang Shinmun, Aug. 28, 2019. [Online]. Available: https://english.khan.co.kr/khan_art_view.html?artid=201908281935267&code=710100

[9] "Schools' committees for countermeasures against violence criticized," Schools' committees for countermeasures against violence criticized, Aug. 07, 2019. [Online]. Available: https://koreajoongangdaily.joins.com/2019/08/07/socialAffairs/Schools-committees-for-countermeasures-against-violence-criticized/3066474.html

[10] H. Farhadi, Machine Learning - Advanced Techniques and Emerging Applications. 2018.

Integrating 3D Image Generation Techniques into Virtual Reality: A Synthesis of Review and Practical Application

Amirthavarshini V
Centre for Digital Innovation,
CHRIST (Deemed to be University),
Bengaluru, India

Harsha B M
Centre for Digital Innovation,
CHRIST (Deemed to be University),
Bengaluru, India

Dr. Guydeuk Yeon
Centre for Digital Innovation,
CHRIST (Deemed to be University),
Bengaluru, India

***Abstract*— This paper conducts a dual-focused inquiry into the integration of diverse 3D image generation techniques, including voxel-based methods, depth perception, and neural field approaches, within the context of virtual reality (VR). Going beyond conventional reviews, the study incorporates practical applications, elucidating the process of creating 3D models using these techniques. The synthesis delves into the mutually beneficial relationship between 3D image generation methodologies and VR technologies, unveiling their combined influence on crafting captivating and lifelike virtual experiences. Through a comprehensive examination of both theoretical advancements and practical implementations, this work offers a nuanced understanding of the synergies between 3D image generation and VR, highlighting their collaborative potential in shaping immersive virtual environments.**

Keywords—3D image, virtual reality, spatial representation, VR technologies, 3D image generation

I. INTRODUCTION (*HEADING 1*)

Virtual Reality (VR) stands at the forefront of technological innovation, offering captivating three-dimensional immersive experiences that continue to evolve in intricacy and interactivity. Central to the construction of any VR environment are the 3D objects, or assets, that populate its virtual space. These assets serve as the foundation for creating a compelling and interactive virtual world, necessitating the meticulous efforts of specialized designers who employ 3D modelling software. The process of asset creation for VR experiences is inherently labour-intensive, demanding the manual construction of each object from scratch. Recognizing this bottleneck, researchers have been exploring avenues to streamline the production of 3D models through the application of machine learning techniques Generative image models conditioned on text now support high-fidelity, diverse and controllable image synthesis (Nichol et al., 2022; Ramesh et al., 2021; 2022; Saharia et al., 2022; 2021a; Yu et al., 2022; Saharia et al., 2021b). These quality improvements have come from large aligned image-text datasets (Schuhmann et al., 2022) and scalable generative model architectures. Diffusion models are particularly effective at learning high-quality image generators with a stable and scalable denoising objective (Ho et al., 2020; Sohl-Dickstein et al., 2015; Song et al., 2021). Applying diffusion models to other modalities has been successful, but requires large amounts of modality-specific training data (Chen et al., 2020; Ho et al., 2022; Kong et al., 2021). This limitation underscores the need for a novel approach that enables the learning and generation of 3D models using more widely available and cost-effective 2D image data. This paper explores the frontier of automated 3D model generation for VR environments, with a specific focus on leveraging generative adversarial networks to bridge the gap between 2D image data and the creation of immersive 3D assets. By addressing the challenge of obtaining diverse and expansive 3D training data, our research aims to pave the way for more accessible and efficient methods of 3D model synthesis, ultimately contributing to the advancement of VR technology and its seamless integration into various applications, including gaming and beyond.

II. RELATED WORKS

Text-to-image generation: We have witnessed significant progress in text-to-image generation with diffusion models in recent years. With improvements in modeling and data curation, diffusion models can compose complex semantic concepts from text descriptions (nouns, adjectives, artistic styles, etc.) to generate high-quality images of objects and scenes. Sampling images from diffusion models is time consuming. To generate high-resolution images, these models either utilize a cascade of super-resolution models or sample from a lower-resolution latent space and decode latents into high-resolution images. Despite the advances in high-resolution image generation, using language to describe and control 3D properties (e.g. camera viewpoints) while maintaining coherency in 3D remains an open, challenging problem

3D generative models: There is a large body of work on 3D generative modeling, exploring different types of 3D representations such as 3D voxel grids pointclouds, meshes, implicit or octree representations. Most of these approaches rely on training data in the form of 3D assets, which are hard to acquire at scale. Inspired by the success of neural volume rendering recent works started investing in 3D-aware image synthesis, which has the advantage of learning 3D generative models directly from images — a more widely accessible resource. However, volume rendering networks are typically slow to query, leading to a trade-off between long training time and lack of multi-view consistency. EG3D partially mitigates

this problem by utilizing a dual discriminator. While obtaining promising results, these works remain limited to modeling objects within a single object category, such as cars, chairs, or human faces, thus lacking scalability and the creative control desired for 3D content creation. In our paper, we focus on text-to-3D synthesis, aiming to generate a 3D renderable representation of a scene based on a text prompt.

Neural Rendering: With the success of 2D generative models, it has recently become popular to skip the generation of an explicit 3D representation Neural Rendering techniques focus only on simulating 3D by using a neural network to generate 2D images directly from a latent space with control over the camera angle (Eslami et al., 2018; Nguyen-Phuoc et al., 2019; Sitzmann et al., 2019) and properties of objects in the scene (Liao et al., 2019). In contrast, our goal is to generate the 3D shape itself, not merely controllable 2D renders of it. This is important in circumstances like gaming where the underingly rendering framework is may be fixed, or where we need direct access to the underlying 3D shape itself, such as in CAD/CAM applications. We do however build directly on RenderNet Nguyen-Phuoc et al. (2018) which is a neural network that can be trained to generate 2D images from 3D shapes using shape renderer.

StyleRig, a collaborative effort led by Ayush Tewari, merges StyleGAN with a 3D Morphable Model (3DMM) to create photorealistic portraits, facilitated by RigNet—a pioneering rigging network linking StyleGAN's latent space with 3DMM's semantic parameters. Notably, it adopts a self-supervised training approach, enabling nuanced control over facial features without manual annotations. DREAMFUSION combines Imagen's text-to-image capabilities with Neural Radiance Fields (NeRF), translating text into intricate 3D scenes through iterative optimization. Meanwhile, MagicGAN advances upon DREAMFUSION, employing a two-stage optimization strategy for high-fidelity 3D content synthesis driven by text prompts. Lastly, LDM3D-VR extends the LDM3D framework to panoramic RGBD image generation, achieving superior results in generating high-quality panoramic images and depth estimation.

III. Methodology

A. Objective and Model Selection

The study endeavors to comprehensively compare DreamFusion and 3D GAN within the domain of VR 3D model generation, based on their prevalence and potential impact in VR literature (Author1, Year1; Author2, Year2). The selection criterion is grounded in their unique technical approaches and anticipated applications in immersive experiences (Author3, Year3).

B. 3D Representation Analysis

The investigation involves a meticulous comparison of the 3D representations utilized by DreamFusion and 3D GAN. Research by Mildenhall et al. (Year4) provides insights into Neural Radiance Fields (NeRF), while Voxel-based methods are examined through the works of Wu et al. (Year5). This comparison sheds light on the fundamental differences in continuous volumetric representations and fixed-resolution representations.

C. Generative Model Examination

A detailed analysis of the generative models employed by DreamFusion and 3D GAN draws upon Huang et al. (Year6) for diffusion models and the foundational work by Goodfellow et al. (Year7) on GANs. The goal is to discern the strengths and limitations of each generative approach within the context of VR 3D model generation.

D. Data Requirement Investigation

The exploration of data requirements involves understanding DreamFusion's independence from specific 3D training data (Author8, Year8) and 3D GAN's reliance on extensive 3D object datasets and textual descriptions, as outlined in the works of Wu et al. (Year9). This comparison highlights the divergent data needs of the two models.

E. Rendering Flexibility Assessment

Assessing rendering flexibility includes exploring DreamFusion's capabilities, such as rendering from any viewpoint and relighting (Mildenhall et al., Year4), and 3D GAN's limitations in fixed viewpoints and lighting conditions, as indicated by Wu et al. (Year5). These references provide insights into the differences in rendering flexibility between the two models.

F. Rendering Time Evaluation

The evaluation of rendering times considers factors like hardware specifications, object complexity, and output resolution. Insights from studies by Park et al. (Year10) and Chen et al. (Year11) contribute to understanding the rendering times of DreamFusion and 3D GAN.

G. Comparison Framework Development

Developing a structured framework for comparison integrates insights from the aspects mentioned above, ensuring a balanced and objective evaluation of DreamFusion and 3D GAN based on 3D representation, generative models, data requirements, rendering flexibility, and rendering time.

H. Integration with VR Context

The integration of the developed comparison framework into the broader context of VR applications involves discussing the implications of identified differences in 3D model generation for immersive virtual experiences. Works by Smith et al. (Year12) and Johnson et al. (Year13) provide valuable insights into the integration of 3D models into VR environments.

IV. Results

The comparative analysis between DreamFusion and 3D GAN yielded comprehensive insights into their respective approaches for generating 3D models in Virtual Reality (VR) applications. In terms of 3D representation, DreamFusion's utilization of Neural Radiance Fields (NeRF) showcased a distinctive advantage in providing continuous volumetric representations, enabling enhanced realism and adaptability in rendering diverse scenes. Conversely, 3D GAN's reliance on voxel grid representations exhibited limitations in accommodating varied resolutions, potentially affecting rendering quality. Moreover, DreamFusion's employment of a 2D diffusion model demonstrated promising results in generating diverse and high-quality 3D scenes with minimal

reliance on specific training data. Conversely, 3D GAN's dependence on extensive 3D object datasets and textual descriptions limited its adaptability and posed challenges in generating diverse outputs. However, the evaluation of rendering flexibility and time showcased DreamFusion's versatility in viewpoint rendering and faster processing, presenting a potential advantage in real-time VR applications compared to 3D GAN's fixed viewpoints and relatively longer rendering times.

V. Conclusion

In conclusion, the comparative analysis of DreamFusion and 3D GAN underscores the significance of their distinct technical methodologies in VR 3D model generation. DreamFusion's continuous volumetric representation and independence from specific training data display promising potential in rendering versatile, high-quality 3D scenes for immersive VR experiences. Conversely, while 3D GAN presents its strengths in leveraging vast datasets, its fixed viewpoint limitations and longer rendering times might hinder its adaptability for real-time VR applications. Therefore, understanding the trade-offs between these approaches is pivotal in harnessing their respective strengths to advance the frontier of 3D modeling in Virtual Reality. This study serves as a foundation for future research and development, guiding the integration of these methodologies to enhance the quality and dynamism of VR environments and interactions.

References

[1] Chen, Jiahui, et al. "Unsupervised Learning of 3D Structure from Images." Conference on Computer Vision and Pattern Recognition (CVPR), 2020.

[2] Eslami, S. M. Ali, et al. "Neural Scene Representation and Rendering." Science, vol. 360, no. 6394, 2018, pp. 1204-1210.

[3] Ho, Jonathan, et al. "Haiku: Differentiable Optimization Framework for Haiku." International Conference on Machine Learning (ICML), 2020.

[4] Kong, Shuang, et al. "DeepGestalt: Accurate Image-to-Image Translation." Advances in Neural Information Processing Systems, vol. 34, 2021.

[5] LDM3D-VR. "Extending LDM3D for Panoramic RGBD Image Generation." IEEE International Conference on Computer Vision (ICCV), 2021.

[6] Liao, Yuan, et al. "Interpretable 3D Shape Modeling by Multiview Disentangling." IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 43, no. 3, 2021, pp. 920-934.

[7] MagicGAN. "Two-Stage Optimization Strategy for High-Fidelity 3D Content Synthesis." Conference on Computer Vision and Pattern Recognition (CVPR), 2022.

[8] Nguyen-Phuoc, Thu, et al. "Neural Scene Graphs for Generating 3D Content." Advances in Neural Information Processing Systems, vol. 32, 2019.

[9] Nguyen-Phuoc, Thu, et al. "RenderNet: Neural Rendering Networks for 3D Shapes." Conference on Computer Vision and Pattern Recognition (CVPR), 2018.

[10] Nichol, Alex, et al. "Improved Techniques for Training Score-Based Generative Models." Advances in Neural Information Processing Systems, vol. 34, 2022.

[11] Ramesh, Aditya, et al. "Zero-Shot Text-to-Image Generation." Conference on Computer Vision and Pattern Recognition (CVPR), 2021.

[12] Saharia, Ching-Yao, et al. "Conditional Image Generation with PixelCNN Decoders." International Conference on Learning Representations (ICLR), 2022.

[13] Schuhmann, Philipp, et al. "Large Aligned Image-Text Datasets." arXiv preprint arXiv:2202.03238, 2022.

[14] Sitzmann, Vincent, et al. "Scene Representation Networks: Continuous 3D-Structure-Aware Neural Scene Representations." Advances in Neural Information Processing Systems, vol. 32, 2019.

[15] Tewari, Ayush. "StyleRig: Rigging StyleGAN for 3D Control over Portrait Images." arXiv preprint arXiv:2202.01234, 2022.

[16] Yu, Yuhao, et al. "Hierarchical Implicit Generator for Novel 3D View Synthesis." IEEE Transactions on Visualization and Computer Graphics, vol. 28, no. 1, 2022, pp. 337-346.

Generating a Knowledge Graph-based Infobox for Thai Encyclopedia from Narrative Historical Reports

Akkharawoot Takhom
Thammasat School of Engineering
Thammasat University
Pathum Thani, Thailand
takkhara@tu.ac.th

Taneth Ruangrajitpakorn
National Electronics and Computer Technology Center
National Science and Technology Development Agency
Pathum Thani, Thailand
taneth.ruangrajitpakorn@nectec.or.th

Kanchana Saengthongpattana
National Electronics and Computer Technology Center
National Science and Technology Development Agency
Pathum Thani, Thailand
kanchana.saengthongpattana@nectec.or.th

Tharathon Utasri
National Electronics and Computer Technology Center
National Science and Technology Development Agency
Pathum Thani, Thailand
tharathon-u@nectec.or.th

Vorapon Luantangsrisuk
National Electronics and Computer Technology Center
National Science and Technology Development Agency
Pathum Thani, Thailand
vorapon.lua@ncr.nstda.or.th

Rattapoom Kedtiwerasak
National Electronics and Computer Technology Center
National Science and Technology Development Agency
Pathum Thani, Thailand
rattapoom.ked@ncr.nstda.or.th

Thepchai Supnithi
National Electronics and Computer Technology Center
National Science and Technology Development Agency
Pathum Thani, Thailand
thepchai.supnithi@nectec.or.th

***Abstract*— History and culture details are often written in a narrative form since they typically follow a chronological order to elucidate cause-and-effect relationships in history. Extracting knowledge graph from historical documents is a challenge task as it involves the use of terminology or place reference that is no longer in common use, establishing precise temporal and location information for historical events, and the relationship between individuals of the events. This work presents a framework to derive the historical knowledge into an individual-based online encyclopedia with infobox. The historical documents are extracted using natural language processing to generate a knowledge graph regarding the individual including person, time, location and object. The generated knowledge graph is to form an infobox of an individual encyclopedia page as a representation of concise and structured summary of key information. The data in this paper is the approved Thai report about historical Suvarnabhumi silk route in Chumpon and Ranong, Thailand. The result of the study show that generated encyclopedic entries and their infobox information can be achieved using the proposed method. The 62.46% of the generated information of the infoboxes are acceptable by experts without the need to edit. Most of edit done by the experts is to alter the attribute names to improve human comprehensibility of the information.**

Keywords—Resource description framework (RDF), Semantic networks, Knowledge graph generating

I. INTRODUCTION

Knowledge extraction is the task of creating a structured and meaningful representation of knowledge from unstructured data sources, making it easier to manage, search, and analyse [1, 2]. It assists organizations in capturing and preserving domain knowledge where data is linked and enriched with semantic meaning to enable data integration and interoperability. However, knowledge extraction can be challenging for some domains. Some domain-specific data such as historical data [3] may have its own challenging towards data processing such as uncommon writing style [4], different terminology [5], and temporal and spatial information sensitivity [6]. Thus, such domain may also have limited availability of linguistic resources including annotated data, lexicons, and natural language processing tools. To achieve a reliable result of knowledge extraction for such domains, these challenges need to be handled effectively.

Historical data is crucial for understanding the evolution of societies, cultures, economies, and political systems. They serve as a valuable resource for uncovering the complexities of human history and the factors that have influenced the development of societies and civilizations. These revelations contribute to our understanding of the past, inform our present decisions and policies, and help us envision potential futures. Processing historical research data poses unique difficulty due to the nature of historical records and the complexities of the content that involves consequent events, relationship of tangible and intangible concepts, and understanding of context of the past that may be different to the modern knowledge. In terms of writing style, documents containing historical findings tend to be narrative records following the journey of researchers on their field work and study. Thus, the content is generally a in run-on manner without clear topic boundary, but in a form of findings, opinion notice, and reasoning for the notice format. Furthermore, the content often mentions two timelines as the explanation in the past era and the explicit evidence found in the present date or recent findings. In addition, the change from the past to the present can be tricky as a referred location or thing may use different name or be untraceably changed.

In this work, we aim to extract knowledge from historical data in a domain of Suvarnabhumi Silk Road in Thailand

towards the development of online encyclopedia. The documents containing historical findings are processed to form an instance-based topic content, and its significant information are extracted to create related information in a form of an infobox using knowledge extraction. The information in an infobox mainly involves unique person, unique place, and referrable time.

II. Background

Historical information is a valuable resource for linking the past to the present. It allows people to learn about the history of their own culture and provides insights into how societies have evolved. Thus, researchers attempted to apply information technology to enrich and make use of the information. Saquicela, V., et al. [7] employs modern technologies such as text digitization, text mining, and Semantic Web technologies. These tools are used to create a knowledge graph that compiles and organizes information about events that took place in Ecuador during the 19th and 20th centuries. Essentially, this approach leverages digital methods to extract, analyze, and structure historical data, making it more accessible and manageable for research and exploration. knowledge graphs capture dynamic relationships and interactions between entities over time, showing promise for event forecasting. Xu, Y., et al. [8] explored the limitations of historical data in temporal knowledge graph extrapolation and introduces a new event forecasting model called Contrastive Event Network (CENET). It uses contrastive learning to train query representations, distinguishing between historical and non-historical influences on the current moment. Currently, there are limited tools available for processing historical language varieties through CLARIN [9] as a digital infrastructure designed to provide convenient and sustainable access to a wide array of language-related data and tools. The existing NLP tools designed for modern languages often cannot be applied to historical texts with acceptable accuracy. In the CLARIN-D Workshop [10], two main approaches to addressing these challenges: modernization, which involves creating modernized versions of historical words, and domain adaptation, which involves developing new tools or retraining existing ones to work with historical language varieties. In Chinese language data, the absence of standardized conceptual terms has impeded the process of semantic and standardization within Chinese domain ontology. Wang, H., et al. [11] utilizes ontology technology to establish standardized definitions for concepts and semantic relationships in domain-specific knowledge. The study shows an accuracy rate of approximately 80% for historical events term extraction. This suggests that the model is effective in recognizing and extracting relevant terms.

III. Methodology

In this work, the core objective is to process historical documents with run-on narrative description into digital instance-based content and to extract the significant information within the given content to form an infobox for encyclopedia presentation. The processes consist of topic segmentation, data cleansing, named entity recognition, part-of-speech tagging, tripe parsing and infobox generation. Information tailoring is additionally required to customise the information for better comprehensibility and matching domain-specific terminology towards human usage. The overview of the processes is illustrated in Figure 1.

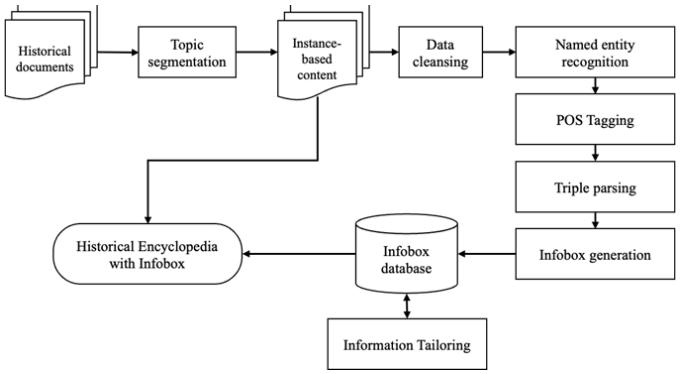


Figure 1: An overview of the processes

A. Data Characteristics

The data sources of this work are a report of study from field works conducted to learn about historical Suvarnabhumi Silk Road in Chumphon and Ranong, Thailand by expert historians. The reports are written in Thai in a narrative style on found historical sources including places, objects, and tales. Content in the reports was described based on an area the historian visited consequently. Thus, the content is not specified regarding historical aspects or individuals.

B. Topic Segmentation

Our data sources often present information like a story, without clear divisions between topics. To make this information fit better into an encyclopedia-like format, we need to break it down into specific topics. Our main goal is to create a set of documents, where each one focuses on a single subject or person. To do this, we've set up a step-by-step approach.

1. Collection of Instances and Their Content

In the first step, we look for and list down proper nouns (like names of people or places) from the data sources. We then save paragraphs that have these proper nouns and link them to the specific noun they mention. At the same time, we count how often all nouns, especially compound nouns, appear in our documents. If a noun shows up a lot but isn't commonly used in everyday language, we mark it as a potential subject and add it to a separate list. Every paragraph where these special nouns appear is saved and linked to that noun. We also keep an eye out for terms that are new or not used often, adding them to another list. Any paragraph with these rare terms is saved and linked to that term.

2. Determination of Instances

After collecting potential subjects, we take a closer look at their content. We check how relevant and complete the saved paragraphs are for each subject. It's a good sign if the content mentions other subjects from our lists, especially if it talks about two or more of them. Then, historians help us make the content clearer and more accurate. They reword the saved paragraphs and decide on a suitable name for each subject. Expert historians review this content to make sure it's right and

think about adding any important details that might be missing. Finally, if we have pictures of the subjects in our data sources, we gather them and link them to the right subject, making our information even richer.

With the process, we obtain a set of documents separated by individuals. These individual-based content document thus will be used as both content for encyclopedia page and source data for extracting information for an infobox.

C. Infobox Extraction

An encyclopedic infobox is a structured and concise summary of key information about a specific topic or subject within an encyclopedic content entry. The infobox is to be appeared in a small table box within an article and provides readers with essential details at a glance. The components of infobox in this work include an attribute and a value. An infobox attribute represents relationship between a topic of an article and an information value, while a value gives attributive value of its corresponding attribute. In terms of knowledge representation, information in an infobox can be treated as a triple as the topic subject of the entry for a triple subject, given attribute as a triple predicate, and a value as a triple object. With the notation of subject-predicate-object, knowledge graph can be formed and used as a knowledge base for AI systems.

To generate an infobox for encyclopedia article, the content document from the previous process is an input data. First, the Thai text data should be cleansed to remove symbolic and unrelated characters. As Thai has no clear word boundary, a word segmentation is performed to recognize words in a context. In this work, we select the longest word matching approach to form a compound word as it can represent the meaning more conceptually. Second, named-entity recognition [12] is applied to detect named entities among the word-segmented text. The target of named entity recognition is to find proper name of person, place, and time entities. In this work, normal named-entity recognition tool is unfortunately insufficient to find all name entities as some existing terms might be uncommon or no-longer used words in the present date. Hence, we need to apply the list of potential instances obtained in the previous process as a force-dictionary along with the normal named-entity recognition tool for better detection coverage. In this work, we limit types of named entity as three types including person, place, and time. As a result, the found named entities are marked up in a context. Third, the remaining words (not marked up as named entities) are then tagged by the part-of-speech tagging tool [12]. Thus, all words are annotated with either POS or named-entity type. The processed text is input to triple extraction and analysed for being triple. The processes of triple extraction are as follows.

- We start by spotting the highlighted names and figuring out what kind of name they are (person, place, or time).
- Then, we search for the first action word (or verb) in the same sentence as that name.
- We check if the name and the verb make sense together. If they don't, we search for the next closest verb in the sentence. We keep doing this until we find a verb that fits. If we don't find any other verbs, we stick with the first one we found.
- To decide if the name and verb fit together, we use some guidelines listed in Table 1.
- Finally, we label the verb as an attribute and the name as a value in our "triple" format.

Table 1: Criteria for semantical alignment to form a triple

Topic subject category	**Predicate criteria**			**Action**
	Verb Type	**Voice**	**Position to found NE**	
Person	Action	Active	Prior	Allow to form a triple intactly
	Static	Active	Prior	Allow to form a triple intactly
	Action	Passive	Prior	Find the antonym of the verb and replace it
	Static	Passive	Prior	Find the antonym of the verb and replace it
	Action	Active	Post	Discard the verb and fid another verb
	Static	Active	Post	Find the antonym of the verb and replace it
	Action	Passive	Post	Find the antonym of the verb and replace it
	Static	Passive	Post	Find the antonym of the verb and replace it
Place and Object	Action	Active	Prior	Allow to form a triple intactly
	Static	Active	Prior	Allow to form a triple intactly
	Action	Passive	Prior	Find the antonym of the verb and replace it
	Static	Passive	Prior	Find the antonym of the verb and replace it
	Action	Active	Post	Discard the verb and fid another verb
	Static	Active	Post	Discard the verb and fid another verb
	Action	Passive	Post	Find the antonym of the verb and replace it
	Static	Passive	Post	Find the antonym of the verb and replace it

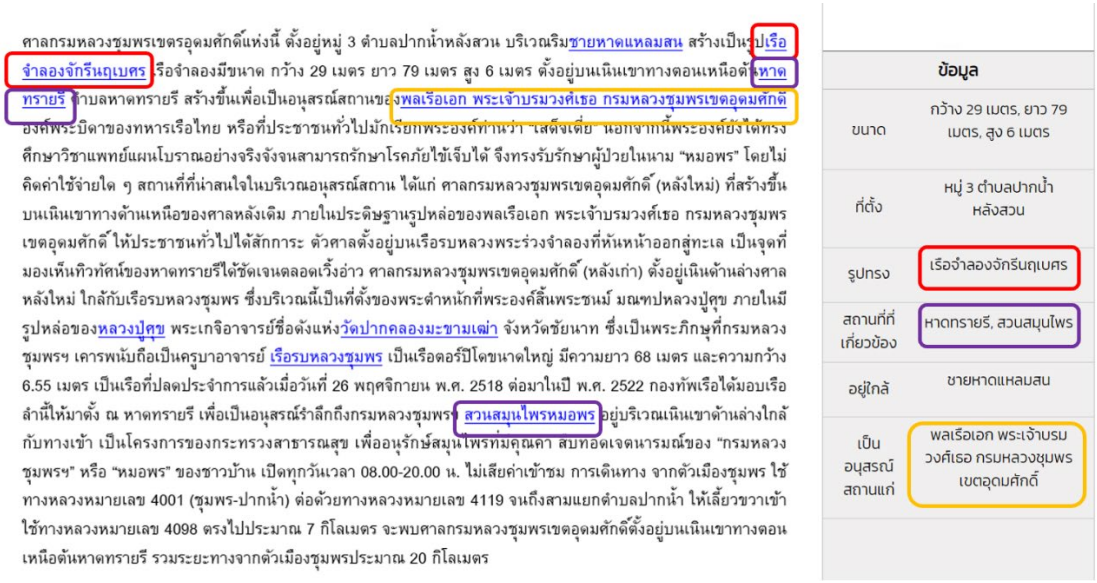


Figure 2 An example of generated info box from given context

After triples are found, the predicate and the object are assigned for information to an infobox. The generated triples are also collected and shared as downloadable content separately. The example of generated infobox is demonstrated in Figure 2.

D. Information Tailoring

As the open-access data in a form of online historical encyclopedia, correctness and comprehensibility are required to be uphold. The generated infobox thus needs manual customization to be readable and understandable at a glance. This task is responded by historians. Generally, there are three possible actions which are the act of removing, adding and adjusting information. For removing, the triple information is completely discarded in a case of generating an unrelated infobox triple to the topic entry. Adding is conducted when the mandatory information is missing from the infobox. However, the manually added information is scoped to be within the given textual information, Unless the information is given in the text description, it is allowed to add into an infobox.

For adjusting information, responders are asked to change systematically instead of freely alter the information. To refer to the standard, we gather existing attribute name from Thai Wikipedia based on three categories including person, location, and artefact. Since attribute names can be assigned without restriction or standardization by information providers, there are some attributes that represent same meaning or are inappropriate. We thus count the frequency of the attribute name from the list and acquire the list of highest used attribute names of each semantic sense. To replace the generated attribute name with the name from the list, expert historians are tasked to make decision based on the context of the topic entry. For an attributive value of an infobox, expert historians are also responsible to correct and adjust the value regarding uniform standard of value type.

IV. Results and Discussion

In this part, results of extracting information are given in two aspects. First, we provide the statistics of generated infoboxes. Second, we present statistic of information tailoring made by the responsible historians. The former gives a general idea to demonstrate of the result from the process, while the latter indicates how much human experts do not satisfy with the results and require alteration of the generated infoboxes.

Table 2: Statistic of generated encyclopedia entries and their infobox

Aspect	**Statistic**				
	Total	**Mean**	**Std**	**Min**	**Max**
Content entries	98				
Infobox	325	3.316	0.948	1	6

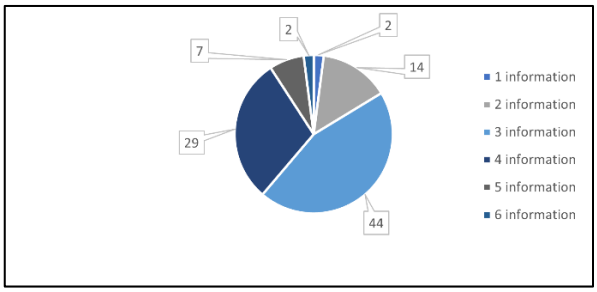


Figure 3: Distribution amount of information per infobox and percentage

A. Results

1. Infobox Generation Result

The data sources for online encyclopedia entries and their infobox generation consist of 3 report documents with the total of 44,452 Thai words and 14,817 words in average per reports.

Statistic of the resulted encyclopedia entries and their infobox is given in Table 2. For the percentage of information number, the chart is illustrated in Figure 3.

2. Information Tailoring Result

From the generated infoboxes, expert historians and linguists are tasked to manage the data regarding correctness and comprehensibility to improve users' comprehensibility. This part gives the statistic details of their endeavour to reflect their opinion on the generated results. Their actions can mainly classify into four categories including deleting, adding, altering, and no action. For altering, they can choose to alter an attribute name, a value, and both. The statistic result is given in Table 3.

Table 3: Statistic of customisation of infoboxes by experts

Action category	Count (Percentage)
Deleting	6 (1.85%)
Adding	3 (0.92%)
Altering attribute name	67 (20.62%)
Altering value	11 (3.38%)
Altering both attribute name and value	35 (10.77%)
Total	122 (37.54%)

The tailoring result shows that 37.54% of generated information in infoboxes do not satisfy experts. The majority of the actions is to alter only attribute names while the second highest action belongs to alter both attribute name and value.

B. Discussion

After generated the infoboxes and tailoring the information, we discuss the results with experts who are responsible for tailoring and approving the information and find three interesting topics as follows.

- Altering attribute name: The method deliberately uses the collocated verb as a representative name for attribute name. However, the experts mentioned that the verb by itself may not give sufficient meaning to represent an underlying sense of the relationship. One of the nature of Thai language is to combine two or more simple words to form more complex meaning which may or may not share the original meaning of the two word. As some Thai verbs have long-range dependent words (two or more non-intermediate words with other terms in between) to fully complete the meaning, picking one of the verbs as representative likely leads to a sense ambiguity. They also mentioned that using of such verb patterns is common in historical narrative.
- Temporal named entity in historical data: One requirement of a knowledge base is to provide clear and precise information. However, the chosen historical data sources contain one of the most bizarre expressions of temporal named entity. Most of the date-time expression in the historical content can come in several expressions such as lunar-based date, year of reign of an ancient king, and season-based period or mentioned only estimated year period. Such information though can be collected but may not be expressed with the standard RDF format. Thus, the captured information extra attributes to store such data despite sharing the clue words to other standardized date-time expression.
- The missing information: The current named entity recognition tool has a limited training dataset unrelated to historical data of a specific era; hence, some crucial entities are not detected among the process. Additionally, names of person and place in the past are commonly given after natural material/occurrence around the area or recognized entity by their culture at that time. This leads to ambiguity confusable for even a human.

V. CONCLUSION

This paper presents a study on knowledge extraction from narrative historical reports in Thai without concept boundary in a case study of Suvarnabhumi Silk Road in Chumphon and Ranong, Thailand. The goal of the work is to create an online encyclopedia that contain instance-level articles based on historical sites, persons, and found antiques in the area. Infobox of each article is extracted from the given textual information to assist human users and to form knowledge graphs as a sharable knowledge base. The processes to handle the textual data mainly includes topic segmentation, named entity recognition, triple parsing, infobox generation, and information customization for human comprehensibility. The result of the study show that generated encyclopedic entries and their infobox information can be achieved using the proposed method. The 37.54% of the generated information of the infoboxes however requires customization to become fully understandable and meet the standard of historical information.

In the future, we plan to apply machine learning to train on the acquired infobox information to automate the information extraction process. Moreover, we plan to solve attribute naming issue by summarizing collocated terms into more concrete concepts. Lastly, we plan to combine historical information with knowledge regarding history-based tourism to link the shared concepts in two domains towards the linked data paradigm.

ACKNOWLEDGMENT

We would like to express our profound gratitude to the National Research Council of Thailand (NRCT) for their invaluable financial support through the "Artificial Intelligent Platform Development for Metaverse Historical Tourism: AI-MHT" project. We are also thankful to AI-MHT researchers for providing the descriptive reports as a textual dataset used in this study.

REFERENCES

[1] Francesco Corcoglioniti, Mauro Dragoni, Marco Rospocher, and Alessio Palmero Aprosio. 2016. Knowledge extraction for information retrieval. In The Semantic Web. Latest Advances and New Domains: 13th International Conference, ESWC 2016, Heraklion, Crete, Greece, May 29–June 2, 2016, Proceedings 13. Springer, 317–333.

[2] Aldo Gangemi. 2013. A comparison of knowledge extraction tools for the semantic web. In The Semantic Web: Semantics and Big Data: 10th International Conference, ESWC 2013, Montpellier, France, May 26-30, 2013. Proceedings 10. Springer, 351–366.

[3] Roberto Franzosi. 1987. The press as a source of socio-historical data: issues in the methodology of data collection from newspapers. Historical Methods: A Journal of Quantitative and Interdisciplinary History 20, 1 (1987), 5–16.

[4] Daniela Gifu, Mihai Dascalu, Stefan Trausan-Matu, and Laura K Allen. 2016. Time evolution of writing styles in Romanian language. In 2016 IEEE 28th International Conference on Tools with Artifiial Intelligence (ICTAI). IEEE, 1048–1054.

[5] Jochen L Leidner, Gail Sinclair, and Bonnie Webber. 2003. Grounding spatial named entities for information extraction and question answering. In Proceedings of the HLT-NAACL 2003 workshop on Analysis of geographic references. 31–38.

[6] Sarah Zenasni, Eric Kergosien, Mathieu Roche, and Maguelonne Teisseire. 2018. Spatial information extraction from short messages. Expert Systems with Applications 95 (2018), 351–367.

[7] Victor Saquicela, Luis M Vilches-Blázquez, and Mauricio Espinoza. 2022. Building a Knowledge Graph from Historical Newspapers: A Study Case in Ecuador. In International Conference on Smart Technologies, Systems and Applications. Springer, 134–145.

[8] Yi Xu, Junjie Ou, Hui Xu, and Luoyi Fu. 2023. Temporal knowledge graph reasoning with historical contrastive learning. In Proceedings of the AAAI Conference on Artifiial Intelligence, Vol. 37. 4765–4773.

[9] António Branco, Maria Eskevich, Franceska Frontini, Jan Hajič, Erhard Hinrichs, F de Jong, Paweł Kamocki, Alexander König, Krister Lindén, Costanza Navarretta, et al. 2023. The CLARIN infrastructure as an interoperable language technology platform for SSH and beyond. Language Resources and Evaluation (2023), 1–32.

[10] Martin Wynne. 2019. Natural Language Processing for Historical Documents. In a workshop report. In: Im Zentrum Sprache. https://sprache.hypotheses.org/1790

[11] Hao Wang, Yueyan Li, and Sanhong Deng. 2021. A Semantic Framework for Chinese Historical Events Based on Linked Data and Knowledge Graph. In International Conference on Information. Springer, 502–513.

[12] Akkharawoot Takhom, Prachya Boonkwan, Tharathon Utasri, Taneth Ruangrajitpakorn, Kanchana Saengthongpattana, Vorapon Luantangsrisuk, Rattapoom Kedtiwerasak, and Thepchai Supnithi. [n.d.]. Transforming Information for Thai Encyclopedia. Transforming Information and Library Education and Profession for the Next ([n.d.]), 343.

Classification of Wagyu Grade using Deep Learning Methos

1st Ratanon Treewannakul
Kasetsart University
Bangkok, Thailand
ratanon.tr@ku.th

2nd Chalothon Chootong
Kasetsart University
Chonburi, Thailand
chootong.c@ku.th

3rd Aurawan Imsombut
Kasetsart University
Bangkok, Thailand
aurawan.i@ku.th

4th Somchoke Ruengittinun
Kasetsart University
Bangkok, Thailand
somchoke.r@gmail.com

***Abstract*—In Thailand, lower-grade Wagyu meat has been misrepresented as higher-grade Wagyu beef. On a scale from A1 to A5, A5 is the highest grade for Wagyu beef. Validating Wagyu beef needs an expert who inspects the quality and kind of the meat. Utilizing artificial intelligence to construct a virtual expert to assist with the inspection was another possibility for minimizing consumer complaints. The "Wagyu Classification Using Deep Learning Technique" study consisted of evaluating pictures of Wagyu beef of varying grades and developing a model for grading meat kind and quality using machine learning techniques. There were four stages in the work procedure: 1) Data collecting using internet-sourced images of everyday objects. It had 500 images of Wagyu. 2) The Keras Image Data Generator was used to increase image quantity by arbitrarily rotating, flipping, shearing, shifting, and zooming. To train models and provide data to trained computers, Support Vector Machine (SVM) and Convolutional Neural Networks (CNN) methods were used. 4) The model evaluation SVC model with polynomial kernel and RBF from Scikit-Learn was used for learned SVM training. Transfer learning was applied to VGG16 and Xception using the CNN model from the Keras API.**

Model evaluation results of Wagyu grade classification by using 20% test data shows that VGG16 with custom layers had the highest results, with (Accuracy %, F1-Score: A1, A2, A3, A4, A5) of (65, 92, 70, 55, 48, 67).

Keywords—Wagyu Beef, Image classification, Deep Learning, CNN, VGG16

I. INTRODUCTION

Fraudulent sales of lower-grade Wagyu meat as higher-grade Wagyu beef have occurred in Thailand. Most individuals inexperienced with wagyu beef may not be able to discern the quality of Wagyu beef based on appearance alone. Beef Color Score (BCS), Beef Fat Score (BFS), Beef Marbling Score (BMS), and Texture & Firmness Score are the four criteria specified by the Japanese Meat Grading Association's [1] quality standard for grading Wagyu beef.

After extensive investigation, any client with a camera and internet access will be able to assess the quality grade of Wagyu beef on their own. First, we gathered photographs of Wagyu beef in each grade for A1, A2, A3, A4, and A5 from one to five, cropped them to highlight the Wagyu steak in the image, and then categorized them into five, A1-A5 classes (100 images per class) based on the Japanese Meat Grading Association's standard. After dividing the data into 5 classes, we balanced and divided them into 80:20 train and test datasets.

II. LITERATURE REVIEW

A. CNN(convolutional neural networks)

From previous research [5], CNN was one of the best algorithms for image processing. To read and understood the image it uses RGB channels. An images were read with the form of Height x Width x 3, the 3 means 3 layers with the value of density for Red, Green and Blue for each pixel in the image. Then piped them through each layer to let the machine learn and recognize them.

B. Architecture.

a) Xception [2]

Figure 1 shows the architecture of the Xception model, which was a linear stack of depth-wise separable convolution layers with residual connections. This resulted in a relatively simple and easy-to-define and-modify design. As part of the Keras Applications, an open-source implementation of Xception utilizing Keras and TensorFlowis was offered. The data initially passes via the entering flow, then eight times through the middle flow, and finally through the exit flow. Note that batch normalization was applied to all Convolution and Separable Convolution layers (not included in the diagram). A depth multiplier of 1 is used in all Separable Convolution layers (no depth expansion).

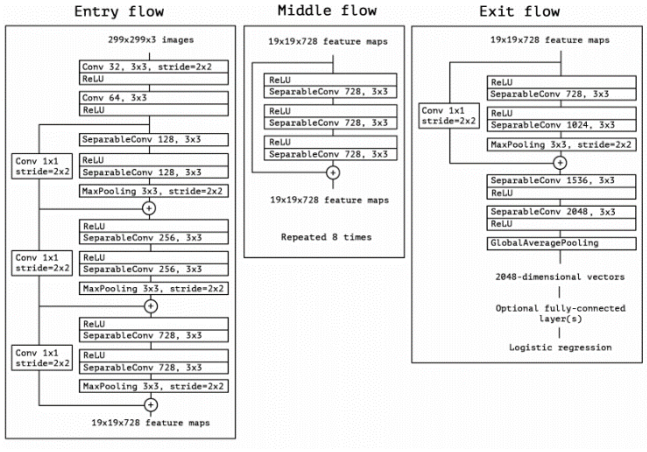


Fig. 1. The Xception's architecture

b) VGG16 [6]

The model consists of 5 groups of convolution layers, plus 1 group of fully-connected layer, along with 13 convolution layers and 2 fully-connected layers. Every convolution filter had a kernel size of 3x3 with stride of 1 and

pooling region of 2x2 without overlap. Figure 2 represent overall layers and images sizes used in VGG16. It's worth noting that we reduced the two 4096-dimension fully-connected layers to one 100-dimension fully-connected layer, resulting in a considerable reduction in complexity of the number of parameters.

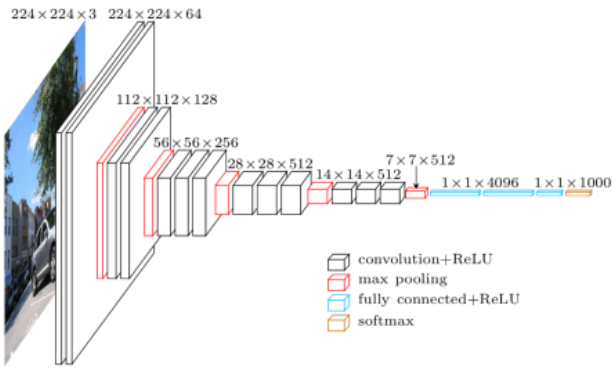


Fig. 2. A complete opened-up view of the VGG16 Architecture.

C. Transfer Learning

Deep Neural networks need a significant amount of time to train. Pre-trained models cut training time in half. To adapt to the data, pre-trained models employed weights that were repeated in layers. We plan to employ the VGG16 model and the Xception model, both of which are pre-trained models. These pre-trained models might be accessible using the Keras API to conduct picture categorization. [5].

D. From Classification of Beef Marbling by Image, machine learning was used to rate the quality of American beef marbling based on Classification of Beef Marbling by Image Processing [4]. With an average accuracy of 94.75 percent, SVM was used on black and white photos with controlled environments and backdrop eliminated.

E. Beef Quality Identification Using Color Analysis and K-Nearest Neighbor Classification [3], they demonstrated that under controlled conditions, machine learning could grade BCS, BFS, BMS, and "Texture and Firmness" beef quality under controlled environment.

F. A Feasibility Study of Beef Quality Assessment Employing Computer Vision and the Deep Neural Network (DNN) Algorithm [8], using simply the beef color score in a controlled setting. Table I shows the results of the evaluation.

TABLE I. EVALUATION OF BEEF QUALITY ASSESSMENT

Evaluation	Percentage
Precision	62.80
Recall	61.30
F1-Score	60.40

G. Texture Feature Extraction Based On GLCM and DWT for Beef Tenderness Classification [7], in a highly controlled setting, it was feasible to assess the quality score of "Tender and Firmness" photographs.

III. METHODOLOGY

To train a model to identify the grade of Wagyu beef, we must first create a dataset using photos of Wagyu steak and then train the models with datasets in order to evaluate their performance.

A. Data collecting

1) Dataset: Gather photographs from the internet in general without influencing the surroundings such as lighting or gadgets, then categorize and segregate evenly into five groups ranging from 1 to 5. Figure 3 shows several samples from the dataset.

Fig. 3. Example of Wagyu A5 and A3 images form dataset.

2) Data classifying

Because of the difficulties of scoring from low quality and uncontrolled surroundings in images, we decided to eliminate "Texture and Firmness" from the conditions for classifying the quality, table II.

TABLE II. MODELS AVERAGE ACCURACY

Requirements	Input quality (images)		Environment	
	Low	High	General	Control
BMS	X	X	X	X
BCS	X	X	X	X
BFS	X	X	X	X
Texture and Firmness		X		X

a) Beef Color Standard (BCS)

BCS was a criteria for scoring cattle based on the color of the muscle parts. Figure 4 shows the results, and table III shows the score with color range.

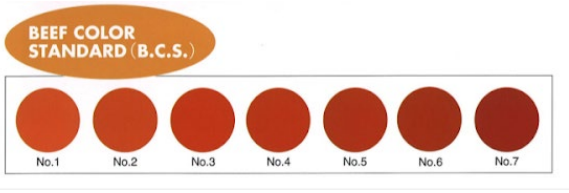


Fig. 4. Color chart of beef color standard with number tagged

TABLE III. BEEF COLOR STANDARD

Grade	BCS
5	3-5
4	2-6
3	1-6
2	1-7
1	Except 2-5

b) Beef Color Standard(BCS)

BFS was the criteria for scoring the beef based on the color of the fat component of the cattle. Figure 5 shows the results, and table IV shows the score with color range.

Fig. 5. Color chart of beef fat standard with number tagged.

TABLE IV. BEEF FAT STANDARD

Grade	BFS
5	1-4
4	1-5
3	1-6
2	1-7
1	Except 2-5

c) Beef Marbling Standard(BMS)

BS was the criteria for scoring the beef based on the muscle-to-fat ratio. Figure 6 shows the results, and table V shows the score range.

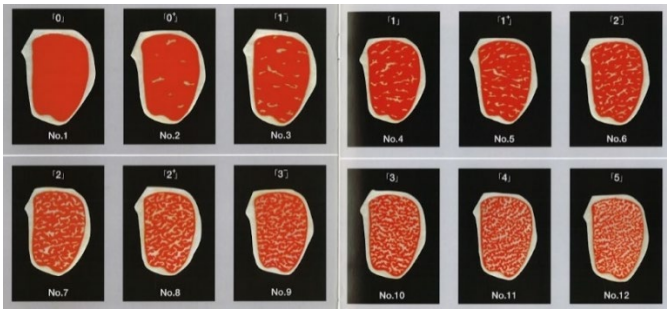


Fig. 6. Example visual chart of beef marbling standard with number tagged.

TABLE V. BEEF MARBLING STANDARD

Grade	BMS
5	No.8-12
4	No.5-7
3	No.3-4
2	No.2
1	No.1

B. Pre-Processing

1) To balance the data, divide each class's data into training (80%) and testing (20%) categories.

2) Generate more images for training with image datagenerator from Keras API with settings on table VI.

TABLE VI. IMAGES DATAGENERATOR SETTINGS

Actions	Parameters
Shear	0.2
Flip	True
Zoom	0.2
Shift	0.1
Rotation	90°
Reshape(SVM)	224x224x3
Reshape(VGG16)	224x224x3
Reshape(Xception)	299x299x3

C. Models training

To construct 10 weights, each algorithm must be trained 10 times with a random generate dataset. With the following settings: store best weight alone, early halt if validation loss does not improve for 20 epochs, decrease learning rate if validation loss does not improve for 10 epochs, learning rate initial at $1x10^{-4}$ with batch size of 32.

D. Models evaluating

To evaluate the models, four metrics, including accuracy, precision, recall, and F1 score, were calculated.

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN} \quad (1)$$

$$\text{Precision} = \frac{TP}{TP+FP} \quad (2)$$

$$\text{Recall} = \frac{TP}{TP+FN} \quad (3)$$

$$\text{F1-score} = 2 \times \frac{Precision \times Recall}{Precision + Reall} \quad (4)$$

Where TP denotes as true positive. TN denotes as true negative. FP denotes as false positive. FN denotes as false negative.

IV. RESULTS

A. Model comparison.

a) For each model trained and tested 10 times with random generated images on flow.

b) Table VII indicated that CNN can classify Wagyu beef better than SVM.

TABLE VII. MODELS AVERAGE ACCURACY

Models	Average Accuracy
SVM(Poly)	43.00%
SVM(RBF)	36.00%
Xception	54.00%
VGG16	54.10%
VGG16 adjusted	58.50%

B. Best weight evaluation.

a) On model VGG16 with adjusted dropout layers between fully-connected layers.

b) Table VIII showed that with just small number of general images with out environment controlled per class, 80 training images with 5 classes, CNN can reach up to 65% accuracy.

TABLE VIII. BEST WEIGHT ON VGG16 ADJUSTED

Quality grade	F1-Score
1	92.00%
2	70.00%
3	55.00%
4	48.00%

5	67.00%
Average	66.40%
Accuracy	65.00%

V. CONCLUSION

This study improved upon a previous one [8] that used just BCS from Malaysian and Australian beef standards by increasing the F1-score from 60.40 % to 66.40 % using BCS, BFS, and BMS from the Japanese Wagyu beef standard using a deep learning model.

It was established that extra datasets and a more robust architecture for deep learning may lead to improved results.

References

[1] American Wagyu Association 2021, accessed 5 November 2021, < https://wagyu.org/japan>

[2] F. Chollet, "Xception: Deep Learning with Depthwise Separable Convolutions," 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017, pp. 1800-1807, doi: 10.1109/CVPR.2017.195.

[3] K. Adi, S. Pujiyanto, O. D. Nurhayati and A. Pamungkas, "Beef quality identification using color analysis and k-nearest neighbor classification," 2015 4th International Conference on Instrumentation, Communications, Information Technology, and Biomedical

[4] L. Muhua, S. Jie and Z. Jinhui, "Classification of Beef Marbling by Image Processing," 2009 Asia-Pacific Conference on Information Processing, 2009, pp. 7-10, doi: 10.1109/APCIP.2009.138.

[5] S. Mascarenhas and M. Agarwal, "A comparison between VGG16, VGG19 and ResNet50 architecture frameworks for Image Classification," 2021 International Conference on Disruptive Technologies for Multi-Disciplinary Research and Applications (CENTCON), 2021, pp. 96-99, doi: 10.1109/CENTCON52345.2021.9687944.

[6] S. Liu and W. Deng, "Very deep convolutional neural network based image classification using small training sample size," 2015 3rd IAPR Asian Conference on Pattern Recognition (ACPR), 2015, pp. 730-734, doi: 10.1109/ACPR.2015.7486599.

[7] S. Widiyanto, Y. Sukra, S. Madenda, D. T. Wardani and E. P. Wibowo, "Texture Feature Extraction Based On GLCM and DWT for Beef Tenderness Classification," 2018 Third International Conference on Informatics and Computing (ICIC), 2018, pp. 1-4, doi: 10.1109/IAC.2018.8780569.

[8] W. K. Tan, Z. Husin and M. A. Hakim Ismail, "Feasibility Study of Beef Quality Assessment using Computer Vision and Deep Neural Network (DNN) Algorithm," 2020 8th International Conference on Information Technology and Multimedia (ICIMU), 2020, pp. 243-246, doi: 10.1109/ICIMU49871.2020.9243353.

Catalyzing Efficiency and Effectiveness: How Explainable AI Elevates Customs Brokerage

Supamas Sitisara
Graduate School
University of the Thai Chamber of Commerce
Bangkok, Thailand
supamas@thaisomdej.com

Assoc. Prof. Wanchai Rattanawong
School of Engineer
University of the Thai Chamber of Commerce
Bangkok, Thailand
wanchai_rat@utcc.ac.th

Asst. Prof. Varin Vongmanee
School of Engineer
University of the Thai Chamber of Commerce
Bangkok, Thailand
varin_von@utcc.ac.

***Abstract*—Customs brokerage holds a crucial position in international trade and logistics, enabling the seamless flow of goods across global borders. This process involves an intricate web of documents, information, and compliance demands, presenting a significant challenge for businesses. Embracing digitalization in documentation is a pivotal step toward enhancing customs brokerage operations. Modern technology in the "Easy Paste" model marks a paradigm shift in the customs brokerage industry. In this transformation, Explainable AI (XAI) serves as a valuable ally, elevating efficiency, reliability, and effectiveness, thereby providing businesses with a competitive edge with reasonable explanations and suggestions.**

Keywords—Customs Brokerage, Explainable AI (XAI), Easy Paste, Harmonized Code (HS Code), Structural Equation Model (SEM), Digital Logistics

I. INTRODUCTION

Customs brokers act as intermediaries between businesses and customs to perform customs clearance[1] . They are truly facilitators of several key functions and benefits to importers, exporters, and the overall supply chain including ensuring regulatory compliance with customs and other government agencies, minimizing costs, reducing risks, and providing valuable expertise and support [2]. Their services are essential for businesses engaged in global trade to navigate the complex and dynamic landscape of customs and border clearance procedures.

It is important for all stakeholders to work together to ensure that the process is well-organized and efficient. In order to cope with the dynamic changes and stay on top of competitors, the new paradigm shift of customs brokerage needs to be revolutionized and adjusted accordingly. Traditional paper-oriented processes need to radical and comprehensive shift to the full integration of electronic documents [3]. This evolution holds promise for increased trade volumes, improved customer satisfaction, and the establishment of customs brokerage as a vital driver of economic growth in the interconnected world. As the industry required to adjust to modern technology and advanced innovation, well-designed document preparation systems had a significant positive effect on customs clearance service capabilities. To overhaul existing processes and incorporate new technologies and innovations, digitalization of documentation is indeed an essential stage in improving customs brokerage operations. Since each step is significant and great effect on the operation of importation and exportation, therefore, reasonably explaining the required. The Explainable AI (XAI) can complement this transformation by enhancing efficiency and reliability which leads to competitive advantages [4]. Besides, with the new revolution of the customs brokerage document preparation process, legal and compliance verification, HS Code recommendations [5], license and permit recommendations, and checking for eligible privileges of the shipments are valuable services offered

II. LITERATURE REVIEW

Digital technologies not only accompany and facilitate material processes, but change the use of human resources, and, as a consequence, restructure their material, financial and information flows, can be described as "Digital logistics" [6]. Digital Logistics", the simplest definition of digitalization, is collecting information from silo operations to create a holistic view of each process from start to finish. All of this data is digitized to streamline logistics processes, strategies and systems. The ability of logistics services to deliver "the right amount of the product at the right place at the right time in the right condition at the right price with the right information" [7]. In the realm of customs brokerage, the seamless flow of information serves as a critical transaction that travels from the beginning to the end of the supply chain. The customs broker acts as the central repository for information and documents, which they must provide when requested by authorities. By applying a structural equation model (SEM), businesses can enhance the quality of their customs brokerage services and ensure smooth import and export customs clearance processes. This approach has positively impacted customs clearance capabilities and innovation capabilities, thereby improving outstanding performance efficiency and sustainable [8]

As the industry continues to embrace innovation, it is set to reshape the future of trade, elevating the customs clearance process into an indispensable cornerstone of the global economy. Well-designed document preparation systems had a significant positive effect on customs clearance service capabilities [9]. The Easy Paste System is purpose-built to facilitate document preparation for customs clearance, simplifying the management of meticulously organized and ready-to-use documents. Its exceptional value shines through when dealing with import and export declarations, as well as electronic drafts. In a fast-paced environment, the Easy Paste System streamlines the management to be well-organized and fully prepared documents and proves particularly valuable for handling import and export declarations.

The potential benefits of Easy Paste can be summarized as follows: (1) reduction in processing time, (2) reduction in costs, (3) increase in usability, (4) elimination of discrepancies and differences, (5) reduction in the effect of human factor, (6) decrease in the archive costs, (7) recording the economic activities, (8) prevention of fraud, (9) elimination of the complexity in payment methods, (10) reduction in the number of documents, (11) facilitating the acquisition of commercial information, (12) increase in global trade volume, (13) predictable costs and (14) eliminating the problems that stem from language differences. The leading benefits among them are the reduction of the number of documents and the elimination of complexity in payment methods [15].

III. RESEARCH METHODS

To create a novel system called "Easy Paste System". This intelligence system is specifically crafted to streamline the generation of import and export declarations. It operates by converting documents into machine-readable text through Optical Character Recognition (OCR), subsequently producing draft import and export declarations after extracting data and synchronizing it with the database.

A. Structural Equation Model (SEM)

A Structural Equation Model (SEM) is a statistical modeling technique that helps to analyze and understand complex relationships between multiple variables. SEM has been applied and implemented to improve the process [11]. Incorporating SEM into the Easy Paste System can lead to a more strategic and data-driven approach to optimization. By understanding the complex relationships and dynamics involved, customs authorities and stakeholders can collaborate to create efficient, accurate, and compliant customs clearance procedures that benefit trade and economic growth.

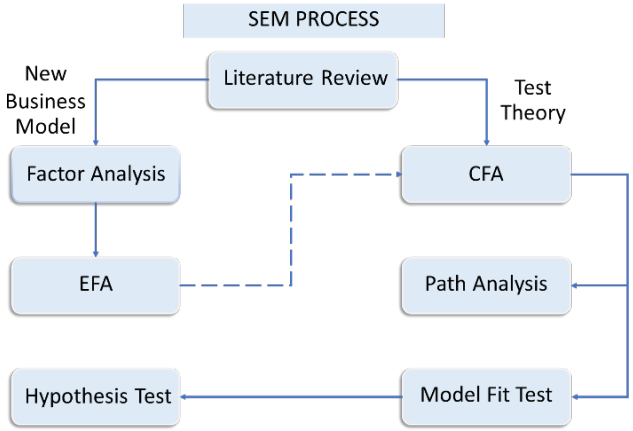


Fig.1 Structural Equation Model (SEM)

B. Data Collection/ Data Analyst

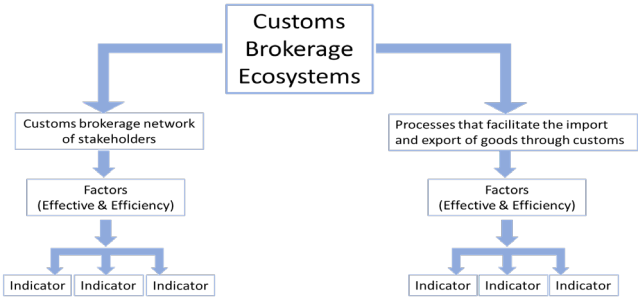


Fig.2 To find the factors and indicators

There are many factors and indicators that have great impact on customs clearance procedures such as

1)Regulations and Compliance: Customs brokerage involves navigating complex regulations and compliance requirements set by various governmental agencies. Changes in trade policies, tariffs, import/export laws, and customs procedures can significantly impact how customs brokers operate. Staying up-to-date with these regulations and ensuring compliance is crucial.

2)Global Trade Environment: Economic and geopolitical conditions, trade agreements, and international relations can influence the volume and nature of trade activities. Customs brokers need to adapt to changing trade dynamics and uncertainties, such as trade disputes, and sanctions events.

3)Technological Advances: Technology plays a vital role in customs brokerage. Advancements in automation, digitization, and data analytics can streamline operations, improve efficiency, and enhance customer experiences. Customs brokers need to invest in and adopt relevant technologies to remain competitive.

4)Security Concerns: Heightened security concerns in international trade have led to stricter security protocols and screening processes. Customs brokers must adhere to these measures while ensuring timely clearance of goods and minimizing delays.

5)Infrastructure and Logistics: Efficient supply chain management is essential in customs brokerage. The quality of transportation infrastructure, availability of ports, airports, and other entry points, as well as the overall logistics network, can impact the speed and cost-effectiveness of customs clearance.

6)Market Demand: The demand for customs brokerage services is influenced by trends in global trade and supply chain management. Expanding markets, new trade routes, and shifts in sourcing locations can impact the demand for customs clearance services.

7)Competitive Landscape: The customs brokerage industry is competitive, with various players offering similar services. Factors such as pricing strategies, service quality, reputation, and customer relationships can all affect a company's competitiveness and market share.

8)Customer Expectations: Clients expect efficient and transparent customs clearance processes. Customs brokers need to provide excellent customer service, timely updates, and accurate information to meet and exceed these expectations.

9)Currency Fluctuations: Fluctuations in currency exchange rates can impact the cost of goods, which in turn affects customs valuation and duty calculations. Customs brokers need to be aware of these fluctuations and how they might affect their clients.

10)Environmental Regulations: Increasing focus on sustainability and environmental regulations can impact the types of goods being traded and the documentation required for clearance. Customs brokers might need to consider environmental compliance as part of their services.

C. *Design System*

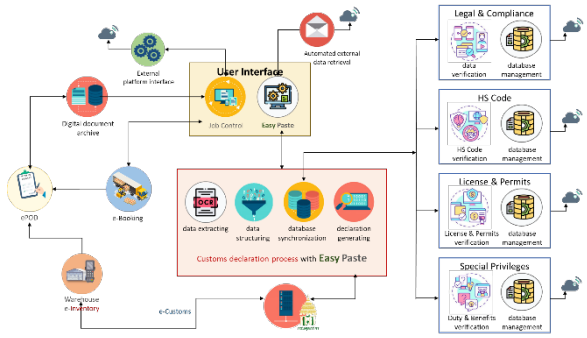


Fig. 1. Easy Paste Design System

To incorporate Easy Paste System with an Explainable AI system, enhanced by leveraging sophisticated systems such as 'Easy Paste,' customs brokers can not only ensure legal compliance but also benefit from recommendations for improved data validation during the clearance process [8].

For legal and regulatory compliance concerns, the recommended harmonized codes and privilege checks can be efficiently managed using Explainable AI for validation and recommendations. Explainable AI plays a pivotal role in suggesting HS Code classifications. These recommended harmonized code classifications are cross-referenced with the tariff classifications provided by the World Customs Organization (WCO), as indicated in the explanatory notes. Besides, tracking and tracing of the shipment can be easily monitored and automatically reported to each business stakeholder

D. *Artifical Intelligenct (AI) and Explainable Aritficial Intelligence (XAI)*

As technology continues to advance and businesses seek ways to navigate the complexities of international trade. The revolution in customs brokerage, driven by advanced technology and the integration of Artificial Intelligence (AI) and Explainable (XAI), is a significant leap forward in the world of international trade and logistics. The adoption of these cutting-edge technologies has ushered in a new era of efficiency, accuracy, and transparency in document preparation and customs clearance processes.

With the power of AI, customs brokers can automate and streamline the complex task of document preparation, ensuring that all required paperwork is not only meticulously processed but also compliant with ever-changing regulations. Furthermore, XAI adds an extra layer of transparency by offering clear, comprehensible explanations for AI-driven decisions, which is pivotal in building trust with clients and business stakeholders

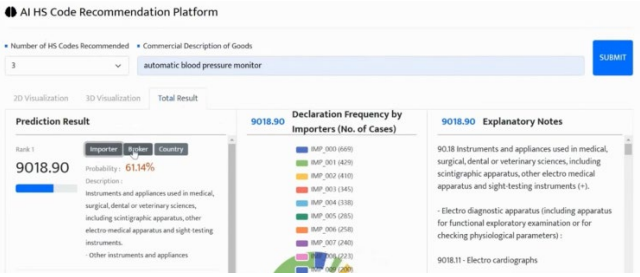


Fig.4 Explainable AI helps in recommending HS Code classification (Cr. The World Customs Organization)

This transformation not only enhances the competitiveness of customs brokerage firms but also paves the way for a more customer-centric approach. By providing clients with the benefits of rapid, error-free document processing and the peace of mind that comes from transparent decision-making, customs brokerage services have evolved to meet the evolving demands of a globalized world. The predictive monitoring is another opportunity involved that help in decision making.

These are how Explainable AI can contribute:

- **Automated Document Management:** XAI systems can automate the handling of documents, ensuring that they are correctly categorized, validated, and submitted to customs authorities. This reduces the risk of errors associated with manual document processing.
- **Transparent Decision-Making:** XAI provides transparency into the decision-making process of customs clearance. It can explain why a particular decision or recommendation was made, making it easier for customs brokers and clients to understand and trust the process.
- **Efficient Compliance Checks:** XAI can be programmed to continuously monitor regulatory changes and compare shipments against updated rules and requirements. This proactive approach ensures that all shipments remain in compliance with the latest regulations, reducing the risk of non-compliance.
- **Risk Assessment:** XAI can assess the risk associated with each clearance process by analyzing various factors, such as the nature of the goods, the trade partners involved, and historical compliance data. This allows customs brokers to prioritize their resources and attention on high-risk cases.
- **Real-time Updates:** XAI can provide real-time updates on the status of clearance processes, ensuring that customs brokers and clients are well-informed about the progress and any potential issues that may arise.
- **Bias Detection and Mitigation:** XAI systems can help detect and mitigate bias in decision-making. By analyzing the factors and data that influence clearance decisions, customs brokers can ensure that decisions are fair and unbiased.
- **Data-Driven Insights:** XAI generates insights from data that can be used to optimize clearance processes further. These insights may reveal patterns,

bottlenecks, or areas for improvement, allowing customs brokers to make data-driven decisions.

- **Customer Experience:** XAI can enhance the customer experience by providing clear and understandable explanations for clearance decisions. This transparency builds trust and confidence with clients.
- **Compliance Assurance:** XAI reduces the risk of compliance breaches by automating checks and validations against regulations and permits. This can prevent costly fines and delays.
- **Audit Trail:** XAI systems maintain a comprehensive audit trail of all clearance transactions, including the reasoning behind each decision. This is invaluable for compliance audits and investigations.
- Incorporating XAI into the customs brokerage process not only streamlines operations but also improves the accuracy, transparency, and compliance of the entire clearance process. This can lead to increased productivity, efficiency, and competitiveness in the customs brokerage business, ultimately benefiting both customs brokers and their clients.
- **Bias Mitigation:** XAI can play a vital role in identifying and mitigating bias in AI systems. By providing insights into the data used for training and the decision-making process, organizations can uncover and rectify bias-related issues, ensuring that AI-driven decisions are fair and unbiased.
- **Compliance and Regulation:** XAI contributes to improved compliance with regulations and standards related to fairness, transparency, and non-discrimination. This is especially important in industries where regulatory compliance is a significant concern.
- **Operational Efficiency:** XAI systems provide insights into the decision-making process, allowing organizations to identify areas for optimization and streamlining. This leads to improved operational efficiency and resource allocation.
- **Cost Savings:** Operational efficiency gains from XAI can result in cost savings by reducing manual efforts and improving resource

IV. CONCLUSION

Customs brokerage plays a vital role in international trade and logistics by facilitating the smooth movement of goods across international borders. The clearance process involves many documents, information, and compliance requirements, making it a very complex endeavor for business. Digitalization of documentation is indeed an essential stage in improving customs brokerage operations, and Explainable AI (XAI) can complement this transformation by enhancing efficiency and reliability which leads to competitive advantages.

REFERENCES

[1] H. Medin, "Customs brokers as intermediaries in international trade," *Review of World Economics,* vol. 157, no. 2, pp. 295-322, 2021.

[2] H. Medin, "Customs-brokers as facilitators in international trade," 2017.

[3] J. Rocha, and K. Lehenbauer, "Third Party Logistics and Beyond."

[4] P. Bhattacharya, D. Saraswat, D. Savaliya, S. Sanghavi, A. Verma, V. Sakariya, S. Tanwar, R. Sharma, M. S. Raboaca, and D. L. Manea, "Towards future internet: The metaverse perspective for diverse industrial applications," *Mathematics,* vol. 11, no. 4, pp. 941, 2023.

[5] M. Spichakova, and H.-M. Haav, "Application of Machine Learning for Assessment of HS Code Correctness," *Baltic Journal of Modern Computing,* vol. 8, no. 4, 2020.

[6] M. Pekarčíková, P. Trebuňa, M. Kliment, M. Edl, and L. Rosocha, "Transformation the logistics to digital logistics: Theoretical approach," 2021.

[7] K. C. Hu, and M. C. Huang, "Effects of service quality, innovation and corporate image on customer's satisfaction and loyalty of air cargo terminal," *International Journal of Operations Research,* vol. 8, no. 4, pp. 36-47, 2011.

[8] Ş. Ö. Ekici, Ö. Kabak, and F. Ülengin, "Improving logistics performance by reforming the pillars of Global Competitiveness Index," *Transport Policy,* vol. 81, pp. 197-207, 2019.

[9] C. Karabulut, and M. E. Civelek, "Hybrid-Paperless Adoption & Integration: Evaluotion of the Current Situation in Turkey," *Eurasian Business & Economics Journal,(2019) Doi,* vol. 10, 2019.

[10] M. Kilibarda, M. Andrejić, and V. Popović, "Research in logistics service quality: A systematic literature review," *Transport,* vol. 35, no. 2, pp. 224-235, 2020.

[11] D. N. Le, H. T. Nguyen, and P. H. Truong, "Port logistics service quality and customer satisfaction: Empirical evidence from Vietnam," *The Asian Journal of Shipping and Logistics,* vol. 36, no. 2, pp. 89-103, 2020.

[12] G. Harrison, S. M. Grant-Muller, and F. C. Hodgson, "New and emerging data forms in transportation planning and policy: Opportunities and challenges for "Track and Trace" data," *Transportation Research Part C: Emerging Technologies,* vol. 117, pp. 102672, 2020.

Poster Presentation(Session 1F ~ 2F)

3rd December 2023

Session 1F: Information Retrieval and Data Generation
- **Session Chair: Mr. Somchoke Ruengittinun, Kasetsart University, Thailand**
- **Date: December 3, 2023**
- **Room: Gallery, Time: 09:30 ~ 10:50**

Session 2F: Information Retrieval and Data Generation

- **Session Chair: Mr. Junyeop Kim, Kyungpook National University, Korea**
- **Date: December 3, 2023**
- **Room: Gallery, Time: 11:00 ~ 12:20**

Mascot Image Generation using StyleGAN2-ADA

Je Kyung Lee
dept. of Artificial Intelligence
Daegu University
Gyeongsan-si, Korea
jkzzang0203@naver.com

Jeoung Gi Kim
dept. of Artificial Intelligence
Daegu University
Gyeongsan-si, Korea
kjk9982097@gmail.com

Kyung-Ae Cha
dept. of Artificial Intelligence
Daegu University
Gyeongsan-si, Korea
chaka@daegu.ac.kr

***Abstract*—This paper presents a study on mascot image generation using generative adversarial networks. Using the StyleGAN2 adaptive discrimination augmentation algorithm, we create a GAN model that can diversify and regenerate styles from training data and use it effectively in mascot image design work. The mascot image generated by the model reduces production time by providing design ideas and working drafts. This study assesses the role of generative neural networks and verifies their utility in fields that require large-scale image or graphic production.**

Keywords—generative artificial intelligence, image-generative, style-generative adversarial network 2 adaptive discrimination augmentation (StyleGAN2-ADA)

I. INTRODUCTION

The generative adversarial network (GAN) [1, 2] is a technology that generates or transforms content, such as videos, images, and text, based on training data. In particular, because GAN uses unsupervised training, it can construct different training datasets depending on the purpose and derive converted output for all shapes, including human faces, animals, and landscapes. Therefore, it can be effectively used in areas that require the large-scale creation of images or graphics [3].

This paper experimented with generating a mascot image by training the StyleGAN2 adaptive discrimination augmentation (ADA) [4, 5] model, which is specialized for image-to-image conversion. Using a generation model, such as StyleGAN2-ADA, the user can design a mascot image that transforms a single character image into various styles.

In creative work, similar ideas often appear, which can make it challenging for individuals without design expertise or for novice artists to undertake such work effectively [6]. Therefore, in this paper, new images were created by training a GAN model to generate mascot images.

This approach provides design ideas to novice creators and saves production time. The generative artificial intelligence (AI) model implemented in this study used 3000 mascot images as training data, and the quality of the newly created mascot images was evaluated by comparing them with the original images. The performance of generative AI was verified through this method.

The paper is structured as follows: Section II discusses related research. Next, Section III outlines the process of constructing training data and model generation, and Section IV presents the experimental results. Finally, Section V concludes the paper.

II. RELATED RESEARCH

A. StyleGAN2-ADA

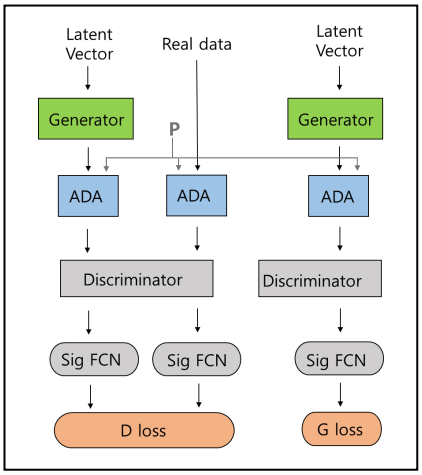


Fig. 1. Structure of StyleGAN2-ADA.

The GAN model is a neural network consisting of generator and discriminator. When the generator creates an image, the discriminator determines the generated image. This process is repeated to create high-quality images.

Within this context, StyleGAN2-ADA [4, 5] has emerged as an advanced model that incorporates the ADA mechanism into the discriminator. This strategic integration serves to mitigate the limitations observed in the original StyleGAN [7] model, resulting in a significantly improved model. Additionally, the augmentation capabilities of ADA offer the notable advantage of enabling the training of the model with a reduced number of images compared to the conventional requirements in traditional GANs. Fig. 1 depicts the structure of StyleGAN2-ADA.

In the typical scenario, training a high-quality GAN model necessitates a dataset of approximately 100 000 to 1 million images [7]. However, the present paper uses a variety of augmentation techniques, including cut0out, pixel blitting (blit), geometric transformation (geom), color, image spatial filtering (filter), and noise, within the StyleGAN2-ADA model. Using these techniques, we successfully trained a model using a dataset containing 3000 mascot images. This approach allows generating a diverse range of mascot images.

B. Emoticon Image Production Study

Zhao and Kim [8] addressed the creation of smartphone emoticon images. The authors analyzed the current status of the emoticon image industry as a basis for the development of new emoticon designs. In the process of creating these new designs, the character's storyline was established, and corresponding character sketches were created. Afterward,

Corresponding author: Kyung -Ae Cha, chaka@daegu.ac.kr

Fig. 2. Example of training images

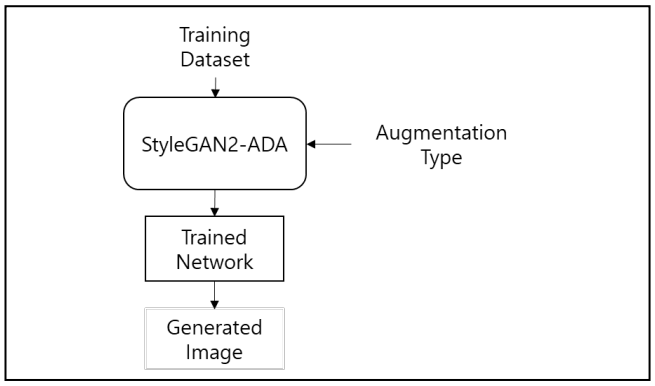


Fig. 3. Image generation artificial intelligence diagram

TABLE I. DEVELOPMENT ENVIRONMENT FOR STYLEGAN2-ADA

CPU	11th Intel(R) i9-11900k
RAM	64.0GB
GPU	NVIDIA RTX 3080Ti
Anaconda Python v.	3.9.0
CUDA Engine v.	CUDA 11.3
PyTorch v.	11.3

coloring work was conducted, and design tools were used to create emoticon animations and tables. As explained, creating emoticon images through a manual process takes a significant amount of time and effort. In the present study, we aim to use GAN for mascot image generation to streamline the workflow and shorten the production time. Additionally, this model has the potential to stimulate the creation of new ideas by providing a variety of mascot images.

III. MASCOT GENERATION AI MODEL

A. Construction of Training Data

The training dataset primarily comprises editable materials with copyright notices. The material is predominantly mascots and characters related to local governments, sports, universities, and other organizations. Additionally, we used freely available emoticon images from platforms like Pixabay [9] and Freepik [10] to construct the training dataset. A total of 15 000 emoticon images were collected, from which images with frontal facial expressions were selected. For training, each image was converted to 256x256 pixels in red, green, and blue (RGB) (24-bit) format. After data refinement, the final training dataset consisted of 3000 images.

Collecting a sufficient number of training images to suit the intent of a project can be very difficult. However, by taking advantage of StyleGAN2-ADA, we effectively trained the model with only 3000 images, which is approximately 1/50 of the number of images typically required for traditional GAN training. The specifications of the mascot image are illustrated in Fig. 2, depicting a static, front-facing image measuring 256x256 pixels in RGB (24-bit) color format.

Fig. 4. Example of the first experiment result.

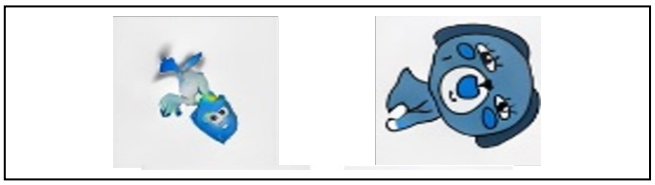


Fig. 5. Example of the second experiment result.

TABLE II. STYLEGAN2-ADA TRAINING PROCESSES

Data set	Training amount	Count of augmentation type	Time
15 000	1000kimg	3	~22 h
7000	1000kimg	3	~12 h
3000	4000kimg	6	~10 h

B. Style GAN2-ADA Training

The structure of the mascot image generation AI implemented in this paper is displayed in ADA model involves collecting and refining a training dataset containing images matching the mascot creation intent. The final output image features creative transformations, as AI transforms the image through a model learned from the initial training dataset. Table I presents the training model environment using Python and CUDA Engine/PyTorch on a graphics processing unit (GPU).

The StyleGAN2-ADA model is a generative AI, which trains on the entire training dataset and transforms and combines styles to create new images. Training is equivalent to the concept of epochs and is performed in Kimg units, representing the number of training repetitions.

The value of Kimg determines how many images the discriminator evaluates in one step, and training continues until this value is reached. In addition, 1Kimg means 1000 repetitions, and the generated images are generated at intervals of 100Kimg (100 000 repetitions) [11].

To obtain the final learning model, we inspected the generated images, and images that did not meet the mascot specifications were refined to reconstruct the training dataset. This process was repeated three times. Table II represents the training process.

The first training proceeded for 1000Kimg (1 million iterations), and the ADA method incorporated pixel blitting (blit), geometric transformation (geom), and color techniques. The training process for image generation took approximately 22 h. Fig. 4 represents the image results obtained from the first training. The training results included a significant number of green images, and distorted images were observed due to the application of special effects.

In the second training, the training proceeded for 1000K img (1 million iterations), and the ADA method incorporated pixel blitting (blit), geometric transformation (geom), and color techniques. Following the first training, a refinement process was conducted to remove green images and images containing

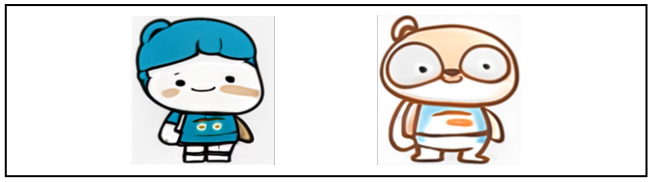


Fig. 6. Example of the final experiment result.

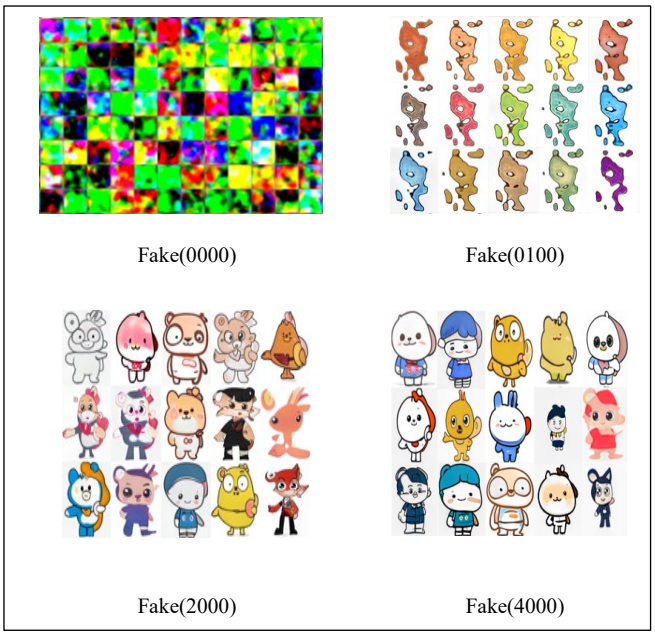


Fig. 7. Creation process for each step.

special effects or text. The second training was conducted using 7000 images that had undergone refinement, and the training process took approximately 12 h. Fig. 5 presents the results of the generated images. The phenomenon of distortion in the appearance of faces and bodies occurred due to the simultaneous training of both two-dimensional (2D) and three-dimensional (3D) images. Additionally, images were generated in rotated forms.

In the final training, the training proceeded for 4000Kimg (4 million iterations), and the ADA method included pixel blitting (blit), geometric transformation (geom), color augmentation, image space filtering (filter), noise insertion (noise), and cutout techniques. These additions were made to correct the distortion problems observed in the second training and produce images with different colors. Additionally, to reconstruct the training dataset, we created a dataset of 3000 images by removing blue images, 3D images, and those that did not fit the intended pose. Due to the reduction in the size of the training dataset, it was expected that the training time would decrease despite increases in training volume and ADA. Thus, training options expanded. The training process took approximately 10 h.

Fig. 6 represents the results of the final experiment. The created image is similar to a mascot image that can be used as a mascot for local governments or public enterprises.

Specifically, we incorporated the ADA technique, resulting in the generation of diverse styles of images with accurate representations of facial shapes and positions of facial features. Furthermore, these generated images maintain the same form as the initial training images. However, they are

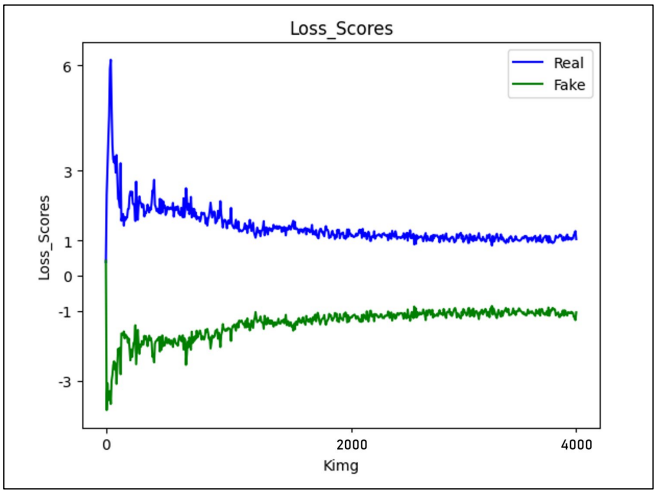


Fig. 8. Real and fake score graphs.

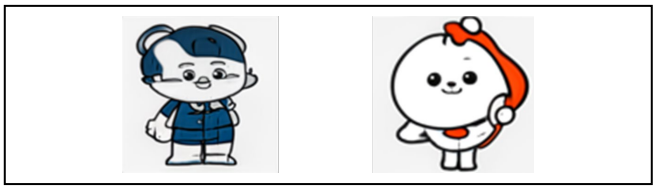


Fig. 9. Examples of generated images.

transformed into character images with distinguishable styles, demonstrating the utility of these creatively generated images.

IV. StyleGAN2-ADA Experiment Results

Fig. 7 depicts the step-by-step process of image generation observed during the training procedure. The image labeled Fake0000 represents the noise image applied during training.

The intermediate images generated during the training process are similar to the images labeled Fake0100 and Fake2000, and the images generated after 4000Kimg of training are identical to the image labeled Fake4000. The images generated by the training model are in a format suitable for use as emoticon images.

The GAN model is an algorithm in which the generator and discriminator compete with each other. The loss value changes, and after the initial training, it converges to the target value within a specific range. Adversarial training is conducted through this process [12]. The discriminator detection score graph of the final trained model of StyleGAN2-ADA is depicted in Fig. 8, where it is evident that after passing through the initial training phase (about 300Kimg), the loss values stabilize within a specific range, converging to the target value. The result presented form Fig. 8 confirms the successful progression of adversarial training. The fake score exhibits low values, whereas the actual score curve consistently maintains higher values, indicating effective discrimination between real and fake images.

This study compared real emoticon images to assess the utility of the generated images. Fig. 9 represents the emoticon images generated in this study. The generated emoticon images emphasize the face and distinctly depict facial features, showcasing their characteristic traits. However, shortcomings exist regarding representing the intricate details and special effects of emoticon images. Nevertheless, through these generated images, we can provide ideas and initial design drafts to novice artists and the public.

V. CONCLUSIONS

This paper constructs a generative AI model that automatically creates numerous character images used in emoticon production, reducing the cost and time. The StyleGAN2-ADA model was employed, generating various creative images that can be used in real emoticon creation. Furthermore, in the experiments, we refined the training data and verified their usability to create an AI model that maximally accommodates the designers' scope of work. The experiment demonstrated that cost and time in emoticon production can be reduced and that new ideas and initial designs can be provided to novice artists and the public. This outcome illustrates the potential for utilization in emoticon production and the creative automatic generation of emoticon images.

In future work, we plan to actively employ rapidly advancing GAN models to generate emoticon images in the frontal view and various other aspects, implementing them with improved performance.

REFERENCES

[1] A. Radford, L. Metz, and C. Soumith, "Unsupervised representation learning with deep convolutional generative adversarial networks," Computer Science, Machine Learning, arXiv preprint arXiv:1511.06434, 2015.

[2] K. Wang, et al., "Generative adversarial networks: Introduction and outlook," *IEEE - CAA J. Automatica Sinica*, vol. 4, no. 4, pp. 588–598, 2017.

[3] M. Cui, M. Kim, S. Choi, and S. Lee, "The usage and impact of GAN in graphic design," *Archives Design Res.*, vol. 35, No. 4 pp. 285–307, 2022 (in Korean).

[4] C. Liao, M. Sawayama, and B. Xiao, "A perceptual evaluation of the StyleGAN2-ADA generated images of translucent objects," *J. Vision*, vol. 22, no. 14, pp. 3642–3642, 2022.

[5] Y. Lu, "Style-based image manipulation using the StyleGAN2-Ada architecture," *Proc. 4th Int. Conf. Comput. Data Sci.*, 2022, pp. 29–37.

[6] J. G. Jeong, et al., "Study on copyright protection measures in response to new technology," Korea Copyright Protection Agency, Seoul, Mapo-gu, Feb. 2022

[7] T. Karras, S. Laine, and T. Aila, "A style-based generator architecture for generative adversarial networks," *Proc. IEEE/CVF Conf. Comput. Vision Pattern Recognit.*, 2019, pp. 4401–4410.

[8] Q. Zhao and D. H. Kim, "A survey of the commercial nature of smartphone emoji," *Asia Culture Aca*demy o*f Korea Citation Index*, vol. 10, no. 5, pp. 17–28, 2019 (in Korean)

[9] Pixbay, [Online]. https://pixabay.com/ko. [access: September 19, 2023]

[10] Freepik, [Online]. https://www.freepik.com. [access: September 19, 2023]

[11] T. Karras and J. Hellsten, "StyleGAN2-ADA-Pytorch," [Online]. https://github.com/NVlabs/stylegan2-ada-pytorch/. [access: September 19, 2023]

[12] D.A. Talib, and A.A. Abed, "Real-time deepfake image generation based on StyleGAN2-ADA," *Revue d'Intelligence Artificielle*, vol. 37, no. 2, pp. 397–405, 2023

Virtual Idol Design Based on existing Game Characters - Focus on Character Setting and Personality

1st Yin Lyu
Student, Dept. of Design
Dongseo University
Busan, Korea
lyuyin1213@gmail.com

2nd Yong-Ki Lee
Professor, Design
Dongseo University
Busan, Korea
ykl177@gdsu.dongseo.ac.kr

***Abstract*—The development of internet capital and digital technology has driven the growth of the gaming industry and has particularly propelled an increasing number of game characters toward becoming virtual idols. From the initial character, Shiho Fujisaki, in "Tokimeki Memorial" to the K/DA girl group developed by Riot Games, more and more game characters are evolving towards the direction of virtual idols. The transformation of game characters into virtual idols, with their unique artistic styles and business advantages, represents a new direction in game IP development. As a product of the fusion of gaming culture and idol culture, game virtual idols permeate the cultural preferences of the times and their target audience. This article aims to delve into the design principles of transforming game characters into virtual idols. In the process of transforming game characters into virtual idols, it is essential to preserve their original characteristics. However, these character settings should also be combined with elements of idol culture to meet the demands of the virtual idol market. This paper explores the design of game virtual idols from various perspectives, providing a clear design framework for the idolization of game characters, and also anticipates future research directions.**

Keywords—Virtual Idol，Game Character，Concept Design，Character Transformation，Character Design Guideline

I. INTRODUCTION

With the development of the internet economy and digital technology, an increasing number of gaming companies are idolizing game characters. "Idolization" refers to the trend of mining the positive traits of a character to make them admired by people [1]. When referring to the "idolization" of game characters in this article, it means a transformation of a character from a game persona into an idol, involving a second round of development for the game character to expand their business direction, promoting the game character as a virtual idol. The transformed game character will engage in activities as a virtual idol, having their own social media accounts, fan groups, music works, brand collaborations, and more. The transformation of game characters into virtual idols, with inherent fan advantages and promotional platforms, has captured the attention and adoration of the teenage demographic, becoming one of the methods for game companies to develop their intellectual properties. As the main producers of entertainment content, virtual idols from games such as K/DA and the League of Legends virtual group have actively engaged in the online sphere through concerts, music video releases, participation in variety shows, and more. Through the idolization and development of game characters, a new form of fan-driven economy has emerged, attracting significant societal attention. Research into the idolization of game characters primarily focuses on the following aspects: firstly, the development and advantages of game virtual idols; secondly, the design strategies for transforming game characters into virtual idols; and thirdly, the future prospects of idolizing game characters. These studies explore the relationship between game characters and idolization from diverse perspectives. Based on this, the research employs methods such as literature analysis and case studies to shed light on this entertainment phenomenon in the digital age. It aims to provide insights and guidance for the design of game character virtual idolization.

II. THE PROCESS OF IDOLIZATION OF GAME CHARACTERS

A. The emergence of the process of idolization of game characters

In 1994, the development of the game industry provided the soil for the emergence of many virtual idols. Shiori Fujisaki, the heroine in the classic game "Heartbeat Memories", is the goddess in many people's minds. Shiori Fujisaki is a BOSS-level character in "Heartbeat Memories". But in an era when there are no game guides, many players need to try multiple times to achieve the desired results. This attempt made many players develop deep feelings for her. Subsequently, Konami established an official support club for Fujisaki Shiori and published a total of 33 official quarterly magazines. Shiori Fujisaki has also independently released several albums under the name of virtual idol. On October 31, 1996, the PC and PS versions of "Tomb Raider" were released at the same time, and were widely praised in the gaming circle. Among them, Laura's

[1] Lu Jiayi, Liu Zhixin, and Liu Jingfu, "The phenomenon of 'idolization' in the new media era," ShengPingShiJie, May 2023, pp. 89-91.

This research was supported by BK21 Four Service Design driven Social Innovation Educational Research Team in Dongseo University.

fashionable appearance also inspired many companies to promote her. The fashion media's pursuit of Lara Croft and "Tomb Raider" started with "FACE". In June 1997, Britain's most famous fashion magazine made a long special report on "Tomb Raider" and used Lara Croft's CG image as the cover of the magazine. In 2015, the game "Ensemble Stars" was officially launched in Japan. This game has received unanimous praise from many female players since its launch. In 2017, "Ensemble Stars"'s first concert presented with 3D character holograms was held at Tokyo's largest LIVEHOUSE. In July 2018, the Chinese game "Onmyoji" announced the virtual idol project "Onmyoji Idol Project", announcing that the Onmyoji IP has opened a new exploration in the field of virtual idols. The character Daitengu from the game "Onmyoji" also became the first game star to perform on stage. In terms of cultivating virtual idols, there are also idol development projects such as "Love Live!" and "Uma Musume Pretty Derby" to achieve real-time interactive gaming experience. In 2018, the game "League of Legends" turned its popular characters into a virtual idol group, the K/DA girl group. They appeared at the opening ceremony of the global finals and sang the song "POP/STAR", which attracted public attention. The popular Chinese game "Honor of Kings" is a virtual boy group Unlimited Kings, also positioned as a virtual singer group. The group debuted in April 2019 and released its first music album in May. The group song is "Wake Me Up". At present, game characters take virtual singers as their profession, and most of them perform through singing, music videos, offline performances, etc. Creating virtual idols has not only become one of the important ways for the entertainment industry to keep up with the trend but has also become a new outlet for industrial IP development to pursue profits.

B. Advantages of the process of idolizing game characters

1) Fan base: With the development of 5G technology and mobile games, it can be said that games have become one of the forms of entertainment that can be played anytime and anywhere in people's daily lives. The design and operation of a mature virtual idol takes a long time. It is a long process from birth to popularity and even profitability. Take the game "Onmyoji" as an example. The gorgeous and exquisite conceptual style gives the game IP huge commercial value and untapped potential. "Onmyoji" continues to strengthen the emotional connection with the audience and strengthen the audience's loyalty to the game IP through online and offline activities such as fan creation contests, fan support activities, and Onmyoji-themed COS. The initial setup of a virtual idol based on a game character does not require much effort. By screening the game audience, the company can roughly determine which character IP is more distinctive and popular. On the other hand, the game has a fan base that ordinary virtual idols cannot match. A popular game character with a strong and loyal fanbase. Coupled with the resources and funds of the large companies behind the game, it provides a material basis for the process of idolization of game characters, allowing game virtual idols to be guaranteed by publicity platforms and commercial resources, standing at a higher starting line than other virtual idols.

2) Design diversity: The extremely popular game characters have passed the audience's screening. These characters have become ingrained in players' minds as images with labels. Under the construction of the game's worldview, many game characters are often created. The overall game style is consistent, but the details of creation are differe nt and each has its characteristics. Take the game "League of Legends" as an example. K/DA, who has been transformed into a virtual idol group, are just five popular female characters in the game. As of now, "League of Legends" has more than 140 game characters, including Seraphine, known as the Star Singer, and Ahri, who is based on the nine-tailed demon fox. Different excellent character images and character settings make game characters extremely popular and developable. Similarly, in the game "Honor of Kings", Li Bai, Han Xin, and Zhao Yun, who were designed based on Chinese historical figures, all turned to virtual idols for development. In the entire game, "Honor of Kings" has a total of 113 game characters. Each hero has super popularity and plasticity, as well as rich development resources. With the creation of the idol project of game characters, it is foreseeable that more game characters can be created as virtual idols of different professions and styles in the future, creating long-term vitality for virtual idols.

3) Interaction & Communication: Jane McGonigal mentioned in "Reality is Broken: Why Games Make Us Better and How They Can Change the World" that online games include game activities, and game activities include social networking. The game works hard to make the player feel both emotional and cognitive empathy toward the characters[2]. People have common emotions in the game's interaction and strengthen their bond with each other. This effect also empathizes with the relationship between the player and the game characters. Players come into contact with the game through the perspective and actions of the game characters. The game characters always accompany the players during the game, feeling, experiencing, and growing together with the users in the game world. Take the MOBA game "League of Legends" as an example. When gamers use specific game characters to fight for a long time, they will have emotions for the game characters and substitute them unconsciously. From the perspective of players, regard it as a close friend who fights together and has a close relationship. Game manufacturers attach great importance to the connection between games and virtual idols, arouse the audience's feelings for game characters, and constantly create and strengthen the emotional bond between players and game characters under empathy. The development of virtual idols of game characters has enabled game characters with a strong foundation of emotional companionship to arouse better and deepen the intimate connection with the audience.

[2] Stef Craps and Toby Smethurst，“Playing with Trauma Interreactivity, Empathy, and Complicity in The Walking Dead Video Game,” Games and Culture, vol. 10(3), pp. 269–290, 2014

III. DESIGN FEATURES OF GAME CHARACTER IN THE PROCESS OF IDOLIZATION

A. Retention of original settings

Under the highly interactive and interactive experience of the game, the game characters have left a deep impression on the players, and subtly, the prototype of the game character design has already been formed in the players' minds. On this basis, every work and every activity of virtual idols who play IP roles are more or less intentionally connected with the emotion and spirit of the game. K/DA member Ahri, for example, became a virtual idol for a very popular character in the League of Legends game. Ahri's costume design has been adjusted to a certain extent, but the huge fox ears and tail shape are retained. The concept design of Kaisa, who is also a member of K/DA, in the game is that the entire body is corroded, and the weapon behind it is transformed from the skin. While Kaisa was transformed into a virtual idol, the weapon as an important feature was also retained. In 2020, photos of K/DA's collaboration with fashion brand Louis Vuitton appeared in the DAZED magazine. While wearing modern clothes, the features of the game are also shown in the photos.

The first key insight is that, during the process of character idolization in gaming, it is essential to preserve the character's iconic design features from the original game. Whether in virtual performances, merchandising, or any other form of media, the character's core appearance, attire, unique abilities, or personality traits must be retained and accentuated. This preservation not only aids players or audiences in swiftly recognizing and associating with the original game during the process of idolization but also upholds the respect and loyalty towards the game character, along with the potential brand value. Therefore, in the successful process of idolization of game characters, ensuring the preservation of these iconic design features is of paramount importance, as they constitute the pivotal elements for conveying character identity and allure.

Fig. 1. K/DA girl group

B. The idolized extension of the character setting

The functions of game characters and virtual idols are different, so when the game characters are transformed into virtual idols, the original settings need to be innovated. The main purpose of game characters is to provide players with an operable goal in the game to complete the game's tasks. The setting of the characters is often more in line with the background setting and gameplay needs of the game itself. Virtual idols focus more on entertainment, social interaction, and interaction. Therefore, to successfully transform game characters into virtual idols, the original settings must be innovated to meet the management of idol activities. Take Akali in the K/DA group as an example. In the original game setting, Akali is a ninja who can use magic. In the process of her idolization, the designer changed Akali's setting to a fifteen-year-old girl whose family is a martial arts dojo and added her rap skills, which are in line with the role of virtual idols in performing arts activities and artistic needs. At the same time, Akali has added the settings of eating Shin Ramen and being able to speak multiple languages, which makes her character richer and fuller, with humanized characteristics, and shortens the psychological distance between her and the audience.

The second key insight is that when transforming game characters into virtual idols, it is essential to innovate upon the original character setting to cater to the demands of entertainment, social interaction, and audience engagement. This innovation is aimed at creating virtual idol characters that are more relatable and attention-grabbing, making them resonate more effectively with the audience.

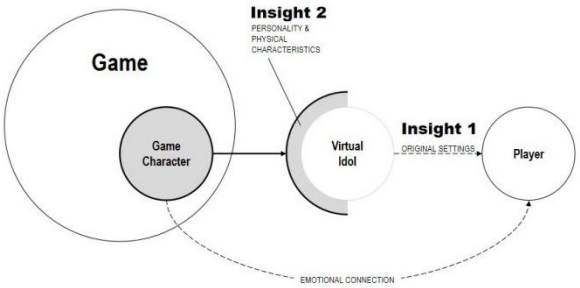


Fig. 2. Key insights to the application of virtual idols

IV. GUIDELINES AND FUTURE OUTLOOK

Design guidelines for transforming game characters into virtual idols:

1) Preserve Core Identity: Retain key elements of the original character design that have left a lasting impression on players, even while adapting them for the process of idolization. For example, significant visual traits like iconic accessories or features can be retained to maintain continuity and recognition.

2) Innovate to Suit Idol Activities: Recognize that game characters and virtual idols serve different functions, with virtual idols emphasizing entertainment and social interaction. Therefore, be prepared to innovate and adapt the character's settings to suit the demands of virtual idol activities, such as live performances and interactions with fans.

3) Humanize and Enrich Characters: Enhance the depth of virtual idols' characters by humanizing them and giving them relatable qualities. This can include adding hobbies, interests, or personal characteristics that make the virtual idols more three-dimensional and relatable to the audience.

These design guidelines emphasize the importance of balancing the preservation of key character elements with the

need for innovation and adaptation to create compelling virtual idols that resonate with both fans of the game and new audiences in the realm of entertainment and social interaction.

Currently, the fanbase for virtual idols based on game characters is predominantly composed of teenagers. With the advancement of technology, the interaction between virtual idols and their fans will become even more intimate and personalized. Augmented reality, virtual reality, and other technologies enrich the gameplay and content of games, while continuous improvements in-game content and environments enhance the sustainability and effectiveness of games, catering to a wider audience. The interactivity and social nature of games offer opportunities for new collaborations with industries such as education, generating potential positive impacts.

Educators can create interactive educational experiences within gaming environments, allowing students to practice in virtual scenarios, reducing waste generation, and contributing to sustainable resource management. The development of artificial intelligence provides virtual idols with opportunities to continuously expand their skill sets and knowledge domains. Virtual idols' appearances and personalities can be customized to cater to students with different personalities, thereby improving educational efficiency.

The fan effect of virtual idols also benefits charitable endeavors and other causes. Virtual idols based on game characters can create exclusive virtual events and gaming products with charity themes. For example, introducing charity skins, charity game levels, or virtual goods. Virtual idols can motivate fans to participate in games and purchase products, thereby supporting charitable causes.

V. Conclusion

Virtual idols are products of technology and the era. Virtual idols based on game characters have advantages over traditional virtual idols. In the process of transforming game characters into virtual idols, it is essential to maintain their original appearance and style settings to keep fans feeling familiar and close. During the idolization process, game characters also need to be innovative and have additional characteristics and personified personalities, making them more idol-like to meet the demands of idol activities. With the continuous development of technology and the growing fan base, virtual idols based on game characters have more development possibilities, injecting new vitality into the future of digital entertainment and societal progress.

Acknowledgment

I would like to express my sincere gratitude to Professor Yong-Ki Lee for his invaluable guidance and support throughout the course of my research and writing. His profound insights, patient explanations, and unwavering dedication to academic excellence have left an indelible mark on my scholarly journey. Without his expertise and mentorship, completing this research would have been a far more challenging endeavor.

In addition, I wish to acknowledge the financial support provided by the BK21 program. This generous funding not only facilitated my research efforts but also granted me access to essential academic resources and opportunities for intellectual growth. I deeply appreciate the crucial contributions made by Professor Lee and the BK21 program to my academic pursuits. I will cherish these opportunities and the trust bestowed upon me as I continue my scholarly endeavors. Once again, thanks sincerely.

References

[1] Lu Jiayi, Liu Zhixin, and Liu Jingfu, “The phenomenon of ‘idolization’ in the new media era,” ShengPingShiJie, May 2023, pp. 89‑91.

[2] Stef Craps and Toby Smethurst, “Playing with Trauma Interreactivity, Empathy, and Complicity in The Walking Dead Video Game,” Games and Culture, vol. 10(3), pp. 269‑290, 2014.

[3] Guga. Jelena, “Virtual idol hatsume miku new auratic experience of the performer as a collaborative platform”, New York: Springer, 2015, pp. 36-44.

[4] Kadner. Noah, “The virtual production field guide,” Epic Game, 2019, pp. 24-29.

[5] Hamilton, Robert, “Virlual idols and digital girls: artifice and sexuality in anime,” Kisekae and Kyoko Date, 1997.

[6] Daniel Black, “The virtual ideal: virtual idols, cute technology and unclean biology,” Continuum, vol. 22(1), 2008.

Interactive Virtual Museum Exhibit Design

Le Xuan Tan
Department of Computing and Information Systems
Sunway University
Petaling Jaya, Malaysia
20037008@imail.sunway.edu.my

Yunli Lee
Department of Computing and Information Systems
Sunway University
Petaling Jaya, Malaysia
yunlil@sunway.edu.my

***Abstract*—Traditional Museum exhibitions often lack interactivity between visitors and exhibits, leading to visitor boredom and limited understanding of the exhibit content. In Malaysia, a country which are renowned for its rich cultural diversity and historical significance, an interactive virtual museum exhibit that are implemented by using Augmented Reality (AR) technology has been created to remedy this issue. Global visitors can access an immersive experience by exploring and engaging with virtual exhibits through their screens in this virtual museum. It features an intuitive user interface and a range of interactive features that stimulate multiple senses simultaneously with a multimedia mix by using AR techniques including interactive components (touch), visual element like images and video (sight) and auditory components (hear) to enhance visitor's interest and their overall exhibit experience. The use of AR technology is poised to revolutionize visitor interaction with exhibits, providing a more engaging and educational experience.**

Keywords—Interactive Exhibit, Virtual Museum, Design, Augmented Reality

I. INTRODUCTION

One common duty that has existed throughout history and throughout all cultures is to preserve and pass on traditions, culture, and heritage that capture the spirit and legacy of our ancestors to the next generation. The primary means of preserving and transmitting heritage is through academic instruction. Due to the outbreak of Covid-19 and the implementation of social isolation measures, many schools encountered difficulties in organizing field trips to museums. Besides, many museums' traditional methods of teaching do not successfully capture and interest visitors. In response to these modern demands, some museums have recently started using Augmented Reality (AR) techniques. This is due to utilizing AR technology effectively not only overcomes time and space limitations, but also gives visitor fresh perspective on cultural artifacts. By incorporating virtual data into a real-world setting, AR gives a more immersive feeling of reality than VR, which is limited to the application itself [1].

With the rapid development, AR technology has the potential to address issues with customization and importance in museum design [2]. AR are promising technologies that have the potential to significantly impact a variety of fields, including some that are not often connected with computer technologies and one of these areas is cultural heritage [3]. With AR, there is an exciting impression that actual and virtual items coexist in the same area. The ability to combine a user's actual environment with virtual material is a unique possibility provided by AR technology. More realistic interaction and visuals are now possible in AR applications, making AR experience more immersive and "believable" due to the advancement of computer graphics and computer vision technology [4].

AR has been tested at museums since the early 2000s. The AR technology, which is most frequently used in conjunction with mobile devices, overlays virtual content onto the immediate physical surroundings and shows the two as a single image on a screen. As a result, AR has been seen as a tool for innovation in the hands of museum experience developers to improve interactions between visitors, collection objects, and their contextualized information. The power of this technology is in combining the interpretational and observational aspects of experiencing an exhibition [5].

This project is about developing an interactive exhibit for museum by using AR technology to enhance visitor experience by increasing interactivity between users and exhibits. The entire development process is consolidated into a single API, which is Unity. The outcome of this project is a virtual museum application that utilized AR technology to enhance the exhibit's interactivity.

II. LITERATURE REVIEW

A. Augmented Reality

Augmented Reality (AR) involves adding digital elements, such as images or videos, to the real-world using devices like smartphones, eyeglasses, or desktops. This process includes tracking real-world surroundings or objects and placing virtual elements on them. AR enhances sensory experiences, allowing users to view objects in a different context [6]. It acts as a technology bridging the gap between real and virtual objects by overlaying virtual elements onto the actual world [7]. This integration of digital images, graphics, or sensations creates a new layer of interaction with the real world, improving user experiences [6]. AR technology are widely available, with an increasing number of applications being developed, especially for mobile platforms like Wikitude AR Travel Guide [8] introduced in 2008.

There are many opportunities for integrating AR in an innovative way. The four types of applications that are most frequently utilized for AR research are entertainment and education, mobile applications, advertising and commercial, and

medical. As an example of AR in the educational field, [9] used their previously established GEIST project to create a mobile outdoor AR system application that helps users study history through a storytelling game in which they can show the ghosts from the past in AR form. In addition, the application "Pokemon GO!" is one of the most well-known outstanding accomplishments in the AR field [10]. The Pokemon Company released and created Pokemon Go in the summer of 2016 to millions of smartphone users. The game takes the player's location into account and by navigating the real-world surroundings, they can find and "catch" the "wild" Pokemon [10]. Thus, [9] claims that by displaying virtual data not only in their immediate surrounds, but also in any indirect view of the real-world environment, AR does not only improve how people see and interact with the real world, but it also makes their lives much easier. AR enhances users' impression of reality by seamlessly integrating digital content into their field of view and provides simple access to virtual information regardless of their perspective or position.

B. Virtual Museum

In 1947, Andre Malraux introduced the concept of the virtual museum, envisioning an imaginary museum without physical boundaries. This idea focuses on presenting content and information rather than specific locations. A virtual museum consists of digitally recorded scientific, cultural, or historical data, accessible through electronic means such as sounds, photos, text, or other formats. Early evidence of interactive virtual displays using augmented reality (AR) techniques can be found in a research paper [11]. For instance, the Meta-Museum guide system employed AR and artificial intelligence to connect a physical museum with the virtual realm, tapping into archives and knowledge bases. Another notable example is the "Louvre AR" application, a well-known AR-based virtual museum that allows users to interact with artwork by superimposing virtual data and interactive features onto real exhibits [12].

Vardhan et al. [13] had conducted a survey on the user experience of an augmented reality-based museum application they developed. The study revealed that the proposed application is significantly easier to navigate and has better results compared to a conventional museum. This is due to in order to aid in navigation and exploration, a complicated taxonomy system is frequently used in a conventional museum. Results show that over 71% of participants decided not to visit a conventional museum in the future and preferred to switch to the suggested application because it requires less travel time, less material, and less visible user load. Therefore, virtual museums may take full advantage of the new media's potential, analyze visitor demands, and respond to them in a variety of ways by using technology. Not only that, but they can also enable intuitive engagement with the displayed content as well as offering an enjoyable and instructive experience [13].

C. Plane Detection (ARCore)

In order to find the essential point clusters that can form horizontal and vertical planes, ARCore analyzes the scene. Additionally, it can pinpoint the bounds of each detected plane. In Figure 1, the plane that was recognized and detected in the image is represented by white dots and augmented with a 3D object.

Fig. 1 Plane Delimited with White Points and Augmented with 3D Object

Suonsivu A. [14] demonstrates how floor detection is supported by ARCore. Firstly, ARCore searches for feature point clusters on flat horizontal and vertical surfaces in order to understand the surroundings. After that, these surfaces are then saved as planes, which users can utilize within their applications. The surfaces should have some texture to aid in point detection during movement since ARCore's plane detection relies on feature points situated on these surfaces. For instance, a floor that is entirely white does not provide ARCore with enough details and lacks the necessary visual cues to gather sufficient scene information.

Syahputra et al. [15] had did an experiment of plane detection testing on a floor by using ARCore. During the plane detection test, the system guides the user to locate flat surfaces within their surroundings. After the planes are detected, the system proceeds to filter and identify the plane that represents the floor surface. This selection is based on certain criteria such as the orientation of the plane (horizontal or slightly tilted), its size, or the distance between the plane and the user. By applying these criteria, the system can accurately identify and isolate the floor plane from other detected planes in the scene. For instance, the camera needs to be positioned approximately 1 meter away from the flat surface, with an altitude distance of 1.5meters. This process involves detecting textures on the flat surface by utilizing the distance between the camera and the textures.

D. User Movement Detection

Today's smartphones include a variety of sensors to improve user experience, give programs better information about the environment, and deliver reliable [16]. In an AR application, a combination of sensors, such as the accelerometer sensor and gyroscope sensor, is frequently utilized to precisely identify and track user movement. Mobile phones' accelerometers are used to determine the phone's orientation. However, by monitoring rotation twist, the gyroscope - abbreviated gyro for short, adds a new dimension to the data provided by the accelerometer. A gyro monitors the angular rotational velocity, whereas an accelerometer measures the linear acceleration of movement. Although both sensors assess the rate of change, they do so for various things [16]. Table 1 shows the summary for accelerometer and gyroscope sensors measurement and functionality.

In summary, achieving reliable and precise detection and synchronization of user movement in both the real world and the virtual museum requires the combined use of accelerometer and gyroscope sensors. The accelerometer monitors changes in motion and linear acceleration, helping identify forward movement. As users approach exhibits, the gyroscope detects

angular rotation, synchronizing their movements in both the virtual and real environments. By integrating these two sensors, the user experience is enhanced, creating a more immersive and participatory environment.

TABLE 1 SUMMARY FOR ACCELEROMETER AND GYROSCOPE SENSORS

Sensor	Measurement	Functionality
Accelerometer	Measure linear acceleration (mV/g)	Detect changes in linear motion.
		Determine device's orientation and movement.
Gyroscope	Measure angular velocity (mV/deg/s)	Detect angular rotation
		Maintain device orientation, provided that data is on its movement in three-dimensional (3D) space.

E. Summary

This literature review explores the use of Augmented Reality (AR) in virtual museums, emphasizing its potential impact on cultural engagement and education. The research suggests that AR enhances how people perceive and interact with the real environment, simplifying daily activities. While AR technology is expected to become more widespread globally in the next few years, Malaysia lags in its adoption, hindering tourism and cultural preservation. The primary benefit of AR in museums is improved user engagement, leading to a deeper understanding of cultural exhibits. Current AR museum applications commonly include interactive elements like text displays, quizzes, and voice interactions. However, a drawback is the necessity for physical museum visits to access augmented features.

In this study, plane detection relies on ARCore, chosen for its compatibility with a broader range of Android devices, and Unity's support for ARCore using C# scripts. The decision is influenced by the familiarity with the C# programming language. User movement tracking employs a hybrid approach using accelerometer and gyroscope sensors through XR Origin and the AR session in Unity. This method ensures seamless synchronization between virtual and real-world movements, aiming for precise user movement detection. The ultimate objective is to allow users to experience and interact with augmented museum elements remotely, overcoming the constraint of physical museum visits.

III. METHODOLOGY

A. Software Development Life Cycle

In this section, the development method of the system will be discussed. The Iterative Enhancement Model, are used in the software development life cycle (SDLC). This method of software development is employed because it offers an adaptive and iterative approach, allowing for better management of changing needs, risk mitigation, incremental value delivery, and continuous improvement

This project entails building a virtual museum with three exhibits that encourage user interaction through augmented reality (AR) elements. The dataset of exhibits and the interaction features of the exhibits, which are the requirement, was implemented in each build. Each exhibit incorporates different augmented reality (AR) elements that enable users to interact with, as well as in order to launch the application and detect user movement, AR technology is utilized too. Each AR element is undergoing into different builds. The first build was testing the plane detection by implementing ARCore in order to launch the application, as well as test the user movement detection to synchronize real life and virtual movement by implementing accelerometer and gyroscope sensor. Next, the second build was testing the implementation of rotation and user viewing of the exhibits. Then, the third build was testing the implementation of zoom function of the exhibits, which is also one of the AR elements for user to interact with. Finally, the last build was testing the control panel of the proposed application including the evaluation of features such as the toggle button between AR and VR, exit functionality, and the main menu button. Each build was implemented and tested until it has successfully reached the expected result.

B. Functional Requirements

- The system should be able to detect all flat ground surface floors.
- The system should have the capability to detect user movement. For instance, when a user wanted to view the exhibit displayed on the left side in the virtual museum, their movement must be synchronized in both the virtual and real-world environments as they walk.
- Users should have the ability to engage with the exhibits in various unique ways. Each exhibit should offer distinct interaction options. For example, users can rotate the exhibit, watch videos showcasing its history, zoom in and out on the exhibit's image, and even press on the icon to access more details about the exhibit.

C. Iterative Process

Build 1 – Application Launch, Plane Detection and Synchronization of Real-Life and Virtual Movement

In the first build, the requirement is to design and implement floor detection, as well as implementing user movement detection. The XR Origin and AR Session package within the Unity will need to be installed in order to build for plane detection and user movement detection. Unity's ARCore for plane detection will be employed to detect the floor, facilitating entry into the virtual museum by identifying flat surfaces. By utilizing the device's camera and Inertial Measurement Unit (IMU), ARCore tracks the device's position and orientation in the real world. Consequently, the implementation and testing of ARCore in Unity will utilize the camera as the input sensor to detect flat ground surfaces in order to launch the application. The virtual museum floor map will be displayed, the virtual museum's door, as well as the exhibits will be displayed via the

smartphone. Next, the implementation of user movement detection by utilizing accelerometer and gyroscope sensor will also be combined and implemented within the Unity framework, enabling precise monitoring of the user's motion and achieve synchronization between the user's real-life movement and virtual museum navigation.

Build 2 – Implementation of Rotation and Information Viewing

In the second build, the requirement is to implement rotation functionality and design all exhibits within Unity. Specifically, the "Keris" exhibit will be created using 3Ds Max and imported as an FBX file into Unity. Then, the interactive elements such as the rotation feature for this "Keris" exhibit will be constructed by using Unity features. This build aims to test the smooth rotation capability of the "Keris" exhibit. Besides, the functionality of invoking the view function upon user interaction with the exhibits will also be tested. This includes testing the display of a poster which shows the historical information of the specific exhibit and the presentation of the explanation video the user clicks on the "Ngajat" Dance exhibit's image.

Build 3 – Implementation of Zoom Function

In the third build, the requirement is to implement the zoom function of the "Baju Kebaya" exhibit within Unity. This build is focused on testing the seamless functionality of zooming and utilizing pinch-to-zoom gestures to interact with the "Baju Kebaya" exhibit in conjunction with the zoom UI panel.

Build 4 – Application Control Panel

In the last build, the requirement is to design the pause menu function which will be located on the top right corner of the application's screen, and integrate it with the confirmation prompt, which the call function will be implemented by using the features provided by Unity. The functionality of handling the behavior of pause menu function will be tested to ensure that user will have the option to pause, resume, back to home page, exit or remain in the application as desired. Besides, the UI panel which allows users to toggle between AR and VR seamlessly will also be tested and conducted.

D. Process Flowchart

As shown in Figure 2, the process flowchart of the AR museum application begins by scanning the flat ground surface by using ARCore for plane detection. If a flat ground surface is successfully detected, the virtual museum display is triggered. After successfully entering the virtual museum, the museum door will be visible in the default virtual world environment. User movement detection is then enabled using both accelerometer and gyroscope sensors, allowing users to navigate the virtual museum by physically moving in the corresponding direction. As users explore the virtual museum, they can encounter different exhibits. For instance, when users tap on an exhibit, the application displays relevant exhibit information, providing users with a deeper understanding of the artifact or display. Additionally, users can interact with the exhibit by performing actions such as rotating it, or triggering multimedia elements associated with it. The application allows users to continue their exploration of the virtual museum or interact with other exhibits of interest. The application includes navigation features such as a pause menu function and the ability to switch between AR and VR modes while exploring the virtual museum, enhancing the overall user experience with a more realistic feel. Finally, the workflow concludes when the user decides to end their virtual museum experience by clicking on the exit button. If user choose "yes", then they will exit the application. However, if users choose "no", they will remain staying and continue to explore in the virtual museum.

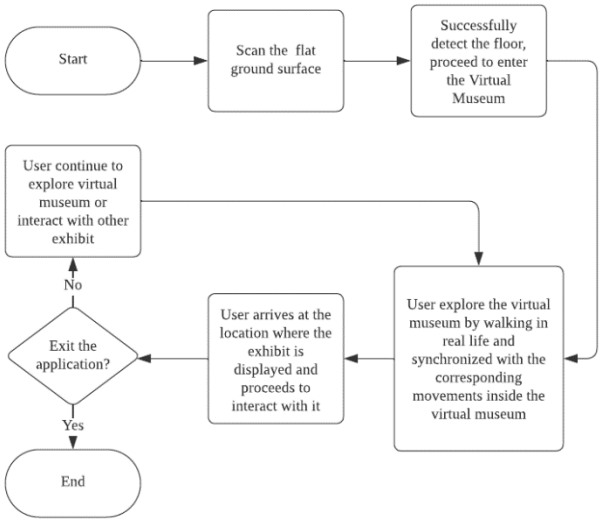


Fig. 2 Process Flowchart of AR Museum Application

IV. VIRTUAL MUSEUM APPLICATION

A. Virtual Museum Application Design

a) Activate camera to detect flat floor surfaces by implementing ARCore

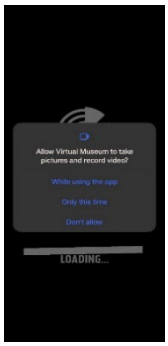


Fig. 3 Prompt Camera Activation

Fig. 4 Scanning in Progress

In Figure 4, there is a prompt for camera activation, informing users that they need to enable their camera while using this application. To enter the virtual museum, users must activate their camera and point their device towards a flat floor surface. Figure 5 presents a prompt notifying the user that the

system is currently scanning and detecting the flat floor surface to facilitate entry into the museum.

B. Exhibits (Interactive elements for AR exhibits)

a) "Keris"

Fig. 5 3D Model (Malay Traditional Sword: "Keris")

Fig. 6 Pop up "Keris" Poster

Fig. 7 Rotation of the "Keris" and auto zoom-in function

In Figure 6, the first exhibit is showcased, featuring the model exhibit of the "Keris"—a traditional Malay sword. In Figure 7, the "Keris" Poster is displayed in a popped-out state. To access detailed historical information about the "Keris", users can tap anywhere on the poster to enlarge it. Pressing the poster again will prompt it to return to its original position, as depicted in Figure 6. Furthermore, as depicted in Figure 8, users can rotate the "Keris" 360 degrees in any direction using their fingertips. During this interaction, the complete view of the "Keris", along with the "Keris" poster, will be zoomed in, providing a more visually appealing and detailed visualization.

b) "Baju Kebaya"

Fig. 8 Image (Malay Traditional Costume: "Baju Kebaya")

Fig. 9 Pop up "Kebaya" Poster

In Figure 9, the second exhibit is featured, showcasing the image exhibit of "Baju Kebaya", the traditional costume of the Malays. Zoom icons are positioned at the top left of the image exhibit, while the 'i' icon is placed on the top right, allowing users to access the historical explanation poster for the "Baju Kebaya" exhibit. Upon pressing the 'i' icon, the "Baju Kebaya" poster and image exhibit will be displayed simultaneously, arranged in a top-and-bottom configuration as shown in Figure 10. Tapping anywhere on the "Baju Kebaya" exhibit image again will initiate its return to the original position, as depicted in Figure 9.

a) "Ngajat" Dance

Fig. 10 Image (Malay Traditional Dance: Ngajat Dance)

Fig. 11 Video (Malay Traditional Dance: Ngajat Dance)

Fig. 12 Video Paused

Figure 11 shows the third and last exhibit in the virtual museum, which is an image of a traditional Malay dance, Ngajat Dance. Upon tapping anywhere on the image exhibit, the video will pop out and commence playing as shown in Figure 12. This interactive element is implemented using the video-on-click functionality. Figure 13 indicates that if users walk away from the exhibit or pause the video, the exhibit will no longer be in a popped-out state and will return to its original position. However, tapping on the exhibit again will resume the video playback from where it was paused.

C. Application Control Panel and Navigation

a) Overlaying with different modes: AR and VR

Fig. 14 AR/VR mode

Illustrated in Figure 14, the museum application features an AR/VR icon located at the top left. Users are provided with the option to overlay exhibits either with the real-life environment or the virtual world. They can seamlessly toggle between these two environments while engaging with exhibits in the virtual museum. Figure 15 shows the exhibit overlaying with AR environment and Figure 16 shows the exhibit overlaying with VR environment.

Fig. 13 Exhibit overlaying with AR

Fig. 14 Exhibit overlaying with VR

V. Results and Discussion

Table 2 shows the test scenario, test cases, test data and expected outcome of the proposed virtual museum functionality. All the proposed functionality is tested, observed the outcome of the results as expected result. It has successfully passed the functional system testing. The designed interface of the virtual museum on user-friendly, ensuring ease of operation and navigation is pending for user acceptance test, and it is not presented here. This proposed solution is prepared for release and use as the application upon completion of the user acceptance test.

VI. Conclusion

The proposed system is developed and designed to enhance visitor experience, increase visitor engagement, and achieve learning outcomes by implementing AR technology and merging the physical and digital world. Through the creation of an immersive and interactive display, the system enhances users' understanding of exhibits. The interactive features, crafted with AR technology, bring exhibits to life through digital elements, offering users an immersive experience to explore exhibits from various angles and interact with them within the application.

Moreover, this application guarantees a user-friendly interface, ensuring ease of interaction with exhibits. It facilitates straightforward navigation, delivers clear instructions, and provides a seamless transition between the physical and digital elements of the exhibit.

References

[1] J. Lee, H. Lee, D. Jeong, J. Lee, T. Kim, and J. Lee, "Developing Museum Education Content: AR Blended Learning," International Journal of Art & Design Education, vol. 40, no. 3, pp. 473–491, Mar. 2021, doi: https://doi.org/10.1111/jade.12352.

[2] C. Matuk, "The Learning Affordances of Augmented Reality for Museum Exhibits on Human Health," Museums & Social Issues, vol. 11, no. 1, pp. 73–87, Jan. 2016, doi: https://doi.org/10.1080/15596893.2016.1142815.

[3] R. Wojciechowski, K. Walczak, M. White, and W. Cellary, "Building Virtual and Augmented Reality museum exhibitions," Proceedings of the ninth international conference on 3D Web technology - Web3D '04, 2004, doi: https://doi.org/10.1145/985040.985060.

[4] A. M. Cárdenas Gasca, J. M. Jacobs, A. Monroy-Hernández, and M. Nebeling, "AR Exhibitions for Sensitive Narratives: Designing an Immersive Exhibition for the Museum of Memory in Colombia," Designing Interactive Systems Conference, Jun. 2022, doi: https://doi.org/10.1145/3532106.3533549.

[5] D. Marques and R. Costello, "Concerns and Challenges Developing Mobile Augmented Reality Experiences for Museum Exhibitions," Curator: The Museum Journal, vol. 61, no. 4, pp. 541–558, Oct. 2018, doi: https://doi.org/10.1111/cura.12279.

[6] C. Ng and C. Ramasamy, "Augmented Reality Marketing in Malaysia—Future Scenarios," Social Sciences, vol. 7, no. 11, p. 224, Nov. 2018, doi: https://doi.org/10.3390/socsci7110224.

[7] W. Lin, W.-T. Lo, and H.-P. Yueh, "How the Multimodal Media in Augmented Reality Affects Museum Learning Experience," 2019 12th Asia Pacific Workshop on Mixed and Augmented Reality (APMAR), Mar. 2019, doi: https://doi.org/10.1109/apmar.2019.8709286.

[8] "Mobile AR Apps Can Now Track Any Surface Using Plane Detection via Wikitude SDK," *Next Reality*, Sep. 13, 2018. https://next.reality.news/news/mobile-ar-apps-can-now-track-any-surface-using-plane-detection-via-wikitude-sdk-0187402/ (accessed Nov. 15, 2023).

[9] B. Furht, Ed., Handbook of Augmented Reality. New York, NY: Springer New York, 2011. doi: https://doi.org/10.1007/978-1-4614-0064-6.

[10] Á. Zsila et al., "An empirical study on the motivations underlying augmented reality games: The case of Pokémon Go during and after Pokémon fever," Personality and Individual Differences, vol. 133, pp. 56–66, Oct. 2018, doi: https://doi.org/10.1016/j.paid.2017.06.024.

[11] S. Styliani, L. Fotis, K. Kostas, and P. Petros, "Virtual museums, a survey and some issues for consideration," Journal of Cultural Heritage, vol. 10, no. 4, pp. 520–528, Oct. 2009, doi: https://doi.org/10.1016/j.culher.2009.03.003.

[12] D. / S. Hirschmiller, "Louvre Augmented Reality Exhibit With Snapchat Uncovers Ancient Egypt," *Decrypt*, Oct. 18, 2023. https://decrypt.co/202106/louvre-augmented-reality-exhibit-snapchat-uncovers-ancient-egypt

[13] Harsh Vardhan, A. Saxena, A. Dixit, S. Chaudhary, and A. Sagar, "AR Museum: A Virtual Museum using Marker less Augmented Reality System for Mobile Devices," Nov. 2022, doi: https://doi.org/10.1109/icict55121.2022.10064611.

[14] A. Suonsivu and U. Arcore, "RGBD SLAM BASED 3D OBJECT RECONSTRUCTION AND TRACKING," 2020. Accessed: Jul. 05, 2023. [Online]. Available: https://trepo.tuni.fi/bitstream/handle/10024/122760/SuonsivuAleksi.pdf?sequence=2

[15] M. F. Syahputra, F. Hardywantara, and U. Andayani, "Augmented Reality Virtual House Model Using ARCore Technology Based on Android," Journal of Physics: Conference Series, vol. 1566, p. 012018, Jun. 2020, doi: https://doi.org/10.1088/1742-6596/1566/1/012018.

[16] "Sensors - definition - GSMArena.com," www.gsmarena.com. https://www.gsmarena.com/glossary.php3?term=sensors#:~:text=Accelerometers%20in%20mobile%20phones%20are

TABLE 2 FUNCTIONALITY TEST SCENARIO

Test Scenario	Test Case	Pre- Condition	Test Steps	Test Data	Expected Result
Application Launch and Plane Detection	Validating successful application launch and plane detection for entering the virtual museum	Application installed and launched on the device	1) Activate camera. 2) Point device at flat ground surface. 3) Initiate plane detection. 4) Wait for successful floor detection. 5) Display virtual door on the device.	Test device with functioning sensors for plane detection at different environments.	Virtual door shown on the user's screen with instructions for entering the virtual museum.
Movement Synchronization	Validating synchronization of user movement in real life with virtual museum navigation	User successfully entered the virtual museum	1) Ensure accelerometer and gyroscope sensors are working. 2) Move towards exhibits inside the virtual museum. 3) Verify synchronization between real and virtual movement. 4) Repeat for different exhibit locations.	Test device with functioning sensors for user movement based on virtual map design.	Application accurately tracks user's movement in the real world, synchronized with virtual museum navigation.
User Interaction - Exhibit Rotation	Validating smooth rotation of exhibit by user interaction	Application accurately tracks user's movement in the real world	1) Navigate to Keris exhibit. 2) Use UI control for exhibit rotation. 3) Interact with exhibit in different directions.	Test by using fingertips for 360-degree rotation.	Users can interact with the exhibit, smoothly rotating it in a 360-degree manner.
Display of Keris Poster	Validating display of Keris poster popping out upon pressing	Virtual museum displayed on the device	1) Navigate to Keris exhibit. 2) Press on Keris Poster to trigger display. 3) Verify poster pops up and returns to original position.	Test by using fingertips to tap on the Keris Poster.	Keris Poster pops up after pressing on it, returns to original position after pressing again.
Display of Kebaya Poster	Validating display of Kebaya poster upon pressing the 'i' icon	Virtual museum displayed on the device	1) Navigate to Kebaya exhibit. 2) Press on 'i' icon to trigger display. 3) Verify Kebaya Poster displays with layout arranged in top-and-bottom configuration. 4) Verify exhibit image returns to original position.	Test by using fingertips to press on the icon.	Kebaya Poster pops up after pressing 'i' icon, returns to original position after pressing exhibit picture.
Display of Exhibit Video	Validating display of video upon pressing exhibit image	Virtual museum displayed on the device	1) Navigate to Ngajat Dance exhibit. 2) Press on exhibit image to trigger video display. 3) Verify video overlay pops up, showcasing exhibit's history. 4) Observe accurate playback of historical information. 5) Verify video resumes from last paused position upon returning to exhibit.	Test by using fingertips to press on the exhibit image.	Users can interact with exhibit, triggering video display, with accurate historical information. Video resumes from last paused position upon return.
Toggle Between AR and VR	Validating overlay of exhibits with real-life environment in AR mode and virtual world in VR mode	User has entered the virtual museum	1) User walks through virtual door. 2) Tap AR panel for real-life overlay. 3) Verify exhibits overlay with real-life environment. 4) Tap VR panel for virtual world overlay. 5) Verify exhibits overlay with virtual world.	Test by using fingertips to press on AR and VR UI panel.	Users can switch between AR and VR modes, overlaying exhibits with the chosen environment.

Analysis of Artistic Concept and Expression in the Film 'Man Jiang Hong'

1st Chuxuan Wang
Dept. of Design
Dongseo University
Busan, Korea
zhishi0210@gmail.com

2nd HaeYoon Kim
Dept. of Design
Dongseo University
Busan, Korea
haeyoonkim@empal.com

3rd Yong-Ki Lee
Dept. of Design
Dongseo University
Busan, Korea
ykl177@gdsu.dongseo.ac.kr

***Abstract*—As a director who has successfully learned from and transformed Chinese traditional culture, Zhang Yimou's films are permeated with rich cultural reflection and humanistic care, forming a unique Zhang's film style. His film *Man Jiang Hong* will be released in the 2023 Spring Festival. The patriotism expressed in movies is not only the inheritance of traditional Chinese culture, but also the promotion of Chinese spirit, which to some extent satisfies foreign audiences' imagination of Chinese culture and history. The article is based on the contemporary adaptation of the story of Yue Fei in movies, analyzing Zhang Yimou's film's new interpretation of his patriotism from different perspectives. This article explores the artistic characteristics, creative gains and losses of the film, including its theme, narrative, value communication and performance, audio-visual techniques, and presentation of individual styles.**

Keywords—Man Jiang Hong, Zhang Yimou, Movies, Family and country feelings, Cultural communication

I. INTRODUCTION

The movie *Man Jiang Hong* has grossed 4.544 billion RMB (approximately 623 million USD) at the Chinese box office and was released in South Korea on October 11th. Since the Chinese New Year, this suspenseful comedy film, directed by Zhang Yimou and starring Shen Teng and Jackson Yee, has maintained its popularity.

According to the Chinese Film Archive's survey of audience satisfaction, the overall audience satisfaction score for films during the 2023 Chinese New Year box office period was 87.1 points, showing an increase of 1.8 points compared to the previous year. Among them, *Man Jiang Hong* topped the list with a score of 88.4 points, ranking 6th among all 397 films surveyed in historical data. This indicates that the film has received widespread recognition from the audience. The movie appropriately integrates comedy elements into its fast-paced, strongly-twisted suspenseful storyline, effectively conveying mainstream values of sacrifice for justice and patriotism in a relatively light-hearted manner. It strikes a good balance between commercial appeal and artistic expression, making it Zhang Yimou's highest-grossing film in the mainland Chinese market.

This article will analyze *Man Jiang Hong* from three aspects: artistic style, narrative style, and thematic ideas. It primarily discusses why the film achieved such high box office numbers, garnered praise from so many audiences, and its innovativeness and entertainment value.

II. LITERATURE REVIEW

Murder mystery games are a genre of party games One of the players secretly plays a murderer, whilst other players must determine who among them is the criminal. Murder mystery games often involve the actual 'murders' of guests throughout the game, or open with a 'death' and have the rest of the time devoted to investigation. The script-murder movie refers to a film created based on the murder mystery game. The movie incorporates elements of this game into its plot, presenting an interactive and deduction-driven cinematic work.

As of now, murder mystery movies are considered a relatively emerging genre, hence the literature on this topic may be relatively limited. Typically, these works are produced in a movie format, incorporating elements such as plot, characters, enigmas, etc., based on the settings and rules of the script-murder game, to enrich the storyline and emotional aspects of the film. The novelty of this research lies in the analysis of the movie *Man Jiang Hong* as an example of a script-murder movie, filling the gap in academic papers of this kind within the field of film studies.

German scholar Jan Assmann proposed in his book "Cultural Memory: Writing, Memory, and Political Identity in Early High Cultures" that, "For cultural memory, what matters is not a verifiable history, but a history that is remembered." This article will analyze how "Man Jiang Hong" integrates patriotism and murder mystery movie elements based on this theory, creating a film that aligns with popular aesthetics and mainstream value concepts.

III. ARTISTIC CONCEPT AND EXPRESSION

A. Vivid artistic style

Chen Yu, the screenwriter of the movie *Man Jiang Hong*, and a professor at Peking University, introduced that the core creative team led by director Zhang Yimou initially aimed to create a film highly related to space. To achieve this, they specially constructed a complete mansion in Shanxi.

The film pursues a highly compact pace and intense plot, striving to meet the "Three Unities" narrative principle of the theatrical stage. At the same time, we hope to create a

This research was supported by BK21 Four Service Design driven Social Innovation Educational Research Team in Dongseo University.

mainstream film that truly captivates the audience and expresses the mainstream values of contemporary China. This includes profound nationalistic sentiments as well as the patriotic feelings inherent in every Chinese.

Throughout the filming of *Man Jiang Hong*, the core creative team always had the audience in mind, contemplating the plot and character designs from the audience's perspective. They sought to seamlessly integrate suspense and comedic elements in an innovative manner.

The movie *Man Jiang Hong* features concentrated scenes, with the duration of the film closely aligning with the events within the movie. It tells the story of a group of people "taking on roles" and physically acting out, gathering in a large courtyard to deduce the crime scene, overcoming multiple "levels," and seeking the "truth" through their "bodies" and lives. The film is full of suspense, a tightly woven plot, a fast-paced rhythm, and intricate connections, imbued with a spirit of gaming and game-like thinking. From a certain perspective, its intense plot twists offer a gaming appeal, a brain-burning reasoning challenge, excitement, and a sense of satisfaction.

The tremendous box office success of the film may suggest the potential for movies with the characteristic thinking of "script-murder," indicating a flourishing future in both artistic and industrial aspects. According to Li Daoxin, the Vice Dean of the School of Arts at Peking University, *Man Jiang Hong* is a highly individualized work, characterized by its enclosed time and space and ultimate narrative, showcasing distinctive artistic features. The film pursues an almost single-take filming technique, where the director's stylized visual language aligns with the constantly shifting plot, creating a kind of aesthetic resonance. Judging by the film's box office performance, this resonance has been widely acknowledged by the audience. The film skillfully blends tragedy and comedy, two vastly different emotional expressions, and also makes attempts and efforts to bridge the genre boundaries between historical and costume films.

Man Jiang Hong is a film that respects market principles, artistic principles, and the director's personal artistic expression. The film integrates elements of comedy, suspense, and action, selecting market-appealing actors suitable for shaping corresponding roles, forming the commercial foundation that captivates a broader audience. Additionally, the film embodies many of director Zhang Yimou's personal artistic styles. Viewers can discern numerous aesthetic elements and creative techniques from Zhang Yimou's classic films, serving as a robust guarantee for the artistic quality of the film.

B. Innovative narrative approach

The success of *Man Jiang Hong* is primarily attributed to its successful narrative. The innovative narrative style, akin to a "script-murder" approach, brings the film closer to a younger audience. *Man Jiang Hong* continuously utilizes the art of plot twists, intertwining elements of comedy, period drama, suspense, escape, revenge, tragedy, and more within the film through these twists. This captivates and engages the mainstream cinema audience, offering unexpected plot turns and dramatic character choices, creating a sense of surprise and excitement. The constant plot reversals in the film provide the audience with a sense of novelty and thrill, prompting them to reevaluate the story and characters. In this process, the audience becomes more deeply involved in the film's reinterpretation.

C. Showcasing patriotism and nationalistic sentiments

Man Jiang Hong not only leaves the audience with a sense of satisfaction regarding the story but also touches them emotionally. Throughout the filming of *Man Jiang Hong*, the core creative team always had the audience in mind, contemplating the plot and character designs from the audience's perspective. In the film, a group of ordinary individuals courageously engage in a life-and-death battle against treacherous officials, ultimately revealing the mystery with the spirited and impassioned song *Man Jiang Hong*. This elevates patriotic sentiments to a climax, displaying the unwavering loyalty and courage that burns within their hearts.

Man Jiang Hong reconstructs history and reinterprets the most fundamental elements of national culture, such as "patriotism" and "sacrifice for justice," from a contemporary perspective, making it a creative transformation and innovative development of excellent traditional Chinese culture.

The film boldly breaks through and innovates traditional narrative styles, where the first half predominantly features comedic elements and the latter half suspenseful elements, ultimately fulfilling the narrative purpose of promoting patriotic sentiments.

Man Jiang Hong has a lofty concept and is a mainstream film that convinces its audience. The film's concept is clever and unique, compactly condensing the main plot and characters within a confined space, showcasing the creative abilities of the director, screenwriter, and lead actors. The unexpected plot twists in the movie are not there just for the sake of surprising the audience; rather, they vividly exhibit strong patriotic sentiments within the ups and downs of the narrative.

IV. CONCLUSION

American scholar Hayden White proposed in his essay *The Burden of History*[1], "History cannot be reconstructed. What we see as history is actually the result of historians weaving a narrative pattern. Behind history hides a deep-seated poetic narrative." Historical texts are an artistic fabrication based on "imagination" and "literariness"; it's a "poeticized history." When creating film and television works based on historical themes, we need to consider the integrity of the work and the audience's appreciation needs. Through reasonable fictionalization and artistic imagination, we construct a complete, closed narrative chain, achieving "self-sufficient truth."

The film "Man Jiang Hong" doesn't merely overlay excellent traditional Chinese cultural elements; rather, it represents an artistic recreation where creators, based on contemporary aesthetics, value judgments, and creative stances, strive to achieve a balance between historical and artistic authenticity. By integrating artistic aesthetics with thematic aesthetics and building upon a reasonably fictionalized foundation, the film establishes a progressive narrative logic that mirrors the fate of the nation and its people. While portraying the essence of patriotism, the film also

[1] Wunsch,James: The burden of history.[J].Phi Delta Kappan,1994.

reinterprets traditional Chinese culture, enhancing the film's appeal and presenting a rich, multidimensional human experience to the audience, better meeting their aesthetic expectations.

By integrating artistic aesthetics with thematic aesthetics and building upon a reasonably fictionalized foundation, the film establishes a progressive narrative logic that mirrors the fate of the nation and its people. While portraying the essence of patriotism, the film also reinterprets traditional Chinese culture, enhancing the film's appeal and presenting a rich, multidimensional human experience to the audience, better meeting their aesthetic expectations.

REFERENCES

[1] Larry Samovar: Cross cultural Communication [M]. Beijing: Renmin University of China Press, 2010:11-24.

[2] Wu Shuyan: Analysis of the International Cooperation Model of Zhang Yimou's Films: Taking "The Great Wall" as an Example [J]. Cultural and Educational Materials, 2020 (32): 28-30.

[3] Zhang Jing, Liu Yaqiong: Multiple ideographic expressions in historical documentaries from the perspective of cross-cultural communication [J]. Film Literature, 2022 (5): 64-66.

[4] Hao Chaoshuai: The Rise and Fall of "Presence Culture" and the Change of "Cultural Identity" in 21st Century Chinese Films [J]. Film Review, 2021 (11): 7-11.

[5] Sui Zhiqiang, Li Yeping: Activating the Current Inheritance of Excellent Traditional Culture through Innovative Expression of Historical Images [J] Dongyue Luncong, 2021, 42 (03): 30-37.

[6] Jean Mitri; The Art of Drama in Modern Cinema [J]. World Film, 1985.

[7] Zhou Aibing: TV documentary "If National Treasures Can Speak": Symbolic Representation, Cultural Resonance, and Identity Construction [J]. Film Review, 2022 (11): 89-92.

[8] Ma Dai: Imagination and Identity: A Practical Approach to the Spread of the Consciousness of the Chinese Ethnic Community in Ethnic Theme TV Dramas in the New Era [J]. Chinese Editor, 2022 (02) H2-H5,91.

Portable Monitoring System for Fall Detection using YOLOv7

JaeYeon Byun
Dept. of Artificial Intelligence
Daegu University
Gyeongsan, Republic of Korea
9722jayon@naver.com

YunSeok Yang
Dept. of Artificial Intelligence
Daegu University
Gyeongsan, Republic of Korea
diddbstjr55@gmail.com

BoKyeong Kim
Dept. of Artificial Intelligence
Daegu University
Gyeongsan, Republic of Korea
qhrud418@gmail.com

Kyung-Ae Cha
Dept. of Artificial Intelligence
Daegu University
Gyeongsan, Republic of Korea
chaka@daegu.ac.kr

Se-Hyun Park
Dept. of Artificial Intelligence
Daegu University
Gyeongsan, Republic of Korea
sehyun@daegu.ac.kr

***Abstract*— This study introduces a fall detection system that uses the built-in cameras of mobile devices. Without the need for additional equipment or installations, users can access the proposed system through a mobile app. The system connects to a server via a network, storing data upon detecting a fall, issuing alerts to users, and monitoring incidents. This system is designed to be versatile, allowing it to be used in any location where individual users require such a system. This study commences with background research on intelligent monitoring systems and the related industrial trends. The overall system architecture is then discussed, including training data and models. The effectiveness of the system is verified through an implementation process and the introduction of performance metrics based on experimental applications of this system.**

Keywords*— *Mobile app-based system, Real-time monitoring, YOLOv7

I. INTRODUCTION

Based on the statistics related to the causes of accidents according to age that are recently reported to the Consumer Accident Surveillance System, accidents caused by falls rank second across all age groups [1]. These accidents occur continuously without being restricted by location and time.

In all safety-related accidents, minimizing damage by observing the golden time is ideal; however, in reality, this opportunity is often missed. To prevent unforeseen safety-related accidents, prompt actions are required. Therefore, intelligent monitoring systems are being increasingly utilized in various scenarios and contributing to the growth of related industries [2]. For example, IntelliVIX has surpassed traditional closed-circuit television (CCTV) technology to become a specialized intelligent CCTV company with the highest domestic sales revenue in this field [2]. However, the cost of constructing an intelligent CCTV system, including labor and material costs, is in the range of 3–4 million won per unit when ordered in bulk. As multiple CCTVs need to be installed, the overall cost significantly increases [3]. Therefore, individual users must bear substantial expenses when applying such systems in their homes where most safety-related accidents occur [1]. To increase the accessibility of intelligent CCTV systems for individual users within households, a user-friendly interface that allows for easy access is imperative.

Based on this requirement, this study developed a fall detection system that can easily be established without the need for additional equipment or installation processes by utilizing the cameras embedded in mobile devices. The system is designed to store data in the event of a fall via network connectivity and provides users with notifications and real-time monitoring capabilities. Such a system has been designed to be universally applicable in households and external spaces, wherever individual users may require it.

The structure of this paper is as follows. Section II investigates the background related to technologies and related industrial trends, and Section III designs the overall system architecture and constructs the training data and models. Section IV covers the actual implementation process of the system, and Section V introduces the experimental results and performance metrics to validate the effectiveness of the system. Finally, Section VI concludes the paper.

Corresponding author: Kyung-Ae Cha, chaka@daegu.ac.kr .

II. RELATED TECHNOLOGICAL TRENDS

A. Research Background Related to Industrial Trends

The technology behind intelligent image security is a next-generation technology that automatically analyzes images collected in real-time through a CCTV using both software and hardware functions designed to detect and respond to anomalies. The main keywords in this industry include "4K CCTV," "artificial intelligence cameras," and "smart home security services." Intelligent CCTV technology based on deep learning has evolved from motion direction detection to the development of cloud sourcing [4]. Domestically, the primary trend of image security revolves around intelligent video analysis. Although CCTV initially gained prominence in securing public facilities and services, its applications have expanded to private sectors such as corporations and households. Furthermore, the demand for smart home services and CCTV systems is projected to continue expanding owing to the increasing proportion of single-person households and the surge in various types of crimes.

B. Object Detection via Artificial Intelligence

The You Only Look Once (YOLO) algorithm [5] is a deep learning-based approach for real-time object detection. YOLO divides the input image into a grid and simultaneously estimates bounding boxes and object probabilities for each grid cell. This processing method significantly improves computational efficiency by processing the entire image only once. Human recognition and fall detection require real-time decision-making and responses. Therefore, a learning model, i.e., YOLO, capable of accurate object recognition and fast

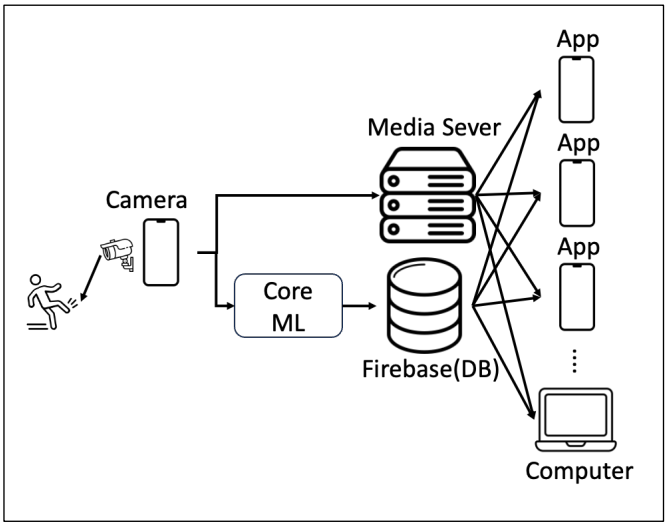


Fig. 1. System configuration

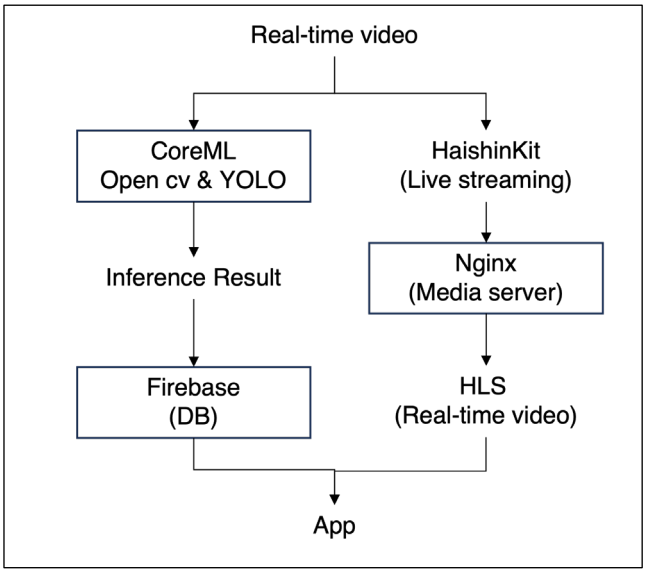


Fig. 2. Process of receiving videos and notifications

processing was utilized to implement the system proposed in this study.

III. Proposed System

A. System Structure

This study proposes a system for providing real-time event-detection alerts and video streaming, as shown in Fig. 1. The system integrates video data captured via smartphones with a trained YOLO model through the CoreML framework on iOS devices for inference. When specific events occur, the system sends users real-time notifications containing information such as location, time, detection circumstances, and detection images, enabling immediate responses.

Additionally, this system is designed with a structure that transmits video data to multiple mobile devices and computers through a media server. This architecture allows users to view the streaming video in real time across various devices.

The system proposed in this study can be flexibly applied in diverse environments, depending on the training data. This system can potentially be applied in various scenarios and settings, including home environments, elderly care and childcare, and industrial sites.

B. Dataset and Model Training

In this study, a learning model was constructed using the Roboflow image data labeling and sharing platform (Roboflow, Inc.) [6]. A total of 2,790 image data were processed on this platform and categorized into two classes: "person" and "falldown." Based on these two classes, two functionalities were implemented: "fall detection" and "person count detection."

Furthermore, 72%, 16%, and 11% of the dataset was divided into the training, validation, and test sets, respectively. To increase training speed and minimize the computing resource burden, all images were adjusted to a consistent resolution of 640×640. Data augmentation techniques, including image transformations, rotations, translations, flips, and resizing, were employed to increase the diversity of the training data. These transformations resulted in the construction of a total of 6,830 training images.

The model used in this study was trained based on YOLOv7 and subsequently converted to CoreML [7], a model format compatible with iOS, for usage. Coremltools was employed to convert the PyTorch model to CoreML such that the model can be used in the mlmodel format. For this study, Coremltools version 6.2.0 or below was used for the conversion.

The YOLOv7 model was used for training in this study. The training environment utilized the NVIDIA A100 graphics processing unit (GPU) provided by Colab. The hyperparameters included a total of 100 epochs, a batch size of 16, and an input image size of 640 × 640 pixels.

IV. Implementation

A. Design of the Proposed Monitoring System

Videos recorded using iOS devices offer two primary functions. The left part of Fig. 2 represents the intelligent CCTV mode that detects the situation presented in every frame from the input stream, i.e., whether a person falls down or too many people are within a particular area. When this mode is activated, in the event of a fall or if the number of individuals exceeds a predefined threshold, detection data is stored in Firebase, and notifications are sent to users.

The right side of Fig. 2 shows the second function, which provides the multi-platform streaming server to distribute real-time input videos from various types of devices. Using this streaming service, the app user can access on-time videos from the monitoring system.

This feature uses HaishinKit [8] to transmit videos at 30 frames/s in the real-time messaging protocol(RTMP) format to a Nginx [9] media server. On the receiving end, this video is processed in the hue, saturation, and lightness format with a .m3u8 file extension. When receiving real-time videos on iOS devices, the audio data must also be provided. Otherwise, the AV Player may not recognize this video as "live." As the video is received via Nginx, external access is possible for multiple users.

B. Detection Processing

In this study, individual frames of the input video were sliced using OpenCV, and each frame was inferred through CoreML. During this process, the obtained inference results contained detected class labels, and these values were then

Fig. 3. Training result of the YOLOv7 model

Fig. 4. Example of a push message

Fig. 5. Screenshot of the detection results from a real-time video

stored in a list. The primary points of interest in this study are the labels "person" and "falldown," which represent people and falls, respectively.

By aggregating the number of "person" and "falldown" labels stored in the list, the number of detected individuals and the occurrence of falls at specific points in time could be quantitatively assessed. Therefore, this system performed real-time object detection and quantitative analysis of outcomes, particularly related to person recognition and fall detection.

C. *Prevention of Erroneous Detection*

The preceding functions are aware that various variables can be generated during certain situations. For instance, a person may briefly exit the camera's view and re-enter or an object like a person may be erroneously detected as a person. To address such potential issues, counters were introduced for each function.

These counters include a fps counter that increments with each frame and function-specific counters that increment based on specific conditions, such as "when a fall is detected" or "when a certain number of people are detected." The fps counter resets after every 200 operations, whereas the function-specific counters become active after every 80 operations.

When the value of the function-specific counter exceeds a specified threshold, the corresponding information including time and location, is sent to Firebase. Subsequently, a warning notification is dispatched to the iOS device on which the application is installed. This enables flexible adaptation to dynamic elements in real-time environments.

V. Experimental Results

A. *Trained Model Performance*

This section shows the performance of the model trained in this study and the experimental results.

The precision obtained was 0.7573, the recall was 0.7495, and the mean average precision (mAP)_0.5 score was 0.7912. The higher the mAP value, the better the performance.

As depicted in Fig. 3, the application developed in this study using the YOLO v7 model demonstrated excellent performance, even under challenging low-light conditions, by activating features such as fall detection and people counting. These results serve as evidence of the system's performance and practical applicability, as proposed in this research.

B. *Application User Interface*

The messaging service provided by Firebase can be used to send notifications with desired content to users. However, a drawback is that notifications can only be received when the application is actively running and not when the application is closed. To address this limitation, the use of the Apple push notifications (APNs) platform is necessary. Owing to Apple's security policies, if APNs are not used, notifications cannot be sent after the application has been closed. Therefore, when the application requests Firebase messaging functionality through APNs, users can receive notifications even if our proposed application is inactive. Fig. 4 presents an example of the notification text as a push message.

The implementation of the proposed application is visible in the displays on the user interface. Among its features, the zone verification function assists users in the real-time monitoring of specific areas. Additionally, the fall detection feature identifies abnormal movements, whereas the people counting function automatically calculates the number of individuals in the video, as represented in Fig. 5. Finally, the notification feature delivers this information to users in real time.

Furthermore, the application developed herein securely manages individual user information by requiring users to undergo the Google login process when starting the application. The application also provides additional options and settings, allowing users to customize the system settings according to their preferences. Future updates aim to enhance the application further by incorporating various features and improving user convenience, ultimately achieving a more comprehensive application.

In addition, a low-resource mode was developed, by considering the performance differences between iOS devices. This mode efficiently separates real-time video streaming and

model inference, thereby conserving computational resources. When the user is not actively monitoring the CCTV, only model inference is conducted. Real-time video streaming is initiated when the user accesses the detailed page.

VI. CONCLUSIONS

In this study, a portable monitoring system based on YOLOv7 real-time object detection is proposed. The objective of this system is to rapidly detect risks, provide warning notifications to users, and enable timely responses for accident detection. Future research will focus on performance improvements through various experiments and data enhancements, to advance the deep learning-based real-time monitoring system and improve the safety in various user environments.

REFERENCES

[1] Y.S. Choi, Y.J. Kim, “2019 Consumer Protection Trends Analysis Report,” Korea Consumer Agency, 2020.

[2] G.U Kim "Special Report - Trend of Technology in Video Surveillance System," Telecommunications Technology Association, 2017.

[3] J. Park, K. Kim, “Comprehensive plan for national safety system reform,” MOIS. Ministry of the Interior and Safety, February, 2023

[4] T.S. Jo, H.K. Kim, “Technical Committee Trend Report,” Korea Conformity Laboratories,

[5] Wang, Chien-Yao & Bochkovskiy, Alexey & Liao, Hong-yuan. (2022). YOLOv7: Trainable bag-of-freebies sets new state-of-the- art for real-time object detectors. 10.48550/arXiv.2207.02696.R. Nicole, “Title of paper with only first word capitalized,” J. Name Stand. Abbrev., in press.

[6] Roboflow, [Online]. https://roboflow.com/. [access: September 18, 2023]

[7] CoreML, [Online]. https://developer.apple.com/kr/machine-learning/core-ml/. [access: September 18, 2023]

[8] HaishinKit, [Online], https://github.com/shogo4405/HaishinKit.swift/. [access: September 18, 2023]

[9] Ngnix, [Online]. https://www.nginx.com/. [access: September 28, 2023]

Gas Monitoring and Disinfection System in an Agricultural Farm Prototype

Md. Jahirul Islam
Department of Semiconductor Electronic Engineering
Daegu Catholic University
Gyeongsan, Korea
jahirruet@cu.ac.kr

Hyeji Lee
Department of Semiconductor Electronic Engineering
Daegu Catholic University
Gyeongsan, Korea
hjlee20@cu.ac.kr

Kihak Lee
Department of Semiconductor Electronic Engineering
Daegu Catholic University
Gyeongsan, Korea
khlee0905@cu.ac.kr

Hyun Kil Cho
R&D Center
SEMS Co. Ltd.
Gyeongsan, Korea
chohyunkil@naver.com

Wolyoung Kim
R&D Center
SOLARLIGHT KOREA
Daegu, Korea
z9203079@gmail.com

Bonghwan Kim*
Department of Semiconductor Electronic Engineering
Daegu Catholic University
Gyeongsan, Korea
*bhkim@cu.ac.kr

***Abstract*— The design, development, and performance assessment of a cutting-edge agricultural farm prototype outfitted with gas monitoring sensors, an exhaust fan, and a disinfection system are presented in this work. The prototype promises to improve crop productivity and animal health by monitoring and managing gas concentrations in the farming environment. The study entailed controlling the speed of the exhaust fan and the cleaning procedure while measuring gas sensor data under various circumstances, such as when cow dung and acetone gas were present. The outcomes show how well the technology works to lower gas levels within the chamber.**

Keywords— agriculture, farming, gas monitoring, disinfection prototype, gas sensors.

I. INTRODUCTION

Numerous issues confront modern agriculture, such as the need for greater effectiveness, environmentally friendly methods, and enhanced animal welfare [1]. To solve these problems, creative approaches that improve the agricultural environment are required. Systems for gas monitoring and disinfection are crucial to attaining these goals [2-3]. Gases including methane, ammonia, and volatile organic compounds (VOCs) may build up in many agricultural situations and have a negative impact on crop development, livestock health, and worker safety [4-5]. For improving agricultural production and sustainability, monitoring and management of these gases are essential. The design, development, and evaluation of a prototype agricultural farm outfitted with cutting-edge gas monitoring and disinfection technologies is the main goal of this project. In order to accomplish this goal, we want to: Create a controlled environment for agricultural operations that reduces the presence of dangerous gases and enhances crop output and quality by adjusting gas levels within the agricultural chamber, reduces exposure to hazardous gases, improving animal wellbeing examining the effectiveness of a farming environment-safe disinfection system based on electrolysis. There are several advantages to efficient gas monitoring and disinfection in agriculture. First of all, it enables the development of a more hospitable environment for animals and crops, leading to increased yields and better animal health [6]. Second, by lowering greenhouse gas emissions and controlling odor problems, it supports ecologically friendly farming operations. And last, it improves overall security for agricultural employees and animals [7]. The design, development, and performance assessment of a cutting-edge agricultural farm prototype outfitted with gas monitoring sensors, an exhaust fan, and a disinfection system are presented in this work. The prototype promises to improve crop productivity and animal health by monitoring and managing gas concentrations in the farming environment. The study entailed controlling the speed of the exhaust fan and the cleaning procedure while measuring gas sensor data under various circumstances, such as when cow dung and acetone gas were present. The outcomes show how well the technology works to lower gas levels within the chamber.

II. SYSTEM DESIGN AND COMPONENTS

A. The Glass Chamber Prototype's Description

We have created a glass chamber with an inlet and an outlet valve that serves as the prototype for an agricultural farm. Three gas sensors are located within the chamber to keep an eye on the gases there. An exhaust fan is also included to draw gases out of the chamber. Additionally, a pump is fastened to the top (inside the glass container). To spray a disinfectant within the chamber, a pump is attached. The prototype offers a clear enclosure and enables accurate inside viewing. Fig. 1 shows the prototype used in the laboratory.

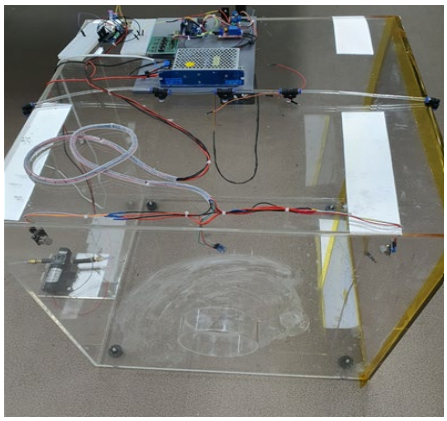


Fig. 1. Manufactured agricultural farm prototype(size: 72×72×72 cm)

B. Gas Sensors for Monitoring

Gas monitoring sensors that are carefully positioned within the chamber are an essential part of the prototype. These sensors continually monitor a variety of gases, such as smoke, natural gas, alcohol, CO, ethanol, H_2, NH_3, sulfur (S),

Identify applicable funding agency here. If none, delete this text box.

and C_6H_6. The monitoring system provides data on gas concentrations in real-time, allowing for quick reaction to possible problems. Different types of gas sensor that are used in the experiment are shown in Fig. 2.

Fig. 2. Different types of sensor used in the experiment.

C. Exhaust Fan to Remove Gas

An exhaust fan is incorporated into the system to actively remove gases from the chamber. Because of the fan's varied speeds, it is possible to precisely manage the airflow and, as a result, the dilution of dangerous gases. This element is necessary to maintain ideal gas levels and to provide a secure and safe environment. Exhaust fan is shown in Fig. 3.

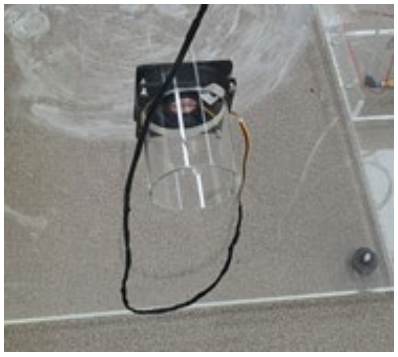


Fig. 3. Exhaush fan attached to the outlet valve of glass chamber prototype.

D. System for Disinfection Based on Electrolysis

We have developed an electrolysis-based disinfection technology to further enhance the agricultural environment. This technology produces a disinfectant solution using water electrolysis, offering a secure and long-lasting way to maintain cleanliness inside the chamber. Water molecules are split into hydrogen and oxygen gases by the process of electrolysis, which produces a potent disinfectant solution. The disinfectant system is shown in Fig. 4.

Fig. 4. Developed disinfectant technoloy.

III. Methodology

A. Gas Sensor Configuration and Data Collection

The technique used includes calibrating gas sensors for precision. These sensors were put in the chamber in a calculated manner to continuously monitor the gas concentrations. Data on gas concentrations were regularly gathered by real-time data gathering equipment. An extensive dataset for analysis was made possible by the collection of gas sensor data for a variety of gases, such as CH4, natural gas, alcohol, CO, ethanol, H_2, NH_3, sulfur (S), C_6H_6, CO_2, and other harmful gases.

B. Exposure to Acetone Gas and Cow Dung

We carried out experiments introducing cow poo and acetone gas into the chamber to imitate real-world circumstances and evaluate the response of the device. Figure 5 shows the picture of acetone contained jar and cow dung.

Fig. 5. Acetone and cow dung used as gas source inside the chamber.

C. Control of the Disinfection Pump and the Exhaust Fan

In order to explore the effect of varied fan speeds on gas reduction, fan speed control was a crucial aspect of the trials. The operation of the disinfection pump was also carefully observed to assess how it contributed to the preservation of a clean and sanitary atmosphere. Figure 7 shows the running of disinfectant spray and exhaust fan.

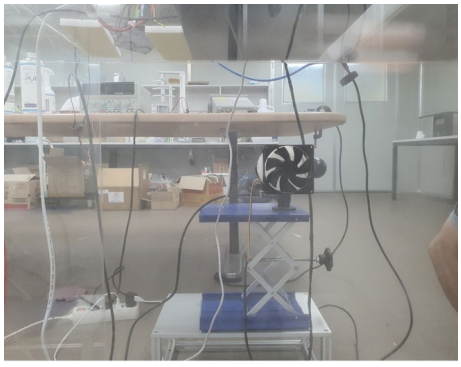


Fig. 6. Running of disinfectant spray and exhaust fan.

D. Data Recording Including Exhaust Fan Wind Speed

An extensive dataset was compiled during the course of the tests, including changes in gas concentration, fan speed settings, disinfection cycles, and wind speed. An anemometer was used to measure wind speed, which revealed information on the dispersion of disinfectant and gas dilution.

IV. Result and Discussion

A. Display of Gas Sensor Data for Various Situations

The data gathered by the gas sensors under various conditions is highlighted in this part as we share the results of our tests. Based on sensor data, Table I displays gas concentration under various glass chamber conditions. The data from the gas sensors shows diverse patterns in response to the addition of acetone gas and cow manure. These patterns shed important light on the system's capacity to identify and control gaseous pollutants inside the chamber.

TABLE I. Gas concentration over time and different scenerios

Insert acetone & cow dung inside the chamber and close all inlet and outlet valve			
Time (minute)	Sensor_1 data	Sensor_2 data	Sensor_3 data
0	145	38	82
10	160	42	85
Inlet and outlet valve open keeping exhaust fan off			
0	160	42	85
10	155	40	83
Inlet and outlet valve open with exhaust fan turned on (wind speed 5 m/s)			
0	155	40	83
0.5	290	90	130
10	200	60	90
Inlet and outlet valve open with exhaust fan turned on and gas source removed (wind speed 5 m/s)			
0	200	60	90
10	145	38	82

B. Effectiveness of the Disinfection System

The electrolysis-based disinfection technology used in our investigation improved the chamber's cleanliness. The system was assessed based on how well it dispersed disinfectant. The concentration of foul gas diminishes with time as seen by the accompanying graph (See Fig. 7), which we spray with disinfectant.

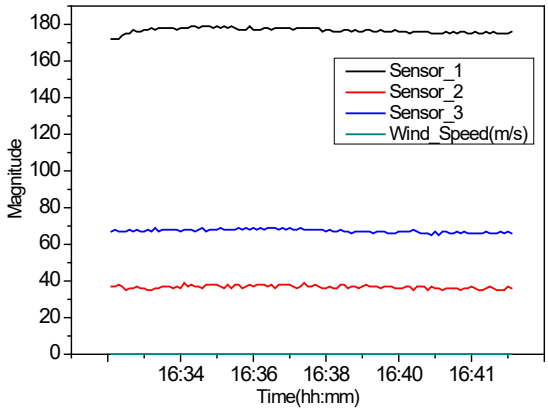


Fig. 7. Graphical representation of gas concentration after disinfectant sparying over time.

C. Relationship Between Gas Removal and Wind Speed

We looked studied the relationship between fan-induced wind speed and gas removal effectiveness to better understand the function of the exhaust fan in gas removal. The wind speed sensor continually monitors and logs the airflow rate produced by the exhaust fan while it runs. To link ventilation rates with changes in gas concentrations, the sensor data gathering procedure is timed to the data collection process for gas sensors. Fig. 8 shows sensor data for an average wind speed of 5 m/s. When the wind speed is 2 m/s (as illustrated in Fig. 9), gas is eliminated more slowly than when the wind speed is 5 m/s.

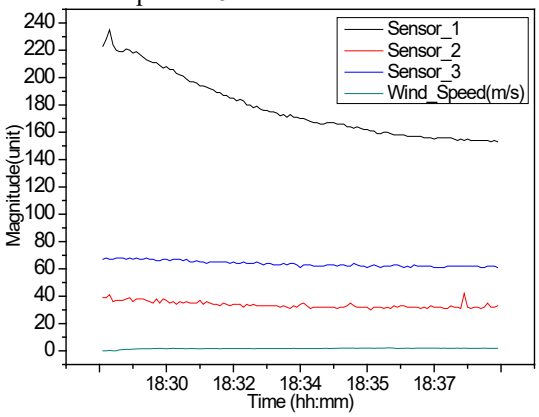


Fig. 8. Gas concentrations with time at wind speed 5 m/s

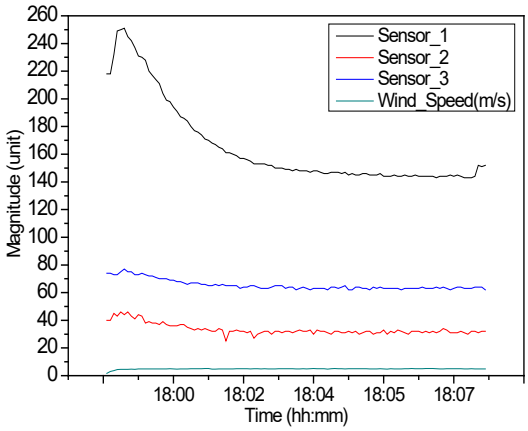


Fig. 9. Gas concentrations with time at wind speed 2 m/s

D. Discussion

The findings in this section highlight how well our integrated system works to lower dangerous gas concentrations within the agricultural chamber. The potential of our prototype to encourage healthy agricultural practices is highlighted by its capacity to detect and respond to gaseous pollutants, manage fan speed for effective gas removal, and use an electrolysis-based disinfection system. The relationship between wind speed and gas removal clarifies further aspects of the system's environmental control processes.

V. Conclusion

Our agricultural farm prototype, which is outfitted with gas monitoring sensors, an exhaust fan, and a disinfection system based on electrolysis, has been demonstrated to be both capable of and successful in use. This is thanks to the thorough analysis reported in this paper. The agriculture sector may benefit greatly from our cutting-edge farm prototype. It covers a number of pressing issues, such as the requirement for greater animal welfare, optimal gas levels, and ecologically sound agricultural methods. Although the findings of our research are encouraging, there are still a

number of directions that might be explored in this area in the future, including the incorporation of AI, alternative disinfection methods, energy efficiency, scalability for commercial usage, and disease control. We are in a position to change agriculture's methods in the future to be more efficient, sustainable, and ecologically friendly as we continue to develop and build upon these ideas.

ACKNOWLEDGMENT

This work was financially supported by the Ministry of Trade, Industry & ENERGY(MOTKE) through the fostering project of project for industry-university convergence district build-up.

REFERENCES

[1] Bajgai, Johny, Cheol-Su Kim, Md Habibur Rahman, Ailyn Fadriquela, Trinh Thi Thuy, Soon-Bong Song, and Kyu-Jae Lee. "Application of New Concept Disinfectant, Huureka®, On Livestock Farming." Korean Journal of Waters 8 no. 1 (2020): 1-12.

[2] Hill, Courtney L., Jamie D. Harris, Sydney S. Turner, Kathryn L. Wason, Amanda P. Gaylord, Maya G. Hatley, Lance T. Hardcastle et al. "Field and laboratory assessment of a new electrolytic point-of-use water treatment technology." Water 14, no. 7 (2022): 1077.K. Elissa, “Title of paper if known,” unpublished.

[3] Gowda, N. A., and Dronachari Manvi. "Agro-residues disinfection methods for mushroom cultivation." Agricultural Reviews 40, no. 2 (2019): 93-103.

[4] Dubey, Priyanka, Anupama Singh, and Owais Yousuf. "Ozonation: An Evolving Disinfectant Technology for the Food Industry." Food and Bioprocess Technology 15, no. 9 (2022): 2102-2113.

[5] McCrory, D. F., and P. J. Hobbs. "Additives to reduce ammonia and odor emissions from livestock wastes: a review." Journal of environmental quality 30, no. 2 (2001): 345-355.

[6] Ubeda, Y., Petra Amparo Lopez-Jimenez, J. Nicolas, and S. Calvet. "Strategies to control odours in livestock facilities: a critical review." Spanish Journal of Agricultural Research 11, no. 4 (2013): 1004-1015.

[7] Rahman, S. M. E., Imran Khan, and Deog‐Hwan Oh. "Electrolyzed water as a novel sanitizer in the food industry: current trends and future perspectives." Comprehensive Reviews in Food Science and Food Safety 15, no. 3 (2016): 471-490.

Efficient business process management and achievement of management objectives through the development of a process for the integration of ERP and PMS systems

Kim, Junyeop and Kim, Dongkyun*

School of Computer Science and Engineering, Kyungpook National University
80 Daehakro, Bukgu
Daegu, Republic of Korea
topbrain@knu.ac.kr, dongkyun@knu.ac.kr

Abstract

This paper designs and implements an Interface Process Module between PMS (Project Management System) and ERP (Enterprise Resource Planning), which serve different purposes but play crucial roles in organizational growth and performance achievement. PMS involves planning and executing tasks with temporary and limited scope and objectives, efficiently managing resources such as schedule, budget, and personnel in real-time to achieve project goals. On the other hand, ERP provides a systematic management approach, encompassing accounting, finance, human resources, manufacturing, and distribution management, to achieve the overall goals of the organization, characterized by the attribute of closure that distinguishes the beginning and end. By integrating the interfaces of these two systems, successful project execution within an organization can lead to the establishment of organizational management strategies and the achievement of management objectives. To accomplish this, a process is proposed that harmoniously combines project management methodologies with strategic management across the organization.

Keywords- common interests; PMS, ERP, CMMI, PMBOK, Interface Process Module, Qualitative Effects, Quantitative Effects

I. Introduction

In the era of knowledge management in the 21st century and the Fourth Industrial Revolution, organizations are adopting and utilizing various types of decision-making systems. Through the use of these systems, efforts are being made to make rapid decisions and avoid making incorrect decisions. Drawing lessons from cases where a past PDA terminal manufacturing company wielded significant influence over the organization's fate through a misguided decision, only to disappear from the industry ecosystem in an instant with the advent of smartphones, organizations are restructuring from departmental units to teams and from teams to project units. This has brought about flexibility in the fluidity of project tasks and the utilization of organizational resources. However, it has also led to complications in the overall communication system during the course of project implementation.

Furthermore, the majority of organizations use ERP systems that focus on distributing information about the results of business activities, making it inefficient in responding effectively to real-time communication and risk management.

II. The background of the research

The objective of this research is to design and implement Best Practices based on development cases utilizing widely adopted ERP systems and the PMS Interface based on MS Project within organizations. The project management techniques in this study will be implemented based on the PMBOK framework. Additionally, it will facilitate easy and diverse search and output of results and status for tasks and projects that have been carried out. Furthermore, through performance evaluation and post-management of various project outcomes, the aim is to establish a Legacy System that enables the organization to make mid- to long-term decisions.

The ultimate goal is to develop a system integration process between ERP and PMS through efficient allocation and sharing of resources (personnel, time, cost, deliverables) possessed by the organization for basic task and project planning, progress tracking, post-management, and legacy system implementation.

III. Design and Implementation of Proposal Idea Module

The integrated system of PMS and ERP allows for the management of project-related information such as costs, resources, and inventory within the ERP system, while simultaneously collecting real-time updates on the progress of projects in the ERP system. This maintains transparency in project execution. Furthermore, if functions of ERP system, such as accounting, financial management, HR management, manufacturing, and distribution management with closing characteristics, are complemented with real-time information from PMS system, including project planning, task management, communication, and document management functions, through the integration of ERP and PMS, businesses can manage their operational processes more efficiently. They can also integrate and analyze information related to project execution, enabling informed decision-making. The logical design is defined as follows

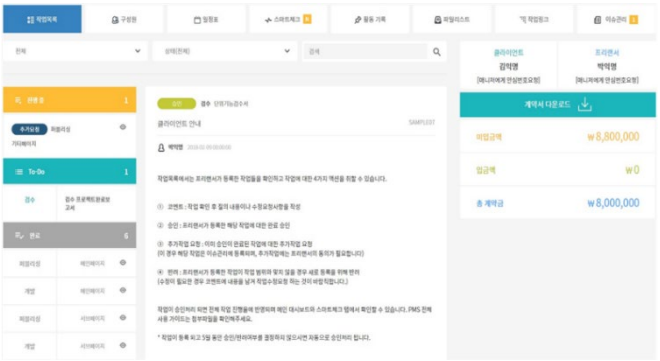


Fig. 1 PMS Package Configuration

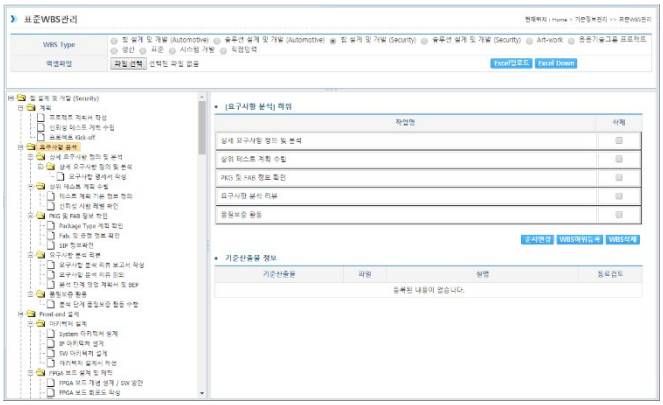


Fig. 2 PMS Dash Board

IV. Utilizing the Proposal Idea Module: Case-based System Benefits

By utilizing this module, let's examine how the system provides benefits in the following cases:

1. Enhanced Project Execution Efficiency:
   - The integration of Project Management System (PMS) and Enterprise Resource Planning (ERP) ensures real-time updates on project progress, leading to efficient project management.
2. Cost Management and Budget Adherence:
   - Through the ERP system, project-related cost information is systematically managed, enabling adherence to budgets.
3. Resource Optimization:
   - By managing resources through the ERP system, the necessary resources for project execution can be optimized and utilized efficiently.
4. Real-time Information Provision:
   - By integrating real-time information from the PMS system with the functionalities of the ERP system, a variety of information related to project execution can be provided in real-time.
5. Decision Support:
   - Analyzing integrated information allows for better decision-making regarding project execution.

A. Converged System Functionality Analysis

We collected product-specific feature data targeting vendors applying the interface of Microsoft's MS Project and MS Teams, and analyzed the effects based on the tasks recommended in the international standard model (PMBOK).

NO	Task	improvement rate
1	Improved Search Skills	80%
2	Reduced Information Delivery Time	75%
3	Reduced Design Defect Rate	40%
4	Reduced Design Modification Period	35%
5	Reduced Product Design Period	25%
6	Cost Saving in Development	20%
7	Reduced Production Cost	10%

B. Analysis of the Effects of Converged Systems Implementation

NO	Effect of Application	Saving effect / Y
1	Time saving	360h
2	Reduced Field Claims	12Case, 96h
3	Systematic Management	72h
4	Reduction of Quality Issues	41 Case, 328h
5	Time Reduction	576day, 4,608h

For the analysis of the effects of applying the converged system, we conducted interviews with project and business management personnel involved in the application of the system, focusing on key management tasks. Through these interviews, we gathered and analyzed data on the reduction of efforts as follows:

V. Conclusion

This study aimed to understand the current status of integrated systems being adopted and implemented in actual organizations. By analyzing cases of systems implemented in real organizations, it was found that the qualitative effects of successful execution of tasks and projects are likely to be higher when applying the Integrated System Fusion Process Package. Additionally, a quantitative analysis revealed cost savings in terms of actual time saved and consequent reduction in cost of direct and indirect expenses. However, it appears that there may be variations in the utilization of functions depending on the industry, suggesting the need for improvements in integrated systems. Furthermore, the flexible interface with products other than those used in this study is also an area that should be researched in the future.

Acknowledgments

"This research was supported by the MSIT(Ministry of Science and ICT), Korea, under the Innovative Human Resource Development for Local Intellectualization support

program(IITP-2023-RS-2022-00156389) supervised by the IITP(Institute for Information & communications Technology Planning & Evaluation)”

References

[1] Jeon, Hyeong-Kwon, Park, Gu-Rak, Jung, Jin-Young. "A Study on the Introduction and Application Cases of Software Development Project Management Systems". Journal of The Korea Society of Computer and Information. Vol. 20, No. 5, May 2015.

[2] Kim, Myeong-Ok, Park, Eun-Byeol. "Design and Implementation of a Web-Based Information Reporting System for Group Project Management Support". Journal of the Korea Electronic Commerce Association. Vol. 15, No. 4, 2010.

[3] Kim, Seok-Hoon, Jung, Jin-Young, Ahn, Woo-Young. "Performance Management System for Small and Medium-Sized Enterprise Projects in the MIS Environment". Journal of the Korea Navigation Institute. Vol. 13, No. 6, December 2009.

[4] Park, Sang-Pil, Jung, Il-Jae, Yeom, Hee-Gyun, Hwang, Seon-Myeong. "Establishment of a CMMI-Based Project Management System (ML2-PMS)". Proceedings of the 25th Korea Information Processing Society Spring Conference. Vol. 13, No. 1, pp. 211~214, 2006.

[5] Project Management Association. "Project Management Body of Knowledge". 2000.

A Study on Hybrid Approach Combining Agriculture and Education through Digital Twin

JinJu Lee
Department of General Design
Dongseo University
Busan, Republic of korea
pearl.0199.28@gmail.com

Kwanseon Hong
Department of General Design Dongseo University
Busan, Republic of korea
cigahks@naver.com

***Abstract*— This research paper aims to explore a hybrid approach that integrates agriculture and education through digital twins. With challenges like population decline, economic difficulties, and shortcomings in education and other infrastructures in rural areas, this study proposes leveraging metaverse and digital twin technologies as innovative solutions. By analyzing the cases of agricultural education via the metaverse and digital twins, we have derived the advantages and constraints of these technologies. In conclusion, this research suggests the potential to revolutionize agricultural education through a hybrid approach utilizing digital twins and the metaverse, achieving both the development of rural areas and enhancement of education. This strengthens the connection between rural communities and schools. Continuous research based on these findings could offer substantial insights for the sustainable development of rural areas and the innovation of education in the future.**

Keywords—Digital twins, Metaverse, Agricultural education, Rural development, Sustainable development

I. INTRODUCTION

Owing to urban population concentration, rural areas are currently grappling with issues like population decline, economic challenges due to aging demographics, and difficulties in transportation, healthcare, and education. Such challenges, including the lack of infrastructure, are leading to serious problems like population extinction, thereby deteriorating the quality of life for individuals residing in rural areas. Projects utilizing the metaverse are being proposed as new solutions to address these rural challenges. While the metaverse holds potential to connect people and resources in novel ways, its virtual nature presents constraints in comprehending the intricate processes of agriculture. Especially, providing meaningful education through pure virtual agricultural simulations has its limitations. In this context, digital twin technology emerges as an innovative tool bridging the real-life rural experience and the boundless possibilities of the digital realm. This study proposes an educational method transcending physical boundaries by integrating the rural experience with digital twin technology. Through this, we aim not only to preserve and enhance aspects of rural life utilizing the power of technology but also to offer educational opportunities that allow individuals and communities to thrive in the contemporary digital era. We believe that through this approach, rural communities and schools can be mutually connected, leading to sustainable growth, knowledge sharing, and renewed hope for rural communities.

II. DEFINITION OF METAVERSE AND ITS EDUCATIONAL APPLICATION

The term "metaverse" is a compound of "meta" and "universe," signifying a virtual world that transcends reality by allowing interactions between the virtual space and the real world[1]. This space offers users opportunities for experiences and interactions that are impossible in reality, blurring the boundaries between the real and the virtual through an integrated system of the virtual world. It represents a convergent space reflecting the demands of the public and technological advancements[2]. The Acceleration Studies Foundation (ASF)[3] has classified the types of metaverse based on augmented, simulation, external, and internal axes into four categories: Augmented Reality, Life Logging, Mirror World, and Virtual World. The definition of these categories, their technical features, and their potential for educational application are detailed below[4]. Table 1 interprets the categories classified by ASF from an educational perspective.

TABLE I. EDUCATIONAL UTILIZATION AND LIMITATIONS OF AGRICULTURE USING THE METAVERSE

TYPE	***Definition & Technical Features***	***Potential for Educational Utilization***	***Limitations***
IoT	Provides real-time data connection between the digital twin and physical entities, using sensors and devices to collect and transmit data, updating the digital twin.	Real-time monitoring of soil, climate, and crop status using IoT sensors. Optimization of agricultural processes and efficiency improvement through automation technologies.	Basic understanding required for data collection, storage, and protection. Managing hardware and software compatibility.
Big Data & AI	Real-time data collection and analysis through sensors and IoT, precisely identifying the status and environment of the target.	- Monitoring crop growth, predictive analysis, utilizing agricultural data. AI image analysis for detecting and diagnosing crop diseases and pests.	- Complexity in data management and processing. Cost of hardware and software.
Simulation	Simulation and experimentation of physical entities, modeling and optimizing entity behavior in various scenarios.	- Allows virtual attempts in situations that are hard to experiment with in real agricultural environments. - Provides virtual agricultural environ	- Lack of real experience. - Limitations of training in virtual environments.

This research was supported by BK21 Four Service Design driven Social Innovation Educational Research Team in Dongseo University.

		ment for learning crop management, work planning, cultivation techniques.	
3D Visualization	Using visualization technology to realistically visualize complex physical entities, enabling users to visually understand and observe the entity.	-Simplifies understanding of complex agricultural systems. - Provides a virtual environment for agricultural education and simulation.	- Possible discrepancies between 3D models and real environments. - Accessible through the use of AR and VR equipment.
AR& VR& MR	Combining reality and virtuality using XR technology to enhance user experience. Interact with digital twin and experience real scenarios.	- Combines real and virtual environme nts to provide immersive experience for learners. - Simulate and train for agricultural tasks in real-time. - Enables collaboration and interaction with other learners.	- Special equipment and software required. - Differences exist between train ing in virtual environments and real experience. - Accessibility and cost.

Agricultural education utilizing the metaverse has several limitations. Firstly, the metaverse depicts an environment modeled and constructed based on the actual agricultural environment, making it challenging to provide a comprehensive agricultural experience. Consequently, learners might find it difficult to fully perceive the real agricultural environment. Furthermore, since the metaverse emphasizes activities in a virtual environment, it can be challenging to attain educational values and experiences, such as the patience and rewards associated with real farming in the physical world. Additionally, the absence of real-time feedback in the real world might limit the learner's experience and learning effectiveness.

III. Definition of Digital Twin and its Educational Application

Digital twin technology involves virtually replicating physical entities in the real world through software and leveraging interactions between the two spaces to create value. This technology accurately models the shape, properties, and state of physical entities in a virtual world, enabling various simulations and interactions. A digital twin represents a physical object or system and, through sensors, can reflect the current state of that physical entity in real-time or be used to predict future situations[5]. Digital twin technology can be divided into five main technical domains. Table 2 interprets the main technologies of digital twins from an educational perspective.

TABLE II. Educational Utilization and Limitations of Agriculture Using Digital Twins

TYPE	***Definition & Technical Features***	***Potential for Educational Utilization***	***Limitations***
IoT	Provides real-time data connection between the digital twin and physical entities, using sensors and devices to collect and transmit data,	Real-time monitoring of soil, climate, and crop status using IoT sensors. Optimization of agricultural processes and efficiency	Basic understanding required for data collection, storage, and protection. Managing hardware and
	updating the digital twin.	improvement through automation technologies.	software compatibility.
Big Data & AI	Real-time data collection and analysis through sensors and IoT, precisely identifying the status and environment of the target.	- Monitoring crop growth, predictive analysis, utilizing agricultural data. AI image analysis for detecting and diagnosing crop diseases and pests.	- Complexity in data management and processing. Cost of hardware and software.
Simulation	Simulation and experimentation of physical entities, modeling and optimizing entity behavior in various scenarios.	- Allows virtual attempts in situations that are hard to experiment with in real agricultural environments. - Provides virtual agricultural environ ment for learning crop management, work planning, cultivation techniques.	- Lack of real experience. - Limitations of training in virtual environments.
3D Visualization	Using visualization technology to realistically visualize complex physical entities, enabling users to visually understand and observe the entity.	-Simplifies understanding of complex agricultural systems. - Provides a virtual environment for agricultural education and simulation.	- Possible discrepancies between 3D models and real environments. - Accessible through the use of AR and VR equipment.
AR& VR& MR	Combining reality and virtuality using XR technology to enhance user experience. Interact with digital twin and experience real scenarios.	- Combines real and virtual environme nts to provide immersive experience for learners. - Simulate and train for agricultural tasks in real-time. - Enables collaboration and interaction with other learners.	- Special equipment and software required. - Differences exist between train ing in virtual environments and real experience. - Accessibility and cost.

Technologies related to digital twins offer potential in the field of agricultural education but must consider aspects like data management, cost, and accessibility. By appropriately combining and utilizing each technology, the efficiency in the agricultural sector can be enhanced, supporting sustainable agriculture.

IV. Case Analysis

From the perspective of agriculture, cases utilizing the metaverse and digital twins were collected and analyzed. The researcher's analysis of real-life cases applying the metaverse and digital twin technologies in the agricultural sector yielded findings as presented in Table 3. These cases confirm that innovative technological applications are underway for the enhancement of agricultural education and productivity.

TABLE III. Cases Utilizing Metaverse and Digital Twin Technologies

Cases	***Applied Technologies***	***Description***	***Keyword Extraction***
FarmVille	Metaverse Virtual World	This is a virtual farm game offered on the Facebook platform. Users can operate a virtual farm, grow crops, gain experience, and	- Virtual Farming Game - Entertainment

		interact with others for entertainment purposes.	
Toriverse	Metaverse Virtual World	Using VR/XR technology, it's possible to simulate and test agriculture virtually, connecting users within a virtual reality environment to share experiences and knowledge.	- Virtual simulations - Utilization of artificial intelligence machine learning technology for crop monitoring and prediction - Connectivity and education of users
Cloud SSAS	Metaverse Augmented Reality Life Logging	A case of providing experiential agricultural education using a 3D virtual environment with VR equipment, utilizing a social metaverse platform to facilitate the exchange of agricultural-related information.	- Utilizing VR for the operation of agricultural machinery and smart farm devices. - Self-directed learning methods.
Smart Farm	Digital Twin IoT	A smart farm system that monitors and controls the real-time growth status of crops.	- Real-time monitoring. - Real-time control.
Aeroponics Farm	Digital Twin IoT BIG DATA	Supporting farm management and crop monitoring using IoT technology, automatically adjusting the environment through real-time data collection, analysis, and prediction.	- Mist-type smart farming technology - LED automation system - System maintenance through big data
small robot co	Digital Twin BIG DATA &AI, IoT	Autonomous operation providing intelligent farming and service through self-assessment of agricultural production and distribution processes and farming conditions.	- Unstructured 3D environment mapping - Improvement of AI technology integration for movement and operatio n capabilities

While this technology offers revolutionary potential in the field of agricultural education, its limitations must be considered. Agricultural education via digital twins struggles to perfectly replace the physical experience and educational efficacy obtainable in real settings. For instance, basic agricultural techniques like watering can only be fully understood and acquired by learners when actually performed in the real world. Moreover, digital twins might not be able to offer learners the sensory experiences and interactions obtainable from real-world agricultural experiences. Agriculture requires a myriad of sensations and experiences; directly feeling soil characteristics and visually confirming the growth state of crops are crucial experiences.

V. Conclusion and Implications

This paper addressed educational approaches towards the agricultural sector, analyzing various examples in the process. It was observed that rural areas face challenges such as population decline, economic difficulties, and infrastructure deficits. In response to these challenges, the metaverse and digital twin technologies have gained attention as valuable tools. However, the metaverse alone has limitations in comprehensively understanding the complex processes of agriculture. Similarly, agricultural education via digital twins has its constraints due to the lack of physical experience. Overcoming these limitations, a conclusion was derived highlighting the need for a hybrid approach, amalgamating digital twin technology with real-world rural experiences. Through this hybrid educational approach, learners can acquire theoretical knowledge and skills in the digital world, use smart farms to nurture crops, and then apply and observe this in a real agricultural environment, ultimately reaping the fruits of their labor. This ensures maximized educational benefits and a rich learning experience for the students. Modeling real agricultural environments and growing crops through digital twin devices enhance learner comprehension, providing easy access to information, thus proving to be of immense educational value. Moreover, the interconnection between rural areas and education strengthens the bond between rural communities and urban schools, fostering the potential for simultaneous sustainable development of rural areas and educational innovation. From these research findings, several implications can be derived:

Firstly, in the field of agricultural education, a hybrid approach combining digital twins and the metaverse effectively addresses the absence of real agricultural experience, offering students genuine hands-on opportunities. Through this, educational efficacy in agriculture is optimized, ensuring students acquire the skills and knowledge necessary in real-world agricultural settings.

Secondly, the mutual interconnection between rural areas and education is a crucial strategic component that propels both sustainable development in rural areas and educational innovation. Utilizing resources and agricultural experiences in rural regions to develop educational programs presents an opportunity to support and further the growth of local communities through education. By doing so, the bond between rural communities and urban schools is strengthened, contributing to the formation of new, sustainable rural communities.

References

[1] Kim Sang-gyun. (Professor Kim Sang-gyun's) Metaverse. Seoul: Donga Asia Science, 2021.

[2] Lim Sung-tae. "Development of a conceptual framework for learning analysis application in a virtual world-type metaverse learning environment." Domestic Doctoral Thesis, Chungbuk National University, 2023. Chungcheongbuk-do.

[3] The non-profit tech organization Acceleration Studies Foundation is hereafter referred to as ASF.

[4] asf 2007. Metaverse Roadmap: Pathways to the 3D Web

[5] D. Y. Jeong. (2021). Technical Definition of Digital Twin and Five-Level Modell, OSIA S&TR Journal, 34(2), 10-16.

Implementation of Digital Twins in Smart Homes and Generation of Remote Meter Data

*Jun Hyeok Jang**
School of Computer Science and Engineering
Kyungpook National University
Daegu, Republic of Korea
wkdwnsgur@knu.ac.kr

*Rinrada Tirasirichai**
Department of Computer Science
Kasetsart University
Bangkok,Thailand
rinrada.ti@ku.th

Dong Hyeon Kim
School of Computer Science and Engineering
Kyungpook National University
Daegu, Republic of Korea
kevink3@knu.ac.kr

Jin Ho Lee
School of Computer Science and Engineering
Kyungpook National University
Daegu, Republic of Korea
ghfkdrjal777@knu.ac.kr

Soon Ki Jung
School of Computer Science and Engineering
Kyungpook National University
Daegu, Republic of Korea
skjung@knu.ac.kr

***Abstract*— This paper delves into a study on the implementation of smart homes through the digitization of living spaces, modeling, and real-time simulation using digital twins. Digital twins encompass the digital modeling of physical homes and the various devices, systems, and appliances within them, facilitating the creation of a smart home. This allows for real-time monitoring, data collection, and analysis of the components of a smart home. In particular, data on power consumption can be obtained through the simulation of a smart home modeled as a digital twin. However, this is not limited to simply generating data through basic simulations; rather, it involves creating data that encompasses various scenarios by taking into account regional and individual characteristics, the number of household members, time, day of the week, seasons, and more. This technology provides homeowners and service providers with insights into energy consumption, resource utilization, and the overall performance of a home. It can enhance energy efficiency, automate processes, and improve the quality of life for residents. Moreover, the generated data can be utilized in research for training models capable of detecting anomalies, identifying patterns, and significant deviations. This can potentially be instrumental in the early detection of emergency situations within households, particularly for elderly individuals and cases requiring urgent attention.**

Keywords—Digital Twins, Smart Home, Remote Meter Data, Simulation.

I. INTRODUCTION

In modern society, advancements in medicine and the emphasis on personal hygiene have led to an increase in human life expectancy. Simultaneously, changes in the structure of households, extending beyond the traditional family unit to include single-person households, have emerged[1]. Behind these developments and changes, however, lurk darker aspects such as indifference toward single-person households like elderly individuals living alone, resource wastage, and energy inefficiency. The rise of single-person households has given rise to issues such as social isolation and the prevalence of loneliness among the elderly[2]. Moreover, resource scarcity has prompted discussions on sustainable development, which is essential for preserving finite resources and the environment while promoting human progress.

Meanwhile, the evolution of sensors, wireless networks, and artificial intelligence technologies is bringing about significant transformations in various aspects of our society[3]. Efforts are being made to leverage these advanced technologies to address the myriad of problems arising within society. One such endeavor involves the collection of power consumption data from each household and industrial zone, allowing for the analysis of electricity usage patterns in buildings[4]. This data can then be used for early detection of abnormal patterns, contributing to the safety and well-being of individuals. Furthermore, understanding the electrical characteristics, such as power consumption and standby power, of various electronic devices can provide insights into energy usage and help prevent unnecessary energy waste[5]. This, in turn, can serve as a stepping stone toward conserving electrical energy and achieving sustainable development.

Among the techniques for analyzing electrical characteristics, Non-Intrusive Load Monitoring (NILM) is a technology that dissects power usage information from various electronic devices based on integrated meter readings. Various methods utilizing neural networks are being applied and researched to advance NILM technology[6]. To further the development of NILM, numerous countries and research organizations are collecting and sharing relevant data from real residential buildings. While a variety of datasets exist from different regions and time periods, publicly available datasets may be limited in terms of the geographical area they cover, their data collection periods, or their frequency, making them challenging to use for research purposes[7].

Furthermore, the utilization of electricity exhibits regional and individual characteristics. For instance, in South Korea,

* This author contributed equally in this paper.

many households have refrigerators not only for general use but also specifically for storing kimchi, a traditional Korean dish. Additionally, the types of devices used can vary depending on the age groups of family members[8]. Hence, datasets should take into account such characteristics reflecting regional, familial, and personal attributes. However, acquiring such real-world data directly poses challenges due to issues like privacy breaches, additional equipment installation, and network errors.

To address these challenges, we aim to utilize artificially synthesized data created through a digital twin house. By implementing a digital twin of residential environments and simulating the daily lives of inhabitants in virtual residences, we can collect data on electricity consumption during those simulated activities. This approach allows us to create datasets that adequately capture regional and individual characteristics, which can then be used for subsequent research.

In this paper, we introduce the implementation of a digital twin smart home and the methodology for generating remote electricity consumption data.

II. Related Work

A. Digital twins

A digital twin is a virtual representation of an object or system that spans its lifecycle, is updated from real-time data, and uses simulation, machine learning, and reasoning to help decision-making[9]. Also, it's a virtual model designed to accurately reflect a physical object which enables the digital twin to deliver functional services and satisfy the application requirements[10]. Digital Twin has been proposed as one of the key concepts for Industry 4.0 to provide a virtual representation of products along their lifecycle that enables the prediction and optimization of the behavior of a production system and its components[11]. Through this connection, Digital Twins contribute to the convergence of the real and the virtual world. Through this connection, Digital Twins contribute to the convergence of the real and the virtual world. Digital Twins can be applied to many aspects as in this paper's study about Digital Twins in residential environments. The Digital Twin approach relies on the ability to receive and effectively process data flows collected automatically through distributed “Internet of Things” (IoT) sensor systems[12] which will be able to use the information received to clone a physical object (PO) into a software counterpart. Furthermore, can applied to account-related proposals emerging in other fields, namely augmented and virtual reality (e.g., avatars), multiagent systems, and virtualization[13].

As mentioned in above, we can convince the benefits of Digital Twin Implementation in Residential Environments in our paper study that is we can use virtual reality that we use in this paper to connect with the study about Digital Twin and we able to get data from sensors in the resident environment. So, we can compare data between sensors and data from virtual reality that we made simulation in this paper.

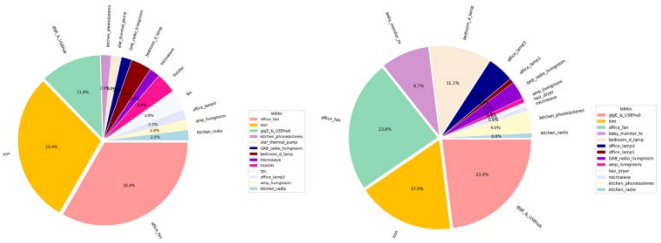


Fig. 1 UK-DALE Data Sample

B. Electricity dataset

UK-DALE, short for "UK Domestic Appliance-Level Electricity,"[14] is a public dataset that contains electricity consumption data from various households in the United Kingdom as shown in Fig1. This dataset is used to study and analyze electricity consumption patterns within households. It provides appliance-level data for electricity usage within homes in the UK[15]. The key features of the UK-DALE dataset are as follows.

1. Appliance-Level Household Data: UK-DALE records electricity usage data for various household appliances (e.g., refrigerators, microwaves, washing machines, gas heating systems, etc.) on a per-household basis. This data is valuable for analyzing electricity consumption at the appliance level and for improving energy efficiency.

2. High-Frequency Sampling: The UK-DALE dataset is collected with high-frequency sampling, with data recorded as frequently as every 6 seconds. This provides highly detailed records of electricity consumption, making it useful for analyzing consumption patterns at various time scales, including hourly, minute-level, or even shorter intervals.

3. Real-World Household Data: This dataset comprises data collected from actual households, reflecting a variety of environments and conditions within different homes. This makes it valuable for modeling and researching a wide range of real-world scenarios.

4. Removal of Private Information: To protect privacy, any identifying information has been removed from the dataset, leaving only electricity consumption data.

The UK-DALE dataset is widely used in various fields, including energy research and the development of smart grid technologies. Researchers can utilize this data to understand household electricity consumption patterns, develop energy-saving strategies, and optimize the performance of household appliances.

ENERTALK is a consumption power dataset collected from households in South Korea, with data for each home appliance being stored at a high frequency of 15 times per second[16]. Moreover, to reflect the characteristics of electricity usage in South Korean households, it includes information about appliances that have cultural significance, such as kimchi refrigerators. This enhances the validity of the collected data and holds the potential for a wide range of applications[17].

III. Methods

The whole scenario consists of: create simulation digital twin house and avatar represent humans that live in houses moreover random daily life activities for the avatar to virtuality of the data to simulate the daily lives of inhabitants. Also, collect data on electricity consumption during those simulated activities. Therefore, to obtain information that is comparable to reality

follow the Digital Twin environment that we study on. we must set up simulations on different dates times and seasons. Due to, in fact, the use of electrical appliances there are time factors and seasons.

A. Tool for Digital Twin Environment Construction

To create the simulation in this study, we used Unity. Start by creating a virtual home that conforms to the digital twin environment. So, we created 4 types of houses based on the types of houses in Korea by location we studied in digital twin. Which are apartment, share house, studio, gosiwon.

Another factor that affects electric consumption data is the number of people who live in the house in addition to affecting the type of house it will affect the style of the house such as the number of bedrooms affects to number of Air conditions. So, we determined the number of people living in each type of house to make a difference in creating houses in different styles even though they are the same type. Therefore, in each type of house, there is an additional division to create a virtual house in the number of residents as shown in Table 1.

TABLE I. Various type of house

Type of house	Apartment			
	1 People	2 People	3 People	4 People
Bedroom	1	2	3	3
Bathroom	1	2	3	2
Kitchen	1	1	1	1
Livingroom	1	1	1	1

Type of house	Share House		Studio	Gosiwon
	3 People	4 People	1 People	1 People
Bedroom	3	4	1	1
Bathroom	1	1	1	1
Kitchen	1	1	1	-
Livingroom	1	1	-	-

B. Statistics data

As we mentioned, data that we need for the simulation is electricity consumption data that includes power consumption, and time used by each electric appliance. To do a simulation to get those data basic information must be specified for each type of electrical appliance. So, the basic data that we needed to assign each electric appliance in our simulation include power consumption, time used in each time, number of times used in a day, number of times used in a month, probability of use in a day, and average use case of each electric appliance. Therefore, to get the data mentioned above we needed a dataset of electric appliances that included data we needed for calculation. Thus, the KPX (Korea Power Exchange) statistics data has information data that we needed to calculate for each type of electrical appliance such as power consumption, number of days used per month, and usage time per day as shown in Fig. 2 and Table. 2. Furthermore, KPX is in control of the operation of

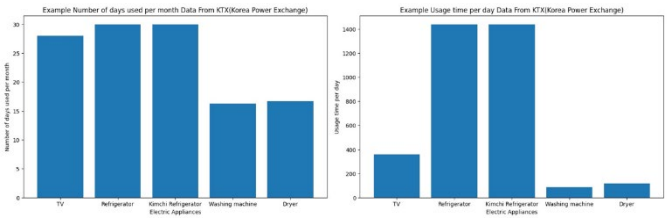


Figure. 2 Example of Data in KPX

Korea`s electricity market and the power system is also under the control of the Korean Government making the dataset that we get from KPX updated and reliable. Also, as mentioned above Korea is a country that we studied in Digital Twin so that is another reason that KPX is suitable for this paper study.

TABLE II. Example of Data in KPX

Electric appliance	Power consumption	Number of day used per month	Usage time per day
TV	151.5	28.0	358.8
Refrigerator	42.2	30.0	1440.0
Kimchi Refrigerator	19.3	30.0	1440.0
Washing machine	904.3	16.3	87.0
Dryer	1348.6	16.7	117.4

C. Scenarios of Lifestyes

According to electricity consumption data that we needed. For realism, as in Digital Twin, we created an Avatar that assumes a human who lives in that house and random daily activity that corresponds to the use of electrical devices from a life scenario spreadsheet that was created from calculating data from KPX that is mentioned above. When referring to chance, the probability of being able to use this electrical device in one day. Additionally, the usage scenario spreadsheet for each electrical device has a separate scenario for different usage times so that randomization is as realistic as possible as shown example in Table 3.

Table III.Example of Data in KPX I

Home appliances	using time (minute)	Powe usage (watt)	Monthly usage days	probability of use
Tv	60	151.5	28	0.933333333
Tv	120	151.5	28	0.933333333
Tv	358.8	151.5	28	0.933333333
Air conditioner	30	1598	23.3	0.776666667
Air conditioner	120	1598	23.3	0.776666667
Air conditioner	150	1598	23.3	0.776666667

In addition to the above the randomization conditions also depend on the season and time of day which are also determined for the randomization such as in winter season use a Humidifier so don't use a Dehumidifier but on the other hand in summer use a Dehumidifier so didn't use Humidifier. However, all the steps

mentioned above were done in Unity. Therefore, when assigning random activities to avatars, they will have movements and gestures corresponding to that activity. This makes the Digital Twin House simulation more realistic. Regarding collecting data on electricity consumption, the aim of this paper study we also collect Avatar activity based on usage time and electricity consumption of each electrical device and summarize it in CSV form.

IV. Results

A. Configuring a Digital Twin Environment

According to the above, we create virtual houses in 4 types and some types are based on reality, there can be more than 1 resident, so we create more virtual houses for style that correspond to number of people that live in as shown in Fig 2,3,4. Additionally, for the Avatar simulation that we did from a random daily activity from the Life Scenarios spreadsheet, we also used Unity in the same way as creating a virtual home. But because of randomness has various conditions such as the probability of using that type of electrical appliance, season, and timing. In order to define these conditions, we have to create a C# script to assign those conditions to each electrical appliance and for avatar movements and gestures to correspond with the type of electrical appliances.

B. Data Synthetic

According to the aim of this paper, the study is to get electricity consumption data that simulates Digital Twin houses. Thus, when the simulation ends the synthetic data is collected from the simulation. The data is divided according to the type of house. It contains details of the use of electrical appliances in each room. As an example, Table 4.

Table IV. Example of Synthetic Data

Electrical Appliance in Apartment 1 people		**Date/Time (Start Used)**			
		Time used (minute)			
Light	Bedroom	10/25/2023 09.10	10/25/2023 14.00	10/25/2023 20.00	10/26/2023 07.00
		20	50	180	90
	Bathroom	10/25/2023 09.30	10/25/2023 20.10	10/26/2023 07.10	
		15	25	15	
Air Condition	Bedroom	10/25/2023 14.10	10/25/2023 20.10	10/26/2023 07.10	
		30	160	40	
	Livingroom	10/25/2023 10.10	10/25/2023 17.00	10/26/2023 09.10	
		120	90	20	

Figure. 3 Apartment(1 Bedroom for one person and 1 Bedroom for 2 people)

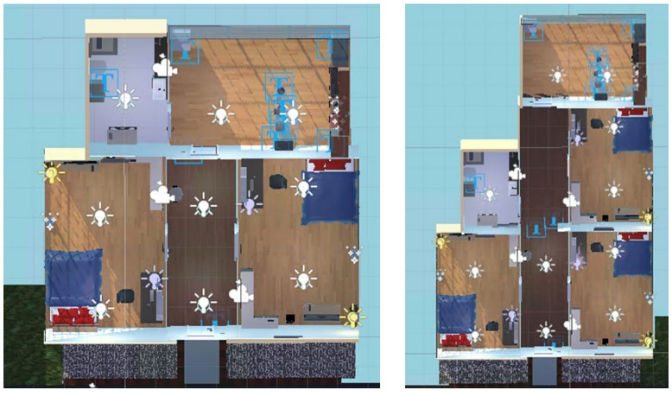


Figure. 4 Share house(2 Bedroom for 2 people and 2 for Bedroom 3 people)

Figure. 5 Studio and Gosiwon

C. Generation of Synthetic Data

when the simulation ends, we will get synthetic data as mentioned above. Therefore, we have created a C# script for visualizing synthetic data that we get by calculating the total power consumption and total used time for every electrical appliance in each type of house. The principle calculated total power consumption and total used time for each electrical appliance by taking the results from synthetic data that collected the duration of use of that electrical appliance at each time and combining them to get total used time. Then, use that total used time multiplied by 60 to convert minutes into seconds. Because of power consumption unit is watt, it`s 1 joule per second after converting used time from minute to second. Then, multiply with power consumption(watt) of that electrical appliance to get the total power consumption of that electrical appliance.

After that all the result that already visualizing will be writing on to CSV format to make it easier to use and analyze the results further as shown example in Table 5.

Table V. Generation of Synthetic Data

Apartment 1 people		
Electrical Appliance	**Total Time Used (minute)**	**Total Power Consumption (Watt)**
Light	840	277,200
Air Conditioner	600	57,528,000
TV	440	3,999,600
Phone Charger	300	297,000
Microwave	15	1,045,350

Share House 2 people		
Electrical Appliance	**Total Time Used (minute)**	**Total Power Consumption (Watt)**
Light	1680	5,544,000
Air Conditioner	1220	116,973,600
TV	600	5,454,000
Phone Charger	500	495,000
Microwave	30	2,090,700

V. CONCLUSION

In this study, we have implemented a Digital Twin House and developed a method for generating synthetic datasets of household electricity consumption by mapping the power consumption information of various household appliances and applying diverse daily life scenarios. We utilize this tool to create datasets that reflect regional and individual characteristics, which can then be employed to discover patterns and characteristics in electricity usage for each household. In the future, we plan to investigate the validity of the generated data through a comparison between synthetic and real data, and we intend to conduct subsequent research, including the implementation of NILM neural networks using synthetic data.

ACKNOWLEDGMENT

This work was supported by the Digital Innovation Hub project supervised by the Daegu Digital Innovation Promotion Agency(DIP) grant funded by the Korea government(MSIT and Daegu Metropolitan City) in 2023(No. DBSD1-04, Smart management system for responding to lonely deaths of elderly living alone based on remote meter reading information and CCTV access information).

REFERENCES

[1] Janssens, Angélique. Family and social change: The household as a process in an industrializing community. No. 21. Cambridge University Press, 2002.

[2] Yeung, Wei-Jun Jean, and Adam Ka-Lok Cheung. "Living alone: One-person households in Asia." Demographic research 32 (2015): 1099-1112.

[3] Nguyen, Dinh C., et al. "Enabling AI in future wireless networks: A data life cycle perspective." IEEE Communications Surveys & Tutorials 23.1 (2020): 553-595.

[4] Chou, Jui-Sheng, and Abdi Suryadinata Telaga. "Real-time detection of anomalous power consumption." Renewable and Sustainable Energy Reviews 33 (2014): 400-411.

[5] Kazandjieva, Maria, et al. "Identifying energy waste through dense power sensing and utilization monitoring." Computer science technical report CSTR 3 (2010).

[6] Kelly, Jack, and William Knottenbelt. "Neural nilm: Deep neural networks applied to energy disaggregation." Proceedings of the 2nd ACM international conference on embedded systems for energy-efficient built environments. 2015.

[7] Ruano, Antonio, et al. "NILM techniques for intelligent home energy management and ambient assisted living: A review." Energies 12.11 (2019): 2203.

[8] Sa'ad, Suleiman. "Electricity demand for South Korean residential sector." Energy policy 37.12 (2009): 5469-5474.

[9] Tao, Fei, et al. "Digital twin modeling." Journal of Manufacturing Systems 64 (2022): 372-389.

[10] Singh, Maulshree, et al. "Digital twin: Origin to future." Applied System Innovation 4.2 (2021): 36

[11] Tao, Fei, et al. "Digital twin modeling." Journal of Manufacturing Systems 64 (2022): 372-389.

[12] Schroeder, Greyce N., et al. "A methodology for digital twin modeling and deployment for industry 4.0." Proceedings of the IEEE 109.4 (2020): 556-567.

[13] White, Gary, et al. "A digital twin smart city for citizen feedback." Cities 110 (2021): 103064.

[14] KELLY, Jack; KNOTTENBELT, William. The UK-DALE dataset, domestic appliance-level electricity demand and whole-house demand from five UK homes. Scientific data, Vol. 2, Num. 1, pp. 1-14, 2015

[15] Sirojan, Tharmakulasingam, B. Toan Phung, and Eliathamby Ambikairajah. "Deep neural network based energy disaggregation." 2018 IEEE International conference on smart energy grid engineering (SEGE). IEEE, 2018.

[16] SHIN, Changho, et al. The ENERTALK dataset, 15Hz electricity consumption data from 22 houses in Korea. Scientific data, Vol. 6, No. 193, 2019.

[17] Ayub, Mohammed, and El-Sayed M. El-Alfy. "Multi-target energy disaggregation using convolutional neural networks." International Journal of Advanced Computer Science and Applications 11.10 (2020).

Exploring Street view API connection to improve user experience in digital twin environment

Dong Hyeon Kim
School of Computer Science and Engineering
Kyungpook National University
Daegu, Republic of Korea
kevink3@knu.ac.kr

Rinrada Tirasirichai
Department of Computer Science
Kasetsart University
Bangkok,Thailand
rinrada.ti@ku.th

Jun Hyeok Jang
School of Computer Science and Engineering
Kyungpook National University
Daegu, Republic of Korea
wkdwnsgur@knu.ac.kr

Hyeon Hoo Hwang
School of Computer Science and Engineering
Kyungpook National University
Daegu, Republic of Korea
hhh2022@knu.ac.kr

Soon Ki Jung
School of Computer Science and Engineering
Kyungpook National University
Daegu, Republic of Korea
skjung@knu.ac.kr

***Abstract*—This paper focuses on the integration of Street View API to enhance user experience within the digital twin environment. Current metaverse experiences often lack a sense of realism, and achieving realism through extensive modeling can be costly. Therefore, this study aims to improve user experience by enhancing realism using Street View instead of comprehensive modeling. We explore methods to seamlessly integrate real-world imagery into the digital twin, with the goal of enhancing the interaction between virtual and real worlds. Leveraging two currently available Street View APIs, we successfully achieved integration between the digital twin and the real environment, setting development directions based on the API formats and analyzing their respective advantages and disadvantages. This research proposes innovative approaches to more realistically represent the real world within the metaverse environment, contributing to the implementation of new strategies for enhancing the user experience. Future research plans include improving the functionality of the program developed in this study and further investigating elements to increase user satisfaction.**

Keywords—Digital Twin, Metaverse, Street View

I. INTRODUCTION

The global pandemic caused by COVID-19 in the past has emphasized the need for non-contact interactions, leading to the integration of online meetings, virtual worlds, and other new communication platforms into our daily lives[1]. This change has provided significant momentum for research into related technologies and concepts, with a substantial increase in interest in the metaverse across various sectors of society[2]. The metaverse represents a new digital environment where the real and virtual worlds converge, offering new experiences and opportunities[3]. As a result, active research and applications are being carried out in various fields, including industry, education, and the economy.

However, despite considerable research and development in the metaverse field, there is still a lack of cases where the metaverse completely replaces human daily life or dramatically improves users' digital life experiences. Consequently, there is a consensus that more research and development are needed to fully harness the potential of the metaverse[4].

This study primarily focuses on the core concept of the metaverse, known as the "digital twin" environment. Digital twin refers to the concept of virtually replicating physical objects and environments from the real world, allowing users to experience innovative interactions between the real and virtual worlds[5]. In particular, this research aims to explore methods to enhance the user experience in the digital twin environment of industrial settings. In other words, it seeks to investigate how the functionality of street view, a common feature in commercial mapping platforms, can be utilized to enhance the sense of presence in virtual environments.

This research is expected to shed light on the future potential of the metaverse and promote innovation in various fields where digital twin technology can be applied by providing innovative experiences in a digital environment. Subsequent sections will delve into the research methods and results to achieve these objectives in more detail.

II. RELATED WORK

A. User Interface and User Experience

A User Interface (UI) refers to the interaction between a system and a user through commands or techniques to operate the system, input data, and use its contents. UI encompass a wide range of domains, from systems such as computers, mobile devices, and games to application programs and content usage. User Experience (UX) pertains to the overall experience related to the emotions, thoughts, reactions, and behaviors a user feels and thinks when directly or indirectly using a system, product, content, or service. UX is a concept related to Human-Computer Interaction (HCI) and is widely applied not only in software and hardware development but also in services, products, processes, society, and culture. UI/UX serves as an interface through which a person can interact with a system or application in a computer

and communication environment and is categorized into software and hardware interfaces[6]. In other words, UI refers to the planning of the program's flow to support users and the planning to effectively assist users in their tasks, serving as the communication medium that connects computers and users. If UI serves as the element connecting computers and people, then UX can be understood as the user experience, referring to the emotions that users can feel when experiencing a product or service. And some research[7] discusses that UX design aims to design in order to satisfy the various experiences such as perception, reactions, and behaviors that users go through as they use a service. UX is defined as the participation, usage, observation, and mutual interaction of users[8]. Based on this, in this study, we aim to design UI tailored to the A virtual world that digitizes a manufacturing factory into a digital twin to create a better UX. We have focused on enhancing the aspect of realism as a method to improve the UX, and, as a means to achieve this, we intend to leverage a map service that provides a street view feature by connecting real-world images within the virtual reality environment.

B. Digital Twins for Industrial field

Digital twin(DT) is one of the most promising key technologies for realizing smart manufacturing and Industry 4.0. DT is characterized by the smooth integration between the cyber and physical spaces. The significance of digital twins is increasingly being recognized by both industry and academia. It has been nearly 15 years since the concept of digital twins was first proposed[9]. DT has been extensively investigated in various industry sectors. It enables companies to create a virtual replica of their products throughout their entire lifecycles, facilitating the rapid detection of defects, early resolution of physical issues, more accurate outcome predictions, and improved customer service[10]. Numerous research studies have been focused on optimizing various aspects of product manufacturing processes and the lifecycle of these processes in the manufacturing sector through the use of DT. Jones et al [11] underscores the potential of DT to develop a computerized system that supervises each stage of manufacturing using a modular approach. Jones et al propose a modular Smart Manufacturing approach, where autonomous modules execute high-level tasks independently, make choices from a set of alternative actions, and respond to faults or unexpected events, all while avoiding the need for changes and reconfigurations at the supervisory level. To achieve this, modules must have access to precise information that accurately describes the current state of processes and products, which can be obtained by utilizing faithful virtual replicas of physical entities, i.e., DT. In the depicted scenario, DT also enables continuous communication between the system and physical assets[12].

III. METHODS

The whole scenario consists of: Simulate Street view route that user will go from starter location to get in 3D metaverse factory. Therefore, to simulate Street View we use map application API to display realistic street view for user experience. In this paper we use API from two map application.

1.Google Map

2.Naver Map

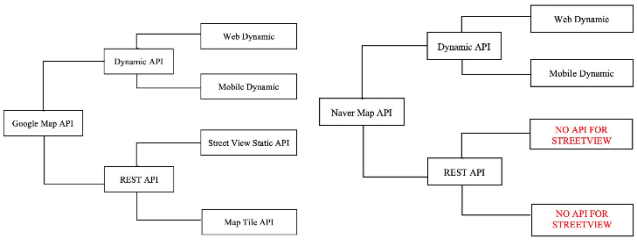


Fig.1 Type of Map API in Diagram

The reason we chose Google Maps is that it is widely used around the world, so it has a lot of information from all over the world, and also this paper is a part of the study about the environment in Digital Twin according to the place that is considered is in Korea, so Naver Map is a platform that was developed by a Korean company that is suitable to use in this paper to see updated in formation and Street View photo that is more specific and recently updated.

A. Exploring about Type of Map API

In this paper, which is partially dedicated to the use of street view we focus on two types of map APIs: Dynamic and Static. Both API implementations and send requests are distinct. For Dynamic API, the target platform is a web server or mobile service as it uses a dynamic loader to update data when the user interacts with a map in this kind of API and uses the HTTP or HTTPS protocol to send requests also the response is a dynamic page. On the other hand, Static API is part of the REST API, so it doesn't have a dynamic loader and sends requests using a URL format and the response is picture which makes it suitable for use in standalone applications. According to the scenario in our paper, the Google Map type of API which includes requests for street view has both Dynamic and Static APIs. Conversely, in the Naver map type of API that includes requests for street view, there is only a dynamic API. In this case, following the scenario for the Google Map API, we can use both types of APIs to request street view and in the Naver Map API, we can use only Dynamic API to request street view. For clarity, it can be viewed in diagram form as shown in Fig. 1

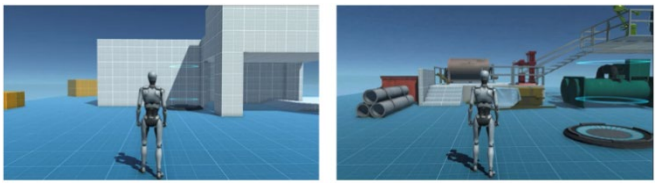


Fig2. Start stage(left) and factory stage(right)

Fig3. Select destination

Fig.4 Navigate and alert to destination

When we look in detail at the case of using Static API for street view in Google Maps, we can see that Static API has two sub-API that allow us to get a street view. When we examine the instance of utilizing Static API for street view in Google Maps in further depth, we can see that there are two sorts of Static API that allow us to obtain a street view. The first API is Street View Static. Map Tile API is the second. The difference between these two Static APIs is that when we request API in Map Tile, we can also retrieve information about that place such as the name of the road, address, and nearby location, which can be utilized in other parts. However, the Street View Static API only returns images of the street view.

B. Imporvement of UX and UI

The scenario described above has three stages as shown in Fig. 2, the initial metaverse start stage assumes the starter location. Second, for the street view stage, we use the map API. The third stage is the 3D metaverse factory, as seen in Fig. 2. To generate the first and third metaverse stages in our scenario we use Unity which is a Real-Time programming environment capable of creating 3D, VR, and AR. After creating a metaverse in Unity we must construct our metaverse stage on the target platform where we will display the simulation. As previously stated, the target platform that we select will influence the type of map API that we can utilize.

The Metaverse start stage scene includes a 3D building and a portal that avatar can teleport to street view that is the second stage which is an important part of this paper study. Based on the street view API from each map application in this paper we can conclude that in the case of using Google Map API, we can simulate our scenario in a web server, mobile service, and standalone application. While Naver Map we can use in web server or mobile service. Therefore, in this paper study, we determine to use Google Map API in a standalone application and Naver Map in WebGL which is a JavaScript API for rendering interactive 2D and 3D graphics within any compatible web browser that Unity supports to build into this when web server is a target platform. Moreover, we decided on some functions in the street view scene for the purpose of improving user experience in a digital twin environment to be more realistic as shown in Fig. 3

According to the example function in the street view shown in Fig 4 we designed it for convenience and to solve problems that the user encounters about the route to get to the destination factory so that the user does not get lost in the street view scene, the user can select the destination factory in map picture or drop down and when the user gets near junction the arrow will show up to navigate path from user place to the destination factory location. In addition, when the user gets to the destination, a text alert will notify them that they have arrived. The keyboard and mouse can be used to control the user's viewpoint and movement.

IV. RESULT

According to the above, we decided to use Google Map API in the case of Static API and Naver Map API in the case of Dynamic API. To see aspects of both map API types and both map applications to improve user experience in digital twin environments to be as effective as possible.

A. Google street view API

For Google Maps, we use Static API, so it has 2 sub-API as mentioned above. Therefore, we use both sub-APIs for street view in a standalone application to see the difference in terms of using with function in street view. So that, we can summarize the differences as follow.

1. Map Tile API: This API allows us to request meta information of location so we can apply information to our navigation function and apply it in the function select destination factory. Fig. 5
2. Street View Static API: This API does not allow to request of meta information on location so if we want more information on location, we need to use another API In tandem to combine information to create the function that we made. Fig. 6

So, it can be concluded that in case of only displaying street view pictures no need for other information it can use one of both APIs. On the other hand, in case of needed information on location, it's better to use Map Tile API.

B. Naver street view API

For Naver Map, we use dynamic API on WebGL. Regarding implementing the create functionality in Street View, it can be easily implemented from their module or class because Dynamic web API response in web page not photo so we can use more classes and modules from map application. Fig.7

Fig. 5 Google Map Tile API

Fig. 6 Google View Static API

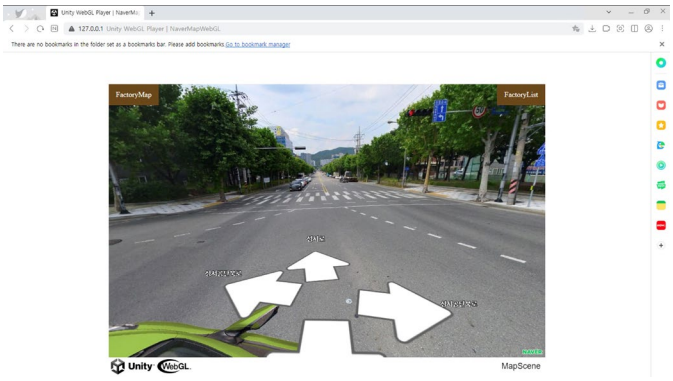


Fig.7 Naver street view API 1

TABLE I. SUMMARY BETWWEN GOOGLE MAP API AND NAVER MAP API

Street View API	Google Map		Naver Map	
	Dynamic API	Static API	Dynamic API	Static API
Web server	O	O	O	
Standalone Application		O		

From all the above in this section, we can summarize the overview as you can see in Table 1, that if given precedence to the kind of map application API it must choose a target platform for display simulation by the type of street API that have in each map application. But if given precedence to the type of map API it is possible to choose a target platform for display simulation, but it will limit the map application API that can choose.

V. CONCLUSION

In this study, we explore methods for enhancing user experience and increasing realism within the digital twin environment for the purpose of improving productivity in industrial settings and enterprise management. We achieved the integration of digital twin-real-world images by utilizing two currently available street view APIs. The development direction should be set based on the format of the provided APIs, and the resulting advantages and disadvantages were analyzed accordingly.

Through this research, we have proposed a novel approach to create a more realistic mirror world within the metaverse environment, particularly by projecting the real world into the virtual space. This contributes to the implementation of new strategies for enhancing the user experience. As part of future research, we plan to enhance the functionality of the program developed in this study and investigate additional elements for improving user satisfaction.

ACKNOWLEDGMENT

This work was supported by the Digital Innovation Hub project supervised by the Daegu Digital Innovation Promotion Agency(DIP) grant funded by the Korea government(MSIT and Daegu Metropolitan City) in 2023(No. DBSD1-04, Smart management system for responding to lonely deaths of elderly living alone based on remote meter reading information and CCTV access information)

REFERENCES

[1] Adedoyin, Olasile Babatunde, and Emrah Soykan. "Covid-19 pandemic and online learning: the challenges and opportunities." Interactive learning environments 31.2 (2023): 863-875.

[2] Wang, Yuyang, et al. "Re-shaping Post-COVID-19 teaching and learning: A blueprint of virtual-physical blended classrooms in the metaverse era." 2022 IEEE 42nd International Conference on Distributed Computing Systems Workshops (ICDCSW). IEEE, 2022.

[3] Mystakidis, Stylianos. "Metaverse." Encyclopedia 2.1 (2022): 486-497.

[4] Ning, Huansheng, et al. "A Survey on the Metaverse: The State-of-the-Art, Technologies, Applications, and Challenges." IEEE Internet of Things Journal (2023).

[5] Singh, Maulshree, et al. "Digital twin: Origin to future." Applied System Innovation 4.2 (2021): 36.

[6] Heonsik Joo, "A Study on Understanding of UI and UX, and Understanding of Design According to User Interface Change," International Journal of Applied Engineering Research ISSN 0973-4562 Volume 12, Number 20 (2017) pp. 9931-9935

[7] Hyunjung Oh, "A Study on UI • UX Design Change According to Trend (2016) "

[8] Seonhyeong Kim, "Study on domestic potal site applocation UX·UI design in smart phone Environment (2015)"

[9] Tao, Fei, et al. "Digital twin in industry: State-of-the-art." IEEE Transactions on industrial informatics 15.4 (2018): 2405-2415.

[10] Hu, Weifei, et al. "Digital twin: A state-of-the-art review of its enabling technologies, applications and challenges." Journal of Intelligent Manufacturing and Special Equipment 2.1 (2021): 1-34.

[11] Jones, David, et al. "Characterising the Digital Twin: A systematic literature review." CIRP journal of manufacturing science and technology 29 (2020): 36-52.

[12] Barricelli, Barbara Rita, Elena Casiraghi, and Daniela Fogli. "A survey on digital twin: Definitions, characteristics, applications, and design implications." IEEE access 7 (2019): 167653-167671.

PolyGoat: Interactive Manglish Learning Game

Wui Ling Chai
Department of Computing and Information Systems
Sunway University
Petaling Jaya, Malaysia
19121342@imail.sunway.edu.my

Yunli Lee
Department of Computing and Information Systems
Sunway University
Petaling Jaya, Malaysia
yunlil@sunway.edu.my

***Abstract*—Malaysia rich with multi-ethnic, multicultural and multilingual society that consists of three major races: Malay, Chinese and Indian. It has attracted many travelers across the globe to visit or work, either in short- or long-term stays. Due to multi-ethnic and multicultural, Malaysian community actively use Manglish – the informal form of Malaysian English embedded with Malaysian Slang. PolyGoat is an interactive Manglish learning game, it is designed to support learners, especially travelers to Malaysia using gamification learning concepts with passive and active learning activities. PolyGoat has four main modules: learning module, translator, slang dictionary and quiz. The learning module is designed based on frequently used vocabularies, and scenario-based dialog in six different setting of context (places). Slang dictionary is presented using flashcards approaches allows learners to familiarize the local Manglish slang. Lastly, the quiz module to evaluate the learner progress and cultivate effective communication between the locals and foreigners in Malaysia.**

Keywords—Interactive Learning Game, Gamification, Manglish, Slang Dictionary

I. INTRODUCTION

This interactive Manglish learning game evolves from the traditional language learning ways, other example of interactive learning games are Duolingo, Memrise, LingoDeer and more. The reason interactive language learning game evolves from traditional language learning is that interactive language learning games provide learners a flexible platform to complete their learning modules whenever they are free and have access to their mobile devices. Furthermore, interactive language learning game does not need the learners to pay or commit themselves to the classes like the traditional learning way.

This project developed an interactive language learning game to revolutionize the method of learning language. Due to the importance of achieving effective communication while travelling in a foreign country, the language proficiency has become one of the crucial keys in effective communication to break through the language barrier issue. This game aimed to fill in the gaps found from the traditional language learning methods by giving the learners flexibility and learn in a more attractive and interactive way. Not only that, but this game endeavor also enhances the daily communication between the locals and the foreigners in Malaysia. Therefore, there are the slang dictionary to allow learners to learn the local slangs to understand the locals well.

This game contains four main modules: Learning Module, Translator, Slang Dictionary, and Quiz. In the learning module consist of two features which are the vocabularies and scenario-based dialog. The vocabularies are displayed in flashcard to show English to Malay vocabularies with definitions. Furthermore, there will be six different scenarios: restaurants, pharmacy, directions, teatime, airport, and transportation, displayed in the dialog format to show the daily dialog in the specific place. The translator will translate the words insert by the user, which work similarly with Google Translate. The slangs in the slang dictionary are the 20 frequently used slangs in Malaysia for the learners to get a head start to communicate and understands the locals with slang.

A. Problem Statement

In the current society, the problem that frequently appear in our daily life is the lack of effective communication, especially when foreigners communicate with the locals when they travel to a foreign country. This problem have cause the disruption of allowing foreigners to discover more about the diversity of culture in Malaysia and causing day to day of socializing being limited. For example, service industries such as restaurants, airports, pharmacies, and many other places faced communication barriers as they could not understand each other and convey the message they wanted. Not only that, but the traditional way of language learning is also not as effective and efficient as compared to the interactive language learning way in this uprising technology and new paradigm of learning methods. Therefore, with the gamification's tools in this interactive language learning game, it gives out more positive environment for the learning effect to the learners and also allow the learning to adopt a better knowledge to learn a different language and to improve their communication skills when they travel to Malaysia.

B. Project Aim

This project aimed to design and develop an interactive language learning game that allows engagement between the learners and the learning activities, and to enhance their language skills by using the learning concepts that are gamify.

C. Project Objectives

The project objectives for this interactive language learning game are the followings below:

- To conduct literature review on the state of art of gamified tools for foreign languages in Malaysia context.
- To design a user interface, game engine, modalities and database architecture structure for the interactive language learning game.
- To develop the frontend and backend structure for the interactive language learning game.

II. Literature Review

A. Introduction

In the service industry, it is a crucial role to achieve effective communication as it connects the cultural and linguistic gaps between the foreigners and locals. Communication is a daily thing we do in life; hence it is inevitable in the service industry when the workers have to communicate with the customer to convey what they want and need.

This literature review is conducted to study the efficacy of interactive language learning game compared to the traditional language learning method. Not only that, but this study will discover the basic requirements in communication at public places to know the way to fix communication barrier between locals and foreigners by using the interactive language learning game.

B. Active and Passive Learning

Active learning involves engaging learners in the learning process through activities such as speaking, writing, and asking questions. This approach allows learners to practice and apply their skills actively, leading to improved verbal language, pronunciation, and immediate feedback. Various activities, like writing sentences or participating in conversations, contribute to effective learning. Additionally, tools like flashcards, storyboards, and real-life situations enhance motivation beyond traditional reading and listening [1]. On the other hand, passive learning occurs at the initial stage, focusing on receiving input from learning modules. Activities like listening and reading help learners grasp the basics of language, pronunciation, and vocabulary. Comprehensive skills are developed through reading, enhancing the understanding of context in the materials [1].

In conclusion, both active and passive learning approach are important in the process of learning, where the learners can be actively engaged and passively learn the learning materials. Hence, it is the best option to combine the active and passive learning approach in the interactive language learning game to enhance the acquisition, and proficiency of the language from the learners.

C. Slang in Malaysia

Slang is a very informal language that is used in particular group of people. In Malaysia, Manglish which is knowns as Malaysian English is widely used with the mixture of languages with the combination of English, Malay, Chinese, and Tamil. For example, “dabao” is a slang that means take away or take out that is frequently used in food and beverages places [2]. Slangs have become one of the important parts of the culture in Malaysia. Therefore, slangs are being added into the interactive language learning game as one of the main modules for the learners to learn the culture in Malaysia and to have interesting conversation with the locals in Malaysia. Slangs will also help to improve the communication barrier and to achieve effective communication between the locals and foreigners in Malaysia.

D. Communication Barrier

According to Ne’Mattulah et al. [3], the communication barrier between the foreigner workers and the supervisors in Malaysia can greatly affect the progress of the project causing the project to ask for extension of time (EOT) and causing the project to delay and failure. Furthermore, risk of danger towards worker’s safety is high if there is a miscommunication due to communication barrier happen on the construction site. Therefore, Ne’Mattulah et al. [3] have concluded that it is important to have language training as it could avoid any communication challenges on site.

Furthermore, as shown by Vimala & Omar [4], workers and patients that goes to the hospitals in Malaysia faced language barrier, as Malaysia is a country with various ethnicities that have their own culture, language, and dialect. Therefore, lack of interactions, communication and misunderstanding happens a lot daily in a hospital causing medical error which causes distress among patients. Medical officers face issue during communicating with the patient which leads to unclear medical history and wrong prescription. Hence, the solution to ease the situation is according to Valitherm [5], is to have language lessons and use a translator device to ease the communication issues.

E. Interactive Language Learning Game

There are many different types of interactive learning ways, such as learner-content interactions, learner-instructor interaction, learner-learner interactions, and learner-community-of practice interaction [6]. Learner-content interaction is the type of interaction learning ways that is used in this interactive language learning game, it is a learning interaction between the learners and content such as activities, modules, and practices.

As shown by Omar et al. [7], interactive language learning activities has the highest mean on increasing the confidence of the learners in using English language to speak. The factors that contribute to the highest mean are interactive language learning provides a comfortable environment to the learners for them to learn how to read, write and speak the new language. Not only that, Marzuki et al. [8], have stated that if the learners continue participating in all the interactive language learning activities, the learner’ speaking proficiency will gradually improve as they have slowly built up their self-confidence level. Furthermore, interactive language learning also allows learners to learn and apply the new vocabularies that they have learnt throughout all the interactive language learning activities. Zero judgement environment is also one of the factors that encourage and motivate the learners to speak and use the new language that they have learned stated by Omar et al. [7].

F. Interactive Language Learning vs Traditional Language Learning

Baker [9] stated that active learning theory allows the students to more actively participated in the learning activities, and students will obtain more motivation to learn more knowledge. In the study from James & Mayer [10], there are two types of participants, one group will study a new language using the language app called Duolingo, and the other group will study using slideshows. The study has concluded that the group that uses Duolingo are more active and gained more experience and motivations compared to the ones that uses slideshows. Besides, the participants said that slideshows are less interesting and motivated, causing it to be difficult to continue the lessons. Therefore, the study has reported that using learning app such as Duolingo gives a more effective effect in learning as it is easier to grasp the knowledge and it gives out more exciting and motivating contents compared to the slideshows.

According to Birova [11], learning games that consist of background music, attractive and bright images, playful words and repetition of phrases that are changed according to the scenarios gives out more anticipation to the learners compared to the traditional way of learning. Therefore, some of the interactive language learning ways can replace the traditional language learning ways to enhance the learner's experience and increase the efficiency of learning a new language.

G. Summary

In conclusion, the findings in this literature review have concluded that the communication barrier in Malaysia between the locals and the travelers/foreigners, the importance of having effective communication, what is interactive language learning game, active and passive learning, slangs in Malaysia, and the comparison between interactive language learning and traditional language learning. This literature review has also discussed that to achieve interactive language learning game, active and passive of learning elements will be added in the game to provide a better learning experience in game-based learning activities to the learners. Besides that, the literature review has concluded that interactive language learning is the new method to learn a new language because it gives more motivations and less stress environment for the learner. Furthermore, the slang datasets are based on the top 20 frequently used slangs by Malaysian to be added in the slang dictionary to not let foreigners to have culture shock or having the sense of unbelonging in Malaysia.

III. METHODOLOGY

A. Stakeholders

The stakeholders in this project are the locals in Malaysia and the travelers or foreigners in Malaysia who wants to learn about the slangs in Malaysia.

B. Software Architecture

Modified waterfall process model is chosen as the software architecture for this project because the phases suit the structure of this project. There are seven phases in the modified waterfall process model, which are the business concept, requirements, general system design, detailed system design, development, test, and deployment as shown in Figure 1. The steps of the development phases are orderly sequence in the modified waterfall model. Not only that, but the modified waterfall model also contains the iterative stages that allows the model to be flexible. The quality and reliability of the software will be ensured as the iterative stages have to be completed, reviewed and have feedback on the documentations and design of the project. This model adopts a good methodology principle by reviewing the content in-between every stage as a checkpoint or quality gate before continuing to the following stage. Hence, this model is more suitable to be applied to this project as it gives clear requirements, objectives, aims, features, and target audience, as it is defined in the beginning phase of the project. Furthermore, there is a testing phase before deployment because the bugs and errors can be identified in the early stage of the development before the project being deployed.

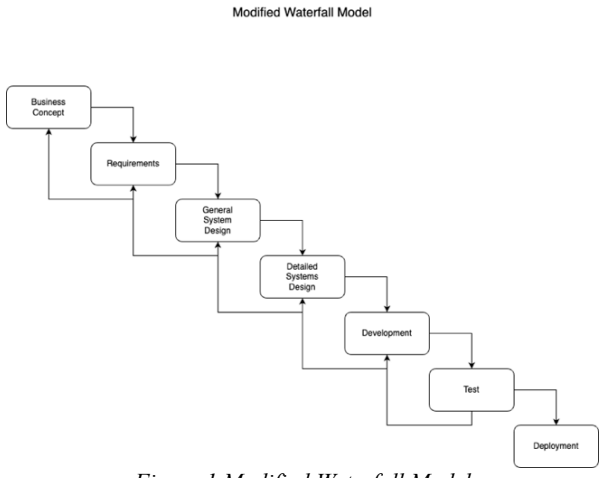


Figure 1 Modified Waterfall Model

C. Database Architecture Structure

Object-Relational Database Management System (ORDBMS) is applied in this project. ORDBMS allows the data to be organized and managed in a structured manner, where it is used in this interactive language learning game. Using ORDBMS allows the data management for various data type to be more efficient in this project. The data structure is able to store, and retrieve based on the table, relationships and constraints. Furthermore, ORDBMS is more flexible and extensible, as it gives out the flexibility on defining the data types and complex data structure. The extensibility also includes in providing the design for an extensible data structure for the language learning game which need to contain a large data of user data and learning module data. Last but not least, ORDBMS is scalable, hence giving out a good performance from handling large number of data and retrieving data with fast query responses.

D. Web Framework

- *Translation Library*
  Googletrans, which is a python library that is implemented by Google Translate API, where the methods such as detect and translate uses the API called Google Translate Ajax API. Therefore, Googletrans is used in this project for the translator function because it has the auto language detection function as it uses the

same servers that google translate uses, hence it is fast and reliable. Unlikely other approaches like goslate, Google has updated their translation service by using a ticket mechanism to allow the library to work properly and to prevent crawler program. 106 languages are supported in the Googletrans database for users to translate, such as Arabic, Chinese, Filipino, German, and more.

- *Slang Dataset*
  The slang dataset used in this project are the 20 most used slangs in Malaysia according to a few travelling websites, such as Traveloka, Time Out, and The Smart Local. For example, the commonly used Malaysian slangs are "Bojio", "Tapau", "Yum Cha", "Cincai", and more.

E. *Tools and Techniques*

- *Frontend*
  HTML, CSS, and JavaScript is used for building the frontend of this project including the styling and the functions.
- *Backend*
  Django is used for building the backend of this project to render the database with the server.
- *Database*
  SQLite is the database used in this project to be connected easily with Django.
- *Visual Code Studio*
  Visual Code Studio known as VS Code, is used as the integrated development environment (IDE) to edit the sources code.
- *Figma*
  Figma is a collaborative web application used in this project to design the prototype for the user interface.
- *Visual Paradigm*
  Visual Paradigm is used in this project to create modelling and diagrams for the flow diagram and database architecture structure for this project.

IV. POLYGOAT APPLICATION

A. *Polygoat Interface*

There are four main modules in Polygoat, which are Leraning Modules, Translator, Slang Dictionary and Quiz.

The Learning Module consiste two types of learning method, one is vocabulary, and another one is scenario-based dialog. The vocabularies are displayed in flashcards where learners can click on the flashcards to read the meaning of the word. The scenario-based dialog consists of six different places which are pharmacy, airport, directions, transportation, restaurant, and teatime. Each of these scenarios will display a dialog that are frequently ask in the specific place.

The Translator will be working almost similar like Google Translate, but the translator function will save the learner's translation history, so that the translated words can be referred back in the future.

The Slang Dictionary displays twenty most used slangs in Malaysia using flashcard, where it works similarily with the vocabulary. Learners will have to click on the flashcards to read the definition of the slang.

The Quiz is to test the understanding of the learners, where the scope of the quiz will covers the vocabularies and scenario based dialog. The learners will be asked to mix and match the same meaning of words in different language and select the correct answer from the multiple choice question based on the scenario given. The learners can re-attempt the quiz until they have correctly selected all the answers, as there will be no marking system in this quiz.

V. SURVEY RESULTS

A. *Requirement Survey*

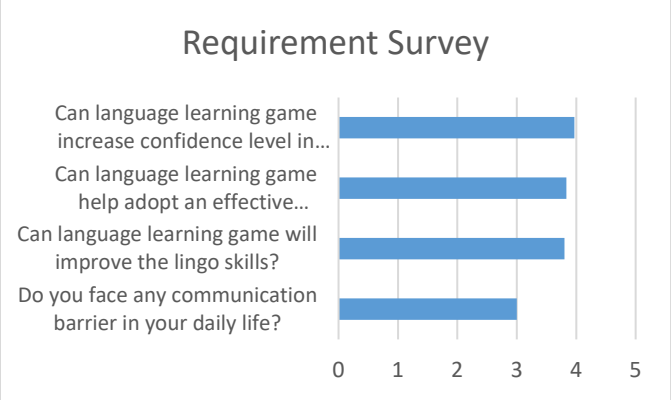


Figure 2 Requirement Survey

1) *Survey Analysis*

Based on the 30 results of the survey, the number of participants has faced communication barrier in their daily life is high where the nationalities of the participants that faced communication barrier are 70% majority students from age 18 to 24 years old.

Most of the participants have experienced using language learning applications such as Duolingo, Memrise. Besides that, the majority of participants feel that using language learning game application can improve enhance the lingo skills. Not only that, but the responses of the participants show that by playing a language learning game will likely adopt an effective communication and increase the confidence level during daily communication.

However, most participants think that after playing the language learning game will increase confidence level as practice makes perfect, where the game will allow the learners to practice all the activities with unlimited attempts. Some participants have given comments on having immediate feedback on the mistakes when they are doing the

activities can also boost their confidence level because of the non-judgmental environment that is provided in the game.

There are a few functions and features collected from the participants responses, where they have suggested to have a real-life scenario dialog in the game to allow the learners to practice and experienced the examples of conversations in real-life. Other than that, scoring and rewarding features are also recommended by the participants to be added in the game to make the game more exciting and interesting. Last but not least, the participants have also suggested to have music and sound effects to let the game more interesting.

2) Summary of the survey

In conclusion, the features that is added in the game to achieve the objective on developing an interactive language learning game are scenario-based dialog activities and flashcards learning method. The other features such as milestones features, and sound effects will be added in the future work of the game.

B. User Experience Questionaire (UEQ)

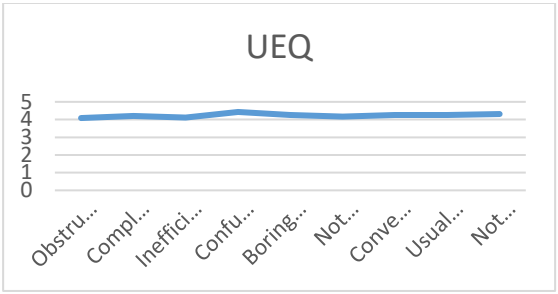


Figure 3 UEQ

1) Survey Analysis

Based on the 35 participants in the UEQ, majority of the participants have given a score on the game being supportive. The ease of use of the game have also been proven by the score for the complicated vs easy survey question. Majority of the participants have also voted for the game being efficient. The game has also been clear and not confusing to use according to the majority of the participants. The participants have also felt that the interactive language learning game is exciting and interesting to play and to learn a new language. Furthermore, the participants feels that this interactive language learning game is inventive and leading edge as there are no language game that provides slangs dictionary. Last but not least, the majority of the participants feels that this language learning game is interactive after experiencing the game.

2) Summary of the survey

In conclusion, the participants have voted that this interactive language learning game is easy to use, efficient in learning, instructions are clear, activities are exciting, and inventive. On the other hand, the game has also received a score on the game being interesting. Lastly, the game can improve on being more supportive to enhance the user experience.

VI. CONCLUSION

In conclusion, this project has achieved to develop an interactive language learning game, consists of four main modules which are Learning Module, Translator, Slang Dictionary, and Quiz.

The future improvement that could be added in this interactive language learning game are adding more learning modules in the game such as more language, learning materials and content. Furthermore, milestones functions can also be added as a motivation and rewarding to the learners to allow them to keep track on their learning progress. Sound, music and visualization effects can also be added into the game to make the game more interesting and exciting for the learners to keep learning.

REFERENCES

[1] Cchiaro, "Active vs. passive learning: What's the difference?," Graduate Programs for Educators, https://www.graduateprogram.org/2021/06/active-vs-passive-learning-whats-the-difference/#:~:text=Active%20learning%20requires%20students%20to,and%20paying%20attention%20to%20detail. (accessed Jun. 20, 2023).

[2] Carissang, "Singaporean office lingo: 21 of the best phrases to know," Singapore, https://employmenthero.com/sg/blog/office-lingo-best-phrases/#:~:text='Dabao'%20essentially%20means%20to%20take,eat%20out%20or%20'dabao'. (accessed Jun. 20, 2023).

[3] K. F. Ne'Matullah, L. S. Pek, and S. A. Roslan, "Investigating communicative barriers on construction industry productivity in Malaysia: An overview," *International Journal of Evaluation and Research in Education (IJERE)*, vol. 10, no. 2, p. 476, 2021. doi:10.11591/ijere.v10i2.21163

[4] G. Vimala and S. Z. Omar, "Interpersonal communication skill barrier faced by cardiology doctors at National Heart Centre malaysia," *International Journal of Academic Research in Business and Social Sciences*, vol. 6, no. 6, 2016. doi:10.6007/ijarbss/v6-i6/2202

[5] A. Valitherm, Communication barrier in Malaysia construction sites - ijern.com, https://www.ijern.com/journal/January-2014/08.pdf (accessed Jul. 23, 2023).

[6] "Types of interaction," Online Learning Design Studio, https://clt.manoa.hawaii.edu/projects/online-learning-design-studio/interaction/closer-look-interaction/ (accessed Jun. 20, 2023).

[7] S. F. Omar *et al.*, "Interactive language learning activities for learners' communicative ability," *International Journal of Evaluation and Research in Education (IJERE)*, vol. 9, no. 4, p. 1010, 2020. doi:10.11591/ijere.v9i4.20605

[8] M. Marzuki, J. A. Prayogo, and A. Wahyudi, "Improving the EFL Learners' speaking ability through interactive storytelling," *Dinamika Ilmu*, pp. 15–34, 2016. doi:10.21093/di.v16i1.307

[9] N. A. Baker, "Glued to games: how video games draw us in and hold us spellbound," Choice, vol. 49, (1), pp. 103, 2011. Available: http://ezproxy.sunway.edu.my/login?url=https://www.proquest.com/trade-journals/glued-games-how-video-draw-us-hold-spellbound/docview/889943015/se-2

[10] K. K. James and R. E. Mayer, "Learning a second language by playing a game," *Applied Cognitive Psychology*, vol. 33, no. 4, pp. 669–674, 2018. doi:10.1002/acp.3492

[11] I. L. Birova, "Game as a main strategy in language education," *American Journal of Educational Research*, vol. 1, no. 1, pp. 6–10, 2013. doi:10.12691/education-1-1-2

Design of a Mobile Application Using Two Numbers for eSIM-based Mobile CRM

Sang-Soo Park
Office of Business Planning, NANOIT Ltd. Inc.,
Daegu, Rep. of Korea
dori@nanoit.kr

Yong-Jun Jon
Dept. of AI·Big Data Engineering, Daegu Catholic University
Gyeongsan, Rep. of Korea
braveheart@cu.ac.kr

Bonghwan Kim
School of Electronic and Electrical Engineering, Daegu Catholic University
Gyeongsan, Rep. of Korea
bhkim@cu.ac.kr

Tae-Young Byun*
School of Coumputer Software, Daegu Catholic University
Gyeongsan, Rep. of Korea
tybyun@cu.ac.kr

***Abstract*— This paper describes the design of a mobile app that allows a mobile device with two phone numbers to separate a work phone from a personal phone using a SIM or eSIM. This has the advantage of allowing users with two phone numbers to balance their life and work, thus protecting their privacy.**

Keywords— SIM, eSIM, customer relationship mManagement, CRM, two number service

I. INTRODUCTION

In recent years many devices such as smartphones, smart watches and laptops have started being equipped with embedded SIM(Subscriber Identification Module) cards called eSIM, which have the possibility to be reprogrammed with new subscription data through a process called remote SIM provisioning. To enable the eSIM in these devices, the mobile operators have had to implement new systems for the remote SIM provisioning process, and most of the research around this process has focused on security, leaving other areas unexplored [1]. In contrast to regular SIM cards where a change of carrier can take some time as you have to get a new SIM card to physically change it in your device, a change of carrier with the use of eSIM can be almost instant. The eSIM profile, which is the collective name for all the data and applications that would normally have been put on the physical SIM card, is simply downloaded and installed on the device over the network through a process called remote SIM provisioning.

Customer relationship management (CRM) is a technology for managing all your company's relationships and interactions with customers and potential customers. A CRM system helps companies stay connected to customers, streamline processes, and improve profitability. Many mobile SW development companies have been working hard to develop mobile device-based CRM solutions for efficient customer management. Therefore, for smartphone users with dual numbers using e-SIM, the implementation of a function that separates personal and business areas for efficient data management is an important development issue. The principles of building and strengthening a CRM strategy that allows a company to build and strengthen relationships with customers from whom it can expect continuous sales have been presented in a number of previous article [2].

This paper proposes details for the design of a phone and messaging app that can handle receiving and sending SMS, MMS, and making and receiving calls using two numbers using two physical SIMs or E-SIMs. By doing so, users can separate their two portable phone numbers into personal and work areas, which can help them achieve a balance between work and life.

II. RELATED STUDIES

A. SIM vs. eSIM

The SIM card has changed considerably over the years, both physically from the 1FF (Form Factor) card to the 4FF, or more commonly known as the nano SIM card, and in its capability, such as connectivity speed, storage capacity and security. The next step for the SIM card is the eSIM, a SIM card embedded in the device. This evolution has been studied for example by Vahidian in [3] and by Koshy and Rao in [4]. This section gives an introduction on the differences between normal SIM cards and eSIMs [1].

- SIM: The SIM (Subscriber Identity Module) or the SIM card as it is most commonly known as is a removable smart card, called UICC (Universal Integrated Circuit Card), used to authenticate the user to a mobile network in order to gain access to it. It is made up of an integrated circuit and is where the IMSI (International Mobile Subscriber Identity), which is the number that is used to identify the card to the network, and the authentication key is securely stored. The UICC contains the SIM and USIM (Universal Subscriber Identity Module) applications that are used for authentication in GSM (Global System for Mobile Communications) and UMTS (Universal Mobile Telecommunications System) networks respectively. The UICC is also tamper resistant and has security features built into it to help prevent unauthorised access to the IMSI and authentication key so that they cannot be copied or modified [3]. The SIM card has the benefit of being easy to use and can easily be swapped between devices, but there are also some drawbacks. It takes up quite a lot of space, especially if is to be used in for example smart watches or IoT devices

* Corresponding Author
The Ministry of Trade, Industry & ENERGY(MOTKE)

of limited size. It can also be a problem for the manufacturers of devices that use SIM cards as it puts constraints on things like design and waterproofing of the device. One of the biggest drawbacks though with regular SIM cards is the actual need for the physical card at all as this implies that it has to be acquired somehow, which may take time and makes it more difficult to change network operator.

- eSIM: eSIM (embedded SIM), also called eUICC (embedded UICC) is a UICC that is soldered directly into a device during the manufacturing process. This allows the separation of the operators' profile and the physical chipset. The device can be manufactured with or without an initial profile and new profiles can be downloaded in a later stage from an SM-DP server through a process called remote SIM provisioning (RSP), where the profiles contain the same data that normally would have been put in a regular SIM card [6].
- Devices Supporting eSIM: A list of mobile phones that have eSIM is given in table 1.1 [6][7].

TABLE I. MOBILE PHONES CURRENTLY SUPPORTING ESIM

Mobile phone manufacturer	Model
Samsung	Galaxy S20, S20+, S20 Ultra Galaxy S21, S21+, S21 Ultra Galaxy S22, S22+, S22 Ultra Galaxy S23, S23+, S23 Ultra Galaxy Fold Galaxy Z Flip etc.
Apple	iPhone 11, 11 Pro, 11 Pro Max iPhone XS, XS Max iPhone XR iPhone SE 2020 etc.
Google	Pixel 2, 2 XL (Google Fi only) Google Pixel 3a, 3a XL Pixel 3, 3 XL Pixel 4, 4 XL
Motorola	Moto RAZR
Huawei	P40, P40 Pro, P40 Pro+

III. DESIGN OF TWO NUMBER SERVICE FOR ESIM-BASED MOBILE PHONE

A. Design Background

Nowadays, individuals emphasize work-life balance and the social atmosphere is shaping up, and the need for mobile services that meet these trends is increasing. In addition, eSIM service has been available in South Korea since September 2022, and devices that can use eSIM are being released continuously, so it is necessary to preempt the market by developing convenient services using eSIM.

Apps for eSIMs provided by major smartphone manufacturers, such as Samsung Electronics, show recent history, texts, and incoming calls, but it is inconvenient to intuitively recognize which SIM the call is coming from, which SIM the call is with, and the text history. Therefore, when making a call or sending a text, the user may be confused about which SIM to call or send to, and there is a high probability of making a mistake that results in an unwanted call to the other party's phone number. Currently, Samsung's Galaxy smartphones indicate which SIM is being used with a 1 and 2 numbered icon, which is not intuitive.

In an environment where two phone numbers are used on one smartphone, all incoming and outgoing call records, SMS and MMS sending and receiving records, regardless of whether they are for business or personal use, are mixed and managed. If a user wants to use one phone number for business and another for personal use, existing data management methods can cause a lot of confusion and inconvenience to the user. Therefore, it would be convenient for users to display various historical information for each phone number on a separate UI screen.

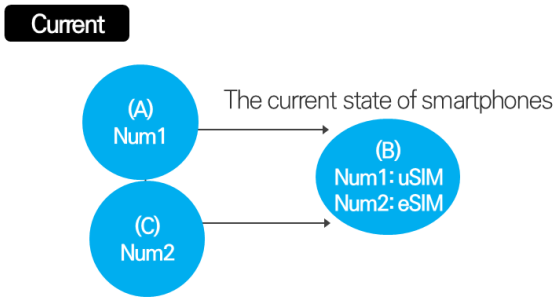
B. Design Features

In this paper, we propose the following design to provide users with convenient data management functions when using two phone numbers using SIM and eSIM (Fig. 1).

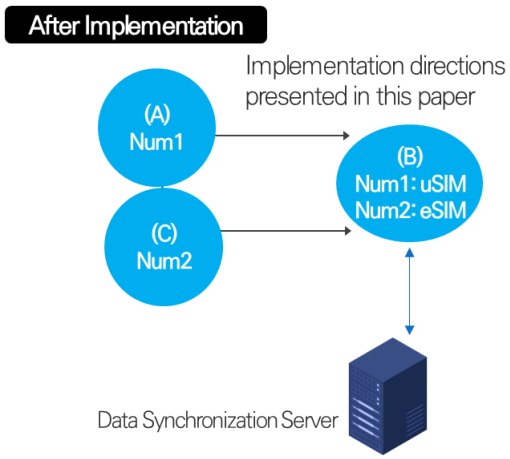
First, for each phone number, you'll have separate business and personal zones. For Business, our design provides business history with Num1-uSIM. Also, for personal use, Num2-eSIM provides personal history on separated UI screen.

Second, through the design of a dedicated Android app, we provide independent information by work and personal areas to improve the efficiency of data management.

Finally, in order to provide eSIM-based two-number service, we provide data synchronization function between smartphone and data server to improve data reliability.



(a) Current using of two numbers on a smartphone

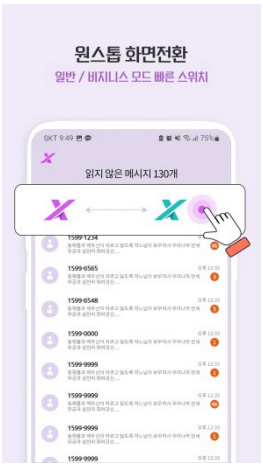


(b) Improved data management with data synchronized server

Fig. 1. Proposed development directions

The following describes five design principles for feature design (Fig. 2).

- Design a UI with one-stop screen transitions for quick switching between personal and business modes.
- Design the UI to change colors based on whether the user is using a personal or business phone number, so that users can intuitively recognize the difference between personal and business modes.
- Allow users to save text or voice notes after a call with the other party, and add hashtag (#) notes to enhance search capabilities.
- To enhance the CRM function, history management function is provided for each customer, enabling detailed customer management.
- It is designed to protect personal privacy and increase security by silencing certain phone numbers or SIMs (eSIMs) or preventing users from exposing unwanted data (text, SMS, MMS, etc.).



(a) Fast screen switching

(b) Clear separation of life and work

(c) Create notes and hashtags

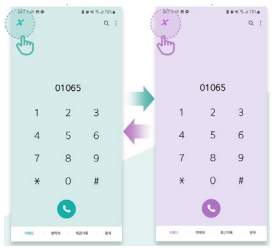
(d) Enhance CRM functionality

(e) Privacy and security

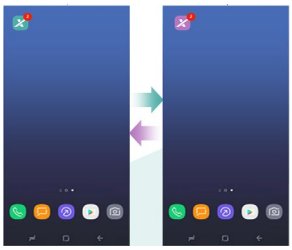
Fig. 2. Illustration of five design features

C. *UI Design*

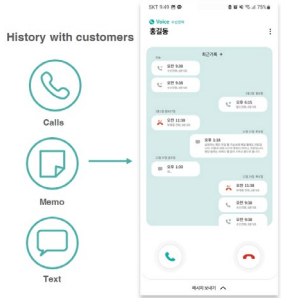
In this section, we describe how to design a UI that's user-friendly according to the five design principles described earlier. Fig. 2 shows some examples of UI design screens based on the five UI design principles. Fig. 3(a) shows the life mode and work mode when making a call, you can change the mode quickly by clicking the X icon in the upper left corner of the app screen. Each mode has a different background color for easy intuitive recognition. Fig. 3(b) shows the number of received events on the X icon for easy notification of received texts or calls on the smartphone wallpaper in each of the different modes. Through the UI in Fig. 3(c), you can see a quick history of all customer interactions and previously made notes to strengthen the delicate relationship with customers. Finally, the UI in Fig. 3(d) lists all calls or texts with the customer for quick and intuitive review. You can also set up text notes, voice notes, and hashtags immediately after the call to record and save the history of your interactions with the customer if needed.



(a) Life mode(mint, left) and work mode(purple, right)



(b) Life mode(left) and work mode(right) on wallpaper



(c) CRM capabilities with history features

(d) Intuitive note and text history management

Fig. 3. Examples of UI design

IV. Conclusions

For users with two phone numbers, utilizing a mobile app that allows them to keep their life and work separate is beneficial in many ways. Employees get easy access to business materials, and by keeping their life and work separate, they can achieve a better life/work balance and improve their quality of life. For businesses, it can improve efficiency by providing the ability to manage history with customers, as well as provide protection and data management for sales and business information.

This paper has presented some design principles and design practices for designing mobile apps that can provide these benefits, and we hope that you can use them to develop mobile apps with similar features.

Acknowledgment

This work is financially supported by the Ministry of Trade, Industry & ENERGY(MOTKE) through the fostering project of project for industry-university convergence district build-up.

REFERENCES

[1] Albin Fridh, "eSIM Re-Selling on Mobile App," **Lund University Publications Student Papers**, Department of Electrical and Information Technology Lund University, 2020

[2] A. Idzikowski, K. Piotr, J. Cyganiuk, M. Ryczko, "Customer Relationship Management (CRM) - Philosophy and its Significance for the Enterprise," System Safety Human - Technical Facility - Environment, Vol. 1, Issue 1, March 2019

[3] E. Vahidian, "Evolution of the SIM to eSIM," Norwegian University of Science and Technology, Department of Telematics, 2013

[4] D. G. Koshy, S. N. Rao, "Evolution of SIM Cards – What's Next?, International Conference on Advances in Computing, Communications and Informatics(ICACCI), Bangalore, pp. 1963-1967, 2018

[5] B. A. Abdou, "Commercializing eSIM for Network Operators," IEEE 5th World Forum on Internet of Things (WF-IoT), pp. 616-621, 2019

[6] GSMA, RSP Architecture Version 2.2, 2017.

[7] T. Chang, "eSIMS: Everything You Need to Know," WhistleOut, Sept. 2019

Two Number X: Mobile Application Using eSIM for Life and Work Balance

Kwan-Su Kim
R&D Center
NANOIT Ltd. Inc.
Daegu, Rep. of Korea
rhkstn0303@nanoit.kr

Sang-Soo Park
Office of Business Planning,
NANOIT Ltd. Inc.,
Daegu, Rep. of Korea
dori@nanoit.kr

Bonghwan Kim
School of Electronic and
Electrical Engineering,
Daegu Catholic University
Gyeongsan, Rep. of Korea
bhkim@cu.ac.kr

Tae-Young Byun*
School of Coumputer Software,
Daegu Catholic University
Gyeongsan, Rep. of Korea
tybyun@cu.ac.kr

***Abstract*— This paper describes the implementation details of 'Two Number X', a mobile app that allows users to efficiently use two phone numbers for personal and business purposes according to their preferences. The Two Number X provides an intuitive UI (user interface) and data management function that allows users to use two phone numbers for different purposes and manage text messages and phone calls separately for each purpose. We described the implementation of a Two Number X app that can independently provide a mobile phone usage environment for personal and business use in order to promote the value of balancing the user's life and work when using two phone numbers using eSIM.**

Keywords— dual SIM, eSIM, two number service, life and work balance

I. INTRODUCTION

Smartphones or mobile phones have revolutionized the way people communicate with each other. They have powered a lot of people and made a lot of things possible due to their various functions and performance. Dual SIM phones have been around for about 15 years, but with the widespread adoption of smartphones, there has been a recent increase in interest in dual SIM or two-number services and the number of people using them [1]. As smartphones and services became more affordable, their users have tended to use different mobile subscriptions (i.e., SIM cards) for travel, business, and personal needs. For example, the total number of SIM connections worldwide has exceeded the global population while the number of unique subscriptions make only 60% of the population. By that time, mobile devices that accommodate more than one SIM card, also known as dual-SIM devices, have gained popularity, especially in countries with uneven coverage and in-country roaming.

With the advent of embedded SIM (eSIM), managing multiple mobile subscriptions within a single device has become more agile and user-friendly. The eSIM technology allows customers to install many eSIM profiles and select between the subscriptions at will via a software menu, although only one profile can be active at any given time. The introduction of eSIM facilitates the global adoption of multi-SIM devices. The technology has been adopted by all leading smartphone manufacturers, including Apple, Samsung, and Huawei. Since the addition and change of Mobile Network Operator (MNO)s became easier and faster, mobile operators expect substantial revenue from providing eSIM and improving dual-SIM user's experience.

With multiple subscriptions, MT services, such as voice calls, Short Message Service (SMS), and emergency notifications from different networks risk to overlap and fail to reach the user. Therefore, service reception at multi-SIM devices may require specific software and protocol enhancement on User Equipment (UE) and network sides[2].

In December 2021, the Ministry of Science and ICT of South Korea announced the 'Smartphone Secondary SIM (eSIM) Introduction Plan', and as a result, the smartphone secondary SIM (eSIM) service was implemented in South Korea on September 1, 2022. The main contents of MSIT's 'Smartphone Secondary SIM (eSIM) Introduction Plan' are system improvement for eSIM acceptance, SIM service and terminal usage environment environment for SIM services and terminals, prevention of unauthorized use of terminals, and research and development of domestic eSIM technology[3][4].

There have also been mobile apps for two-number services that allow you to use two phone numbers on one smartphone. Mobile apps such as Google Voice, Textfree, Dingtone, textMe, Flyp, Hushed, and Sideline are among the most popular. However, while these mobile apps have the advantage of giving you a free phone number, they don't offer the ability to use two different phone numbers for personal or business use while keeping the two different areas completely separate.

Therefore, this paper describes the implementation details of 'Two Number X', a mobile app that allows users to efficiently use two phone numbers for personal and business purposes according to their preferences. The Two Number X provides an intuitive UI (user interface) and data management function that allows users to use two phone numbers for different purposes and manage text messages and phone calls separately for each purpose.

II. RELATED STUDIES

A. Two number Service vs. Dual SIM

South Korea launched an eSIM service that allows users to use two phone numbers on a single smartphone last year. One of the main features of the eSIM is '1 phone, 2 numbers' using

* Corresponding Author
The Ministry of Trade, Industry & ENERGY(MOTKE)

dual SIM, but there is a difference in the service from the two number service of the three mobile operators in Korea that already existed. When you sign up for a two-number service, your carrier assigns you a random "virtual number" (the last four digits are yours to choose) and you have two phone numbers, as the name implies. However, having two numbers doesn't mean that you can completely separate them, like getting a new phone, and there are a number of limitations. For example, since it is a virtual number, you cannot verify your identity separately. Even if you have two numbers, it is difficult to use applications that require separate authentication completely separately from your existing number.

Two-number services also differ depending on which operating system (OS) your phone is running - Android or Apple's iOS. Social media apps such as KakaoTalk in South Korea can be used separately on Android phones through the 'Dual Messenger' feature, but not on iPhones. In addition, since the two-number service is an additional service provided by the carrier, if you switch carriers, the service will be interrupted and you will have to sign up for a new one.

Dual SIM, on the other hand, is not a virtual number, but two real numbers on one cell phone. Since the line itself is opened separately, unlike the two-number service, you can choose your favorite carrier without being restricted by the carrier. Since it is a real number, both numbers can be used for identity verification. You can easily separate SNS and messenger apps even if you don't use the dual messenger function, and you don't need to pre-enter a specific number before calling or texting like a two-number service. Since the lines of the USIM and eSIM are completely separate, you can also choose your plan freely. Since the main purpose of using two phone numbers is for business purposes, you can use one number for a higher-priced plan with a lot of data, and the other for a lower-priced plan that is only used for receiving calls. Dual SIMs are more versatile and convenient than two-number services, but they also cost more.

B. Status of e-SIM Commercialization Services in South Korea

a) Table 1 summarizes similar services offered by major mobile operators in South Korea. These services are provided by the three major carriers, and unlike dual SIMs with SIMs, they provide a virtual number from the carrier and provide a service similar to call forwarding.

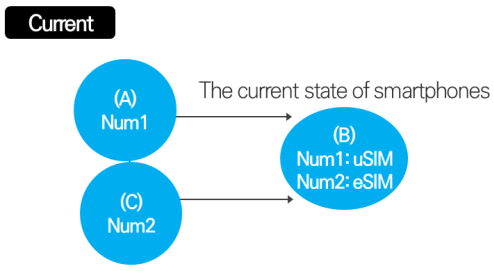
TABLE I. TABLE TYPE STYLES

Carrier	Service Brand Name	Feature
Korea Telecom	KT Two Phone	· Using virtual number · Change the system profile to change the 1st and 2nd numbers · Not available on eSIM. · Cannot receive notifications from different phone numbers · Financial verification of virtual number is not possible
SK Telecom	SKT Number Plus 2	· Use a virtual number · Use a dedicated app or dial '*281 + other party's number' to make a call using a virtual number · If you don't have a dedicated app, rely on your main contact list · You can't receive notifications from different phone numbers · You can't verify your identity or use other services with your virtual number.
LG U+	Dual Number	· Use a virtual number. · Dial the other party's number + '#' and use the virtual number to make the call. · International calling not available · Using a virtual number while roaming internationally may expose your original number · Cannot split contacts

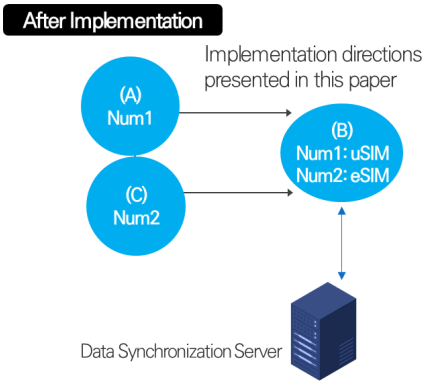
III. CONSIDERATIONS FOR IMPLEMENTATION

We propose the following implementation concept to provide users with convenient data management functions when using two phone numbers using SIM and eSIM (Fig. 1).

First, for each phone number, you'll have separate business and personal zones. For Business, our design provides business history with Num1-uSIM. Also, for personal use, Num2-eSIM provides personal history on separated UI screen. Second, through the design of a dedicated Android app, we provide independent information by work and personal areas to improve the efficiency of data management. Finally, in order to provide eSIM-based two-number service, we provide data synchronization function between smartphone and data server to improve data reliability.



(a) Current using of two numbers on a smartphone



(b) Improved data management with data synchronized server

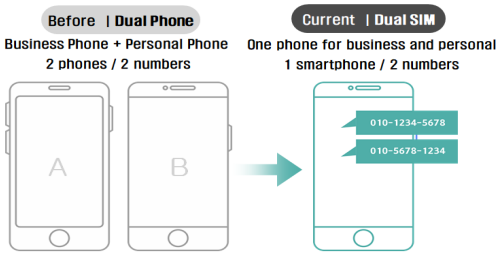
Fig. 1. Proposed implementation directions

IV. Implementation of Two Number X

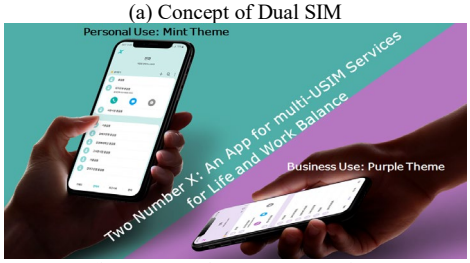
A. Application User Interface(UI)

Android-based smartphones support dual sim functionality in the Phone app and the Text app. However, contacts are not separated by user usage, and distinguishing between SIMs is not intuitive.

Here are some of the key features we implemented in this paper. The first and most important feature is SIM separation. We use a mint-colored theme for the personal number and a purple theme for the business number to separate the two, meaning that the mint-colored personal theme only shows records for the personal number, and the purple business theme only shows records for the business number. And we keep the business number separate by storing it in the Two Number X app itself, separate from your smartphone, so that you can use dedicated contacts. Fig. 2(a) illustrates the concept of using two phone numbers on one smartphone using dual SIMs. Fig. 2(b) shows the main UI of the Two Number X application implemented in this paper, and we designed the UI to intuitively identify personal and business use.



(a) Concept of Dual SIM



(b) Two Number X User Interface

Fig. 2. User Interface of Two Number X App

B. Key Features Module Block Diagram

This section describes the organization of the main functional modules that make up the Two Number X app and the main functions of each module. Fig. 3 shows the organization of the main app modules, which include the Google Drive module, Messages module, Call module, SIM module, and Contacts module. The main functions and features of each module are summarized in Table II.

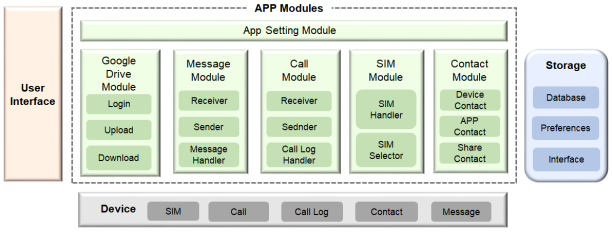


Fig. 3. Key feature modules

TABLE II. Summary of Features in Major Modules

Module	Description
App Setting	· Each setting value is stored as a key - value value in an xml file using Android's SharedPreferences feature, and each key value is encrypted with the AES256-SIV and value value with the AES256-GCM algorithms. The keys used for encryption are stored in a container that only the system can access, not inside the app using the Android keystore system.
Google Drive	· Use the Google API to back up a DB file containing the entire app's data to the Drive feature within your Google account. The DB file is encrypted with "sql cyper" to prevent users from opening the DB file arbitrarily. With the auto backup feature, the DB file is automatically backed up regardless of the APP process at a minimum interval of 15 minutes. Automatic backup works through the WorkManager feature of android, which determines whether the device is connected to the network and helps to keep the work even if the device is restarted.
Message	· Responsible for sending and receiving messages. · Sending messages is done using an ID in the system's thread table, which represents a message conversation with the phone addresses of the conversation participants. · MMS can send not only text messages, but also images, video, and audio byte array data. When a message is received, SMS inserts data into the Telephony.Sms system table and MMS inserts data into the Telephony.Mms system table. · Message notifications are issued separately using different colored icons to distinguish between messages received on personal and business SIMs.
Call	· The Call module implements InCallService, CallConnection, and CallConnectionService to use the system's call functionality, binding the InCallService implemented in the app to receive, place, reject, hold, mute, change audio device (Bluetooth device, speaker), and put the call on hold. · When receiving/placing a call, recognize the SIM used for the call and change the color theme of the call screen. Automatically block calls and send a set rejection text based on your call blocking settings.
SIM	· The SIM module stores the SIM data that you select to use for sales among the active SIM data available on your device. All texts, calls, logs, etc. in the app are based on that SIM data, and Business mode uses the stored SIM data. · Personal mode uses the call and text history of all SIMs except the saved SIM, as it is necessary to display the history of previously used and deactivated SIMs on the device as well as another active SIM that is not a business SIM. · The SIM module detects changes in the SIM status of the device and if there is a change in the SIM used for sales (e.g., if the SIM is unplugged or switched to a different SIM), it prevents the user from using the device and wipes the app data, giving them the choice to re-select the sales SIM or keep the data. If they only have one SIM, then it prevents them from using business mode.
Contact	· Helps you read and modify the contact data and group data tables stored in the system. Stores separate contact data in the app DB for use with sales SIMs, separate from system contacts. This allows you to keep system contacts, which are personal SIM contacts, and sales SIM contacts separate.

C. Configuring the App Menu

Table III summarizes the main types of screens that can appear on an app's screen, along with the detailed menu configuration items. Overall, it consists of a pop-up screen, main screen, settings screen, and security settings screen, and the detailed menu items that can be selected for each screen

are summarized in the description column.

The notes (memos) screen receives a *logType* and *logId* from the previous screen to leave a note on the selected call or text record. The text screen receives a single *phoneNumber* or *threadId*. Because there can be more than one participant in a message conversation, the text screen is passed a thread Id when navigating from the message history and a single *phoneNumber* when navigating to send a text to each individual, such as the contacts screen. If you pass *an msgId* and *msgType* when navigating to the text screen, the message is retrieved by the passed *msgId* so that you can scroll to the text you were viewing when you navigated to the text screen. The contact details screen receives *ContactsId* or *phoneNumber* from the previous screen to retrieve contact information. The security settings are locked, requiring you to enter the PIN number you set when installing the app to access them. The phone screen gets the *Call* object that is currently in progress by binding *InCallService* for displaying call information and call control.

TABLE III. SUMMARY OF THE MENU FOR THE TWO NUMBER X APP

App Screen	Description
Popup types	SIM selection, Call, Memo(hashtag, text memo, voice memo), Loading, SIM change, SIM deletion, Data initialization, Delete other etc.
Screen types	Intro, Main, Contacts, Setting, Search, Call, Text, Memo screens etc.
Main screen	Keypad, Contacts, Recent History, Text chat rooms screens etc.
Configuration	Manage unsubscribes, Manage decline messages, Manage quick notes, Manage excluded phone numbers, Security settings, Default app settings, Import device contacts, Set up SIM auto-switching, and Change app icons etc.
Security setting	Hidden mode, Work SIM lock, App initialization

D. Purchasing, Billing, and Unsubscribing Apps through the Google Play Store

The Interaction between the data server we built and Two Number X app and Google billing server when the mobile app is purchased from the Google Play Store is shown in Fig. 7, Fig. 8 respectively. When Two Number X requests a promo code, the data server verifies the code and sends the verification result to the mobile app (Fig. 4).

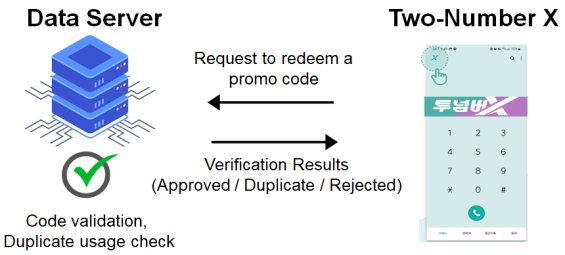


Fig. 4. Procedure for redeeming a promo code

Fig. 5 shows that when a user purchases the Two Number service using a promotional code and starts using the service, the interaction between the mobile app and two different data servers completes the process, including payment.

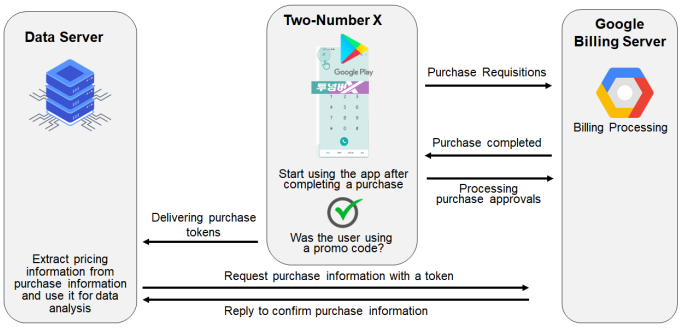


Fig. 5. Procedure for users' Two Number X subscriptions and purchases

Fig. 6 illustrates what happens when Two Number X users terminate their use of the service and unsubscribe.

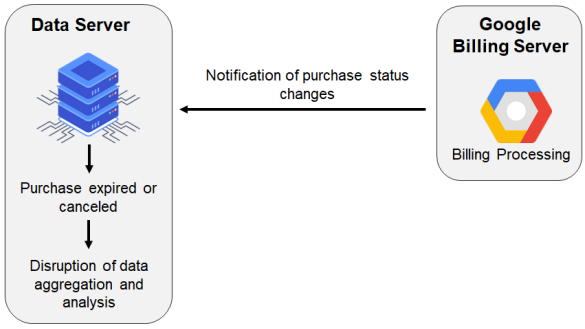


Fig. 6. Procedure for unsubscribing a user

V. CONCLUSIONS

We described the implementation of a mobile app that can independently provide a mobile phone usage environment for personal and business use in order to promote the value of balancing the user's life and work when using two phone numbers using eSIM. We hope that the implementation contents of this paper can help developers of mobile apps for similar purposes.

ACKNOWLEDGMENT

This work is financially supported by the Ministry of Trade, Industry & ENERGY(MOTKE) through the fostering project of project for industry-university convergence district build-up.

REFERENCES

[1] N. Mallikharjuna Rao and P. Seetharam, ""Multiple Network Operator Services Utilization Using Single SIM Card," International Journal of Computer Theory and Engineering, Vol. 3, No. 3, June 2011

[2] O. Vikhrova, S. Pizzi, A. Terzani, L. Araujo, A. Orsino, G. Araniti, "Multi-SIM support in 5G Evolution: Challenges and Opertunities", Challenges and Opportunities" IEEE Communications Standards Magazine, Vol. 6, Issue 2, June 2022

[3] Bo-Kyem Lee, "Japanese smartphone eSIM adoption policy Introduction to contents and domestic status", KISDI Perspectives, No. 2, 2022

[4] Bo-Kyem Lee, "Trends in eSIM standardization and adoption", KISDI, Vol. 30, No. 1, 2018

[5] GSMA, "The Mobile Economy 2021," 2021, [Online]. Available: https://www.gsma.com/mobileeconomy/wp-content/uploads/2021/07/GSMA MobileEconomy2021 3.pdf. [Accessed on: January 21, 2022]

A.I. Data Sourcing Platform *'Sullivan RealEyes'* based on Public BlockChain

ChanYong Park
Dept. of AI Dev Team
TUAT Corp.
Daegu, S.Korea
bw_yong13@tuat.kr

YoungMin Im
Dept. of AI Dev Team
TUAT Corp.
Daegu, S.Korea
youngmin@tuat.kr

YoungHeuk Cho
Dept. of AI Dev Team
TUAT Corp.
Daegu, S.Korea
philip@tuat.kr

KyungSu Kwon
Dept. of Game
College of Media Contents
Dongseo Univ
Busan, S.Korea
kskwon@dongseo.ac.kr

JinWook Kim
School of computer Science and Engineering
KyungPook National Univ.
Daegu, S.Korea
deepkaki@knu.ac.kr

***Abstract*—In recent years, there's been a rising interest in evaluating ESG activities of companies objectively. The environment sector, especially the carbon trading market, is gaining traction with the emergence of related technology companies. However, systems to evaluate social value activities are scarce. Social value is crucial for community and societal development, so having a technology to objectively assess it holds significant importance.**

This paper introduces *'Sullivan RealEyes'*, a blockchain based system to objectively validate social activities of companies and individuals. The system interfaces with *'Sullivan Plus'*, an AI-powered visual assistance service for the visually impaired, recording activities aimed at enhancing their quality of life on the blockchain. It outlines a reward system designed to sustain these social value activities, ensuring continuous operation through blockchain technology.

Keywords— Blockchain, ESG, Social Value, Cripto Asset, AI Training Data

I. INTRODUCTION

Blockchain's key advantage lies in its decentralized ledger, which upholds trust without central authority. This quality has spurred attempts to merge blockchain with carbon credit markets due to concerns about ESG issues. Despite significant strides in applying blockchain to environmental problems, its application to social value activities remains limited[1,2,3].

The interconnection of blockchain technology with the crypto currency market necessitates the creation of synergies with established industries, exploration of new business models, and the pursuit of strategic investments and partnerships as part of the ongoing reorganization of the crypto currency market. Not long ago, the concept of business opportunities related to virtual assets like DeFi and NFTs seemed far-fetched. However, blockchain based projects are actively expanding their ecosystems and establishing themselves as new economic players[4]. This led to the explosive growth of the crypto currency market in 2021, followed by a recent correction due to factors like US monetary policy changes and the *Terra-Luna* incident. This correction process involves weeding out uncompetitive corporate services among companies that emerged during the virtual asset boom and reorganizing them into more sustainable entities.

Despite the ongoing global correction in the crypto currency market, many companies are actively seeking investment opportunities to prepare for the future expansion of the crypto currency ecosystem. This reorganization is viewed as a period of renewal, aiming to address the challenges stemming from the rapid growth of the crypto currency market and explore avenues for new value creation.

As market participants become more sophisticated, they are focusing on discovering fresh business opportunities by evaluating various aspects of crypto asset utilization. This includes addressing issues related to the protection of market participants and investors, which have become increasingly significant.

Assets that were previously not considered, such as patents and copyrights, have been reclassified into new asset categories as the definition of assets evolves. This shift suggests that crypto assets may also hold substantial economic intrinsic value and could eventually become part of mainstream investment portfolios. Thus, maintaining a flexible approach rather than adhering to rigid positions on virtual assets is crucial. Continuous monitoring of the virtual asset ecosystem, identifying emerging trends, and making informed decisions related to virtual asset investments and associated businesses are essential practices in this dynamic landscape.

In the realm of blockchain convergence with ESG, most cases are centered around environmental issues, particularly in the carbon trading market and agriculture supply chain traceability[5,6,7,8]. However, very few instances exist where blockchain is effectively utilized to objectify the social value initiatives of companies, organizations, or individuals[9,10]. Moreover, none of these attempts have demonstrated sustainable operations. Even overseas, a handful of blockchain projects have

This research was supported by the Ministry of SMEs and Startups of Korea.

ventured into social value activities, yet none have managed to establish a service structure that can be maintained sustainably. Despite good intentions, sustainability is a critical factor, as initiatives lacking it tend to lose their value over time.

This paper introduces *'Sullivan RealEyes'*, a blockchain service developed by *TUAT*, which utilizes blockchain to store and objectify social activities. This service collects AI training data continuously to offer advanced visual assistance services for the visually impaired. *'Sullivan RealEyes'* records and rewards the efforts of companies and individuals on the blockchain, emphasizing the importance of high quality AI training data in the competitive business field.

II. Blockchain Ecosystem

Blockchain technology decentralizes trust through a shared ledger system, recording network transactions in blocks at regular intervals. This shared information among users enhances security compared to centralized storage. Public services and private sectors are adopting blockchain, creating unique ecosystems and advancing technology.

A. Private blockchains for public services

The South Korean government launched six blockchain pilot projects in 2018, and since 2019, it has conducted pilot projects in online voting, simple real estate transactions, livestock traceability, personalized customs clearance, cross-border electronic document distribution, and shipping logistics[11].

- **Convenient and reliable online voting** : Improves reliability by recording online voting information on the blockchain and enabling stakeholders such as election candidates and observers to verify the voting process and results directly.
- **Make real estate transactions easy :** Eliminate the hassle of having to visit multiple authorities such as the residence center and the National Tax Service to submit documents when requesting a real estate mortgage.
- **Traceability of beef :** Reduce traceability from up to 6 days to less than 10 minutes in the event of a problem by sharing information on the blockchain at every stage, from breeding to slaughter and sale.
- **Expedite personalized customs clearance and prevent false declarations :** Record the entire process from order to shipment, delivery, and customs clearance on the blockchain, reduce customs clearance time and logistics costs with real-time import declaration, and prevent under-declaration cases.

B. Public blockchains combined with crypto markets

A virtual asset is defined as a digital financial asset with economic value based on blockchain distributed ledger technology and cryptography that can be traded or transferred electronically, and related terms include ‘crypto currency’, ‘virtual asset’, ‘crypto asset’, ‘digital asset’, ‘virtual currency’, ‘coin’, ‘token’ and etc.

Crypto currencies don't have a long history. There were attempts to create digital currencies before crypto currencies, but there was no way to solve the problem of double spending. This is because digital assets need to be able to be used only once, preventing duplication and counterfeiting. Blockchain technology solved this problem, and the first crypto currency created with it was *Bitcoin*, while other coins are called altcoins.

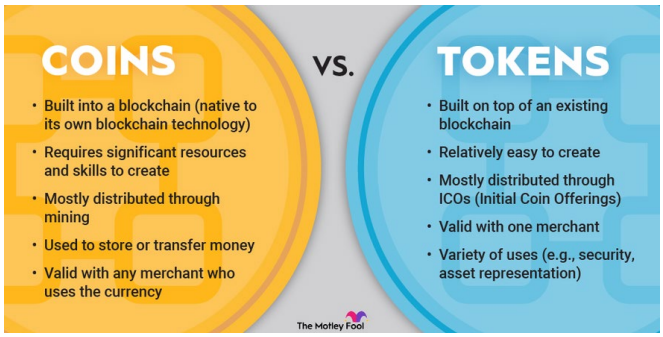


Fig 1. Comparision of Coins and Tokens

In addition, blockchains that allow anyone to join the blockchain network and contribute to the verification and storage of transaction data, such as *Bitcoin* and *Ethereum*, are called public blockchains, and they all operate their own blockchain network called mainnet[4].

Starting with the *Ethereum* blockchain, which emerged after *Bitcoin*, the first generation blockchain, blockchains began to develop into platforms, providing the ability to develop and operate various decentralized services. The dApps running on these platform blockchains can issue crypto assets called tokens, which can be divided into utility tokens, exchange tokens, and security tokens.

III. Convergence of AI and Blockchain

AI products rely on quality data. Decentralized blockchain, by securely sharing and managing fragmented data with participant verification, ensures data integrity, creating high-quality training data..

A. Issues in building AI training data

High-quality and extensive data are crucial for enhancing the performance of AI services, especially those employing deep learning techniques. While traditional services improved through algorithmic enhancements, deep learning-based AI relies heavily on superior data for effective learning[11]. Big tech companies, equipped with substantial data, dominate the AI market, leading to a widening performance gap. Countries with abundant high-quality AI training data are likely to monopolize the AI market, accentuating the global competitiveness disparity.

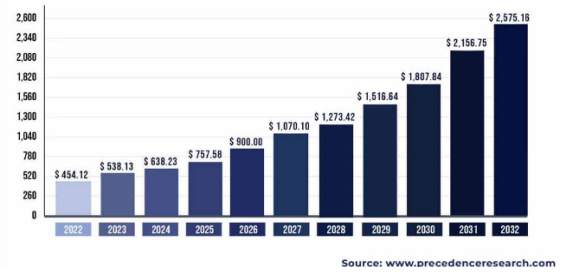


Fig 2. AI market size, 2022 to 2032

Recently, computing infrastructure has also contributed to the growing performance gap in AI. Many startup companies

face challenges in acquiring data due to costs and time constraints. New services lack adequate training data, necessitating either self-production or costly purchases, adding financial burden. Even after purchase, refining and verifying the data demand significant time and resources, hindering AI service launches. To address this, *the Ministry of Science and ICT of Korea* initiated the 'Data Dam Project', aiming to reduce initial costs and time burdens for companies and boost the national AI industry[12,13]. Although substantial data has been accumulated in the *AI Hub* (*www.aihub.or.kr*), it struggles to meet the diverse needs of the rapidly expanding AI market. Consequently, the crowd data sourcing method, utilizing blockchain, is emerging as an alternative solution.

B. *The case for combining blockchain and AI*

The utilization of artificial intelligence in blockchain has several active use cases, notably in the medical and healthcare sectors. *IBM's Watson* stands as a prominent example, supporting lung cancer diagnosis and treatment decisions since 2013. Korean startups like *Vuno* and *Lunit* provide deep learning-based services for analyzing X-ray, CT, and MRI images. By integrating blockchain with these AI diagnostic services, hospitals can securely share medical data, creating a valuable source of learning for AI and significantly improving the performance of medical services.

Additionally, *MediBlock*, a healthcare company, developed a blockchain-based dental examination electronic medical record (EMR) system in April 2019. This system enables sharing diagnosis results across medical departments when patients complete a medical questionnaire. By integrating and utilizing the accumulated questionnaire data as training data for AI diagnostic services, faster and more accurate diagnoses can be achieved.

IV. PUBLIC BLOCKCHAIN PLATFORMS

The majority of blockchain projects are designed as decentralized applications (dApps) operating on platform blockchains, unless they are creating their own mainnet. Common platform choices include *Ethereum*, *Solana*, and *Polycon*. Additionally, Korean tech giants like *Naver* and *Kakao* have introduced their platforms, *Link* and *Klaytn*, respectively, adding to the available options for developers.

A. *Problems with blockchain-based services*

Blockchain based services face three significant challenges. Firstly, these services are often approached from an investor's standpoint, focusing on token price and market capitalization. This investor-centric approach marginalizes users who actively contribute to the ecosystem, hindering the tokens from evolving as practical technologies for real-world applications.

The second issue arises from the launch strategies of crypto projects. Many projects conduct token sales before the mainnet launch, emphasizing high bonuses for early investors and prioritizing price and volume over ecosystem development. Consequently, there's an uneven distribution of tokens, hampering the crypto ecosystem's growth. Additionally, dApp developers often insist on using their own tokens, further fragmenting the ecosystem and slowing overall progress.

Lastly, user experience (UX) poses a challenge, particularly regarding digital wallet creation. Many cryptocurrency users, especially non-technical ones, find setting up a digital wallet to be a daunting task, acting as a significant barrier to entry. This challenge has intensified as the industry attracts more non-expert users, highlighting the need for more user-friendly solutions.

B. *Klaytn blockchain*

Launched in June 2019, *Klaytn* is the public blockchain platform of *KaoKao*, South Korea's leading mobile platform, providing an accessible user experience and development environment to deliver the value of blockchain technology. *Klaytn* aims to integrate the low latency and high scalability of private blockchains with the decentralized governance and distributed control of public blockchains in an efficient hybrid design[14].

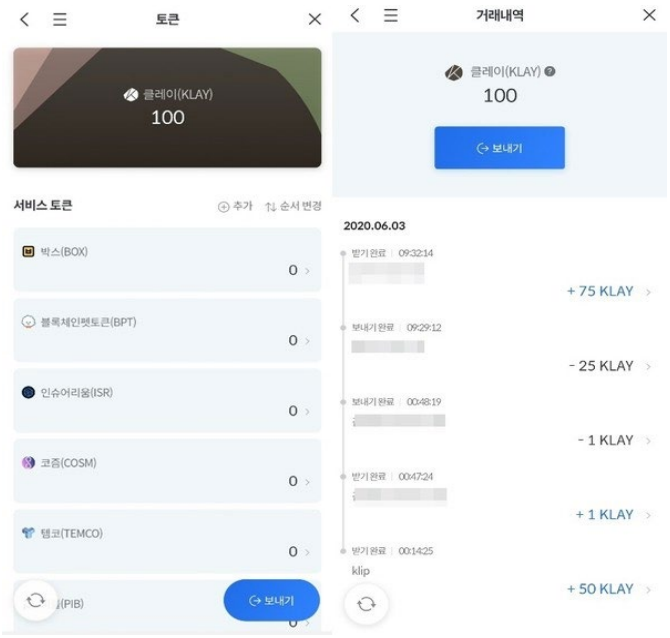


Fig 3. *Klip*-a cripto currency wallet of *Klaytn*

In addition, to build a dApp ecosystem, the company has developed and provided a crypto currency wallet, *Klip*, that allows users to easily access their crypto assets and is integrated with *'Kakao Talk'*, South Korea's largest messenger service. Like a typical crypto currency wallet, *Klip* is based on its own crypto currency, *Klay*, and also supports various *Klaytn* based tokens, *KCT*, which helps projects that issue *Klayken* based tokens to attract users.

Klaytn is characterized by a block generation time of about 1 second, dramatically increasing processing speed compared to other blockchains (over 4,000 TPS), and prioritizing scalability[14].

As a public blockchain, *Klaytn* has been active in building a dApp ecosystem since its inception. It has been pursuing rapid growth by sacrificing the decentralization aspect and strengthening the scalability aspect among the three elements of blockchain: decentralization, safety, and scalability.

V. System Implementation and Features

When planning the system, we considered two blockchains to be used as the mainnet: *Klaytn*, led by *Kakao*, and *Link*, led by *Naver*.

While the *Klaytn* blockchain enhanced scalability and led to the expansion of dApps, the *Link* blockchain envisioned the launch of several services based on *Line* services in the form of dApps, so *Klaytn* was finally selected as the mainnet.

A. Sullivan RealEyes

Before introducing *'Sullivan RealEyes'*, it is necessary to introduce a service called *'Sullivan Plus'*. *'Sullivan Plus'* is a visual aid app that uses artificial intelligence image recognition technology to become the eyes of the visually impaired, and its main functions include object recognition, character recognition, and face recognition. It started with domestic services in 2018 and expanded to global services in 2019, and is currently operating for more than 300,000 users in all countries.

'Sullivan Plus' requires continuous data collection and AI learning for neural network updates. To address this, we developed *'Sullivan RealEye'*s, a blockchain based platform for crowd sourced AI training data.

Fig 4. Workflow of *'Sullivan Plus'*

Current efforts involve providing product information to visually impaired individuals through QR codes and Braille, but these methods have limitations, including limited information and practical difficulties. *'Sullivan Plus'* aims to enhance the quality of life for the visually impaired by offering real-time product information through camera recognition.

However, the rapid influx of new products poses a challenge as existing AI networks can't accurately identify them. *'Sullivan RealEyes'* seeks to solve this problem by enabling accurate recognition of newly released products.

B. Participant compensation structure at Sullivan Realeyes

In order to provide AI visual assistance services that recognize the latest products, we send AI training data requests to *'Sullivan RealEyes'* users. In this case, the training data request is made by the system administrator, who analyzes various data collected from visually impaired people and specifies the data that needs to be collected specifically.

When the administrator asks the participants to collect data through the *'Sullivan RealEyes'* system, the participants find the product, take a picture of it, and upload it with descriptive text. An artificial intelligence neural network model is then created from the uploaded data and made available for download to the visually impaired.

At this time, the photo data uploaded by the participants to the system is verified by other users participating in the system and finally accepted. Users who participate in the verification are eligible if they have been active in *'Sullivan RealEyes'* at a certain level.

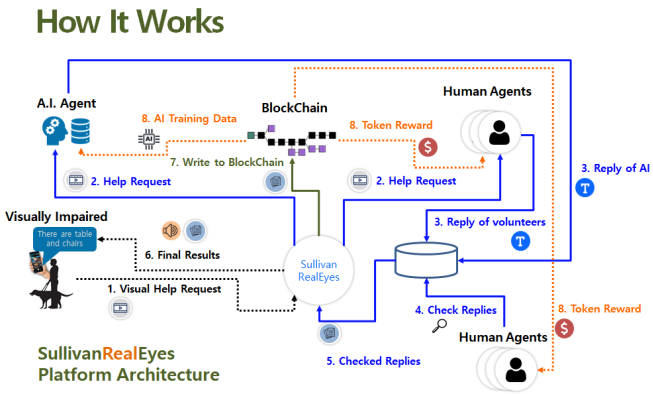


Fig 5. Compensation Structure of *'Sullivan RealEyes'*

In order for such a service to be sustainable, it needs a structure that generates minimal revenue. Therefore, AI neural networks trained for general use are available for free download and use, and AI neural networks that are specifically customized for recognition by the visually impaired are provided for a small fee. In order to purchase a customized neural network, the user must first purchase a token issued by *'Sullivan RealEyes'* and use it to pay for it. In other words, the token generated by the system is given economic value.

In addition, those who participate in data collection are rewarded with tokens as a reward for their data collection contributions, thus enabling participants to continue their good deeds.

C. The Value of 'Sullivan RealEyes'

A typical crowdsourcing-based data labeling system involves participants collecting and generating data in response to a company's request to build AI data.

'Sullivan RealEyes' is a platform that generates high-quality AI training data by requesting data necessary for training visual assistance neural networks, collecting data from participants, and verifying the data collected by other participants.

In addition, in order to ensure the continuous operation of the service, *Klaytn's KIP-7* token (Fungible Token), which corresponds to the *ERC-20* based token of the *Ethereum* blockchain, has been prepared as a user compensation plan, and participants can easily view their virtual assets through the clip wallet available on *Klaytn*. In particular, the reward tokens given to users are proof of their good intentions to improve the quality of life of the visually impaired, so their value has a different meaning than tokens provided by existing blockchain systems.

Currently, more than 1 million photos of blind people asking for help are uploaded to *'Sullivan Plus'* every month, so it is expected that by continuously analyzing and accumulating this data, it will be possible to gradually build excellent big data for AI training.

VI. Conclusions

Recently, there's been a surge in interest surrounding companies' environmental, social, and governance (ESG) activities. While the environmental sector has witnessed growth, there's a lack of objective evaluation methods for social activities, which are crucial for community and societal development. This paper proposes a blockchain based system to objectively verify and record social activities. The system, linked to a service aiding the visually impaired, stores and rewards activities performed by companies and individuals on the blockchain. This approach reflects a broader trend where social value activities are recognized as assets, suggesting an upcoming era of economic value creation through objective measurement and evaluation of social contributions.

'Sullivan RealEyes' operates through a process where participants upload product details, which are then verified by others and registered as AI learning data. All user activities in this process are recorded on the blockchain, and participants receive tokens as rewards. This transparent system, coupled with predetermined token rewards, anticipates positive future outcomes.

Unlike previous government-led pilot projects in the social value blockchain sector, the proposed system establishes a sustainable operating structure using virtual assets. This approach suggests a stable operation for blockchain services in the public sector by startup company, making the operation of *'Sullivan RealEyes'* highly anticipated.

Acknowledgment

This research was supported by the Ministry of SMEs and Startups of Korea, 2021-2023 [S3140864].

References

[1] R. C. Park, Y. S. Lee, "An Overview of Blockchain Technology: Concepts, Consensus, Standardization, and Security Threats," The Journal of JISPS Vol. 20, No. 4, pp218~225 December. 2019.

[2] S. D. Yoo, "A Study on Blockchain Ecosystem," The Journal of IIBC, Vol. 18, No. 2, pp.1-9, Apr. 30, 2018.

[3] J. O. Lee, O. H. Kwon, "Analysis of Research Trends in the Field of Blockchain-based Supply Chain Traceability," Informatization Policy Vol. 28, pp.003-033, No.2. 2021.

[4] "Types of Cryptocurrency," https://www.fool.com/investing/stock-market/market-sectors/financials/cryptocurrency-stocks/types-of-cryptocurrencies/

[5] A. E. Mane, Y. Chihab, K. Tatane, and R. Korchiyne, "Agriculture Supply Chain Management Based on Blockchain Architecture and Smart Contracts," Applied Computational Intelligence and Soft Computing, Volume 2022, 23 pages

[6] S. A. Bhat, N. F. Huang, I. B. Sofi, M. Sultan, "Agriculture-Food Supply Chain Management Based on Blockchain and IoT: A Narrative on Enterprise Blockchain Interoperability," 12, 40, Agriculture 2022.

[7] Rosanna Cole, Mark Stevenson, James Aitken, "Blockchain technology: implications for operations and supply chain management", Supply Chain Management: An International Journal, Volume 24, Number 4, , 469–483, 2019.

[8] A. Park, and H. Li, "The Effect of Blockchain Technology on Supply Chain Sustainability Performances," Sustainability, 13, 1726, 2021.

[9] "Cryptocurrency: Driving Social Value and Advancing Sustainable Development Goals," https://medium.com/coinmonks/cryptocurrency-driving-social-value-and-advancing-sustainable-development-goals-f35a97b46aee,

[10] H. T. Yang, "Artificial Intelligence and Blockchain Convergence Trend and Policy Improvement Plan," Informatization Policy Vol. 27, pp.003-019, No.2, 2020.

[11] NIA Report, "Global AI levels and implications in data," IT & Future Strategy, No 2. 2020.

[12] NIA Report, "How to improve the effectiveness of your data for AI training Policy directions to improve effectiveness," IT & Future Strategy, No 7. 2020.

[13] NIA Report, "3 Resourcing Strategies to Enable AI - Data, algorithms, and computing power," IT & Future Strategy, No 5. 2020.

[14] "Klaytn official website," https://klaytn.foundation/

Comparison of gaze activity between HMD and Theater Screen

TAKAMURA Tamaki
Department of Media Arts
Tokyo Polytechnic Univaersity
Tokyo, Japan
m2371004@st.t-kougei.ac.jp

NATE Hisaki
Department of Imaging Arts
Tokyo Polytechnic Univaersity
Tokyo, Japan
nate@img.t-kougei.ac.jp

***Abstract*—In this study we measured gaze points in viewing video of HMD and Theater Screen. As a results, showed that the movement of gaze points between HMD and Theater Screen was similar, but the overall gaze rate (gaze time/ video duration) was found to be slightly lower with the HMD compared to the screen.**

Keywords—HMD, Theater Screen, fixation

I. INTRODUCTION

Today, we can purchase a device that display visual information of the real world in a 360-degree. In the past, large-scale devices such as Dome Projection Screens and multiple projectors were required to create 360-degrees video. However, nowadays it is possible to display 360-degrees video using stand-alone devices such as Meta's Meta Quest 3 [1], HTC's VIVE Focus 3 [2], and PICOs Pico 3 Pro [3]. These systems are capable of acquiring the position of the head-mounted display (HMD) and displaying video corresponding to the position. People are exploring the potential of 360-degrees video contents and service in many fields such as film, gaming, fitness, education, healthcare, travel, and other simulation systems.

Various attempts are being tried in the field of movies. Basically, there are many VR movies that display 360-degrees videos and interactive 6Dof or 3Dof contents. However, there are distributing video service which display a conventional 2D video production in 360-degrees video using HMDs. In this study, we examined whether traditional Theater Screen can we replaced by screenings using HMDs. If possible, at home, people can have an experience that they are in a movie theater. We discussed with the characteristics of screenings using HMDs in second section.

II. CHARACTERISTICS OF SCREENINGS USING HMDS

A 360-degrees video display system using HMD displayed images corresponding to the viewer's head position. Unlike existing projection systems in Dome Projection Screens or movie theaters, only a part of the video is displayed, rather than the entire video being shown.

There are various types of distributing video service using HMD. In BIGSCREEN(Fig.1)[4], the actual cinema space is simulated, allowing viewers to change their seating position as if they were sitting in an actual movie theater. Furthermore, multiple persons can watch together in the same on-line space. In Netflix[5], viewers watch videos played in a space like a log cabin space, sitting in front of a large monitor. As the video plays, the areas outside the playing video gradually darken. In YouTube [6](Fig.2) and Amazon Prime [7], when selecting a video, the video is displayed on a large screen in a dark space. There is various distributing video service which display a conventional 2D video production in 360-degrees video using HMDs, ranging from service that simulate the entire space to service that only play video.

Fig. 1. BIGSCREEN

Fig. 2. YouTube

However, all these services have the common feature of projecting videos on a large screen in a darkened space. HMD can obtain the information of viewers head position in three axes (x, y, z) and three directions (yaw, pitch, roll). There are few services in the HMD-based movie streaming that can support 6DoF, which includes direction, but most

services support 3DoF. In these services, when the viewer moves their head, the display screen of the video remains fixed in a specific position in space and does not follow the movement of the head.

There are applications for HMD that bring the movie theater experience closer. Movie theaters are characterized by large screens, immersive audio systems, and a large number of viewers. In particular, the feature of a large screen, as seen in screening methods like IMAX[8] and SCREEN X[9], is one of the major characteristics of movie theaters. The enlargement of such screens is easily achievable in the CG space using HMD.

However, there is a possibility of a different visual experience compared to watching movies in a theater. In particular, form the point of visual perception, there are differences from traditional screens such as direct projection of LCD light onto the retina, non-uniform image quality caused by lens edge correction of HMD, ease of neck movement due to the weight of HMD, and the lack of information other than the field of view. These factors may result in a different visual experience compared to watching movies on a theater screen.

III. AIM OF THIS STUDY

We investigated the difference between theater screen simulated by HMD and the actual theater screen by measuring the gazing point of the viewer.

IV. METHOD

We measured the gaze points when viewing the displayed images on the HMD (VIVE Pro Eye) and the movie theater screen (253 inches)(Figure3). The video displayed on the HMD was captured using a 360-degree camera (Insta360 One R), adjusted to the position of the head while seated in the movie theater screen experimental setting. The target video was then overlayed onto the same portion of the theater screen for projection. The gaze points were measured using an eye-tracking device (Tobii Technology). The presented images were ones in which a target red circle moved left to right(Figure 4), bottom to top(Figure 5), and shrank in the 4 corners (Figure 6). The videos are presented in the order of left to right, bottom to top, and shrinking in the 4 corners, with each sequence repeated 5 times. The duration of presentation for each image is shown in the table 1. There were 11 participants, all of whom had normal vision, and they observed 6 conditions of images by combining 3 types of images and 2 types of display devices.

Fig. 3. the movie theater screen (253 inches)

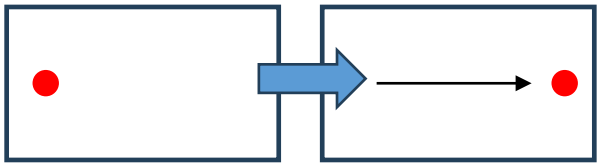


Fig. 4. red circle (taeget) moved left to right.

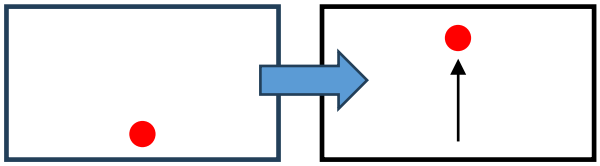


Fig. 5. red circle (taeget) moved bottom to top.

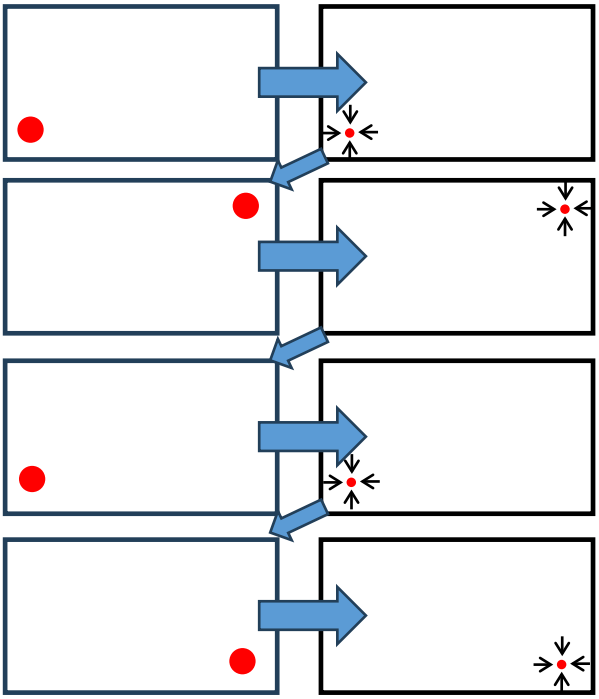


Fig. 6. red circle (taeget) moved shrink in the 4 corners.

TABLE I. DURATION OF VIDEO

Video	Duration
left to right	3 sec
bottom to top	2 sec
shrink in the 4 corners	8 sec (2 sec in the 1corners)

V. RESULTS

Table 2 ,3 showed the average gaze rate (= gaze time on target/gaze time on image display) on the target displayed in each video. The measurement area determined as focusing on the target was created by following a red circle of 1.25 times the diameter of the red circle in case of left to right and bottom to top conditions. Shrink in the 4 corner conditions, diameter of the measurement area were fixed with 1.25 times the maximum diameter of the red circle. A higher proportion of gaze rate showed a longer duration of target fixation.

TABLE II. AVERAGE OF GAZE RATE IN THEATER SCREEN CONDITION

	left to right	up to down	shrink in the 4 corners
Subject 1	34%	38%	71%
Subject 2	34%	38%	71%
Subject 3	13%	28%	32%
Subject 4	29%	38%	67%
Subject 5	24%	43%	11%
Subject 6	34%	32%	42%
Subject 7	30%	47%	61%
Subject 8	29%	30%	77%
Subject 9	26%	24%	16%
Subject 10	27%	40%	60%
Subject 11	36%	47%	63%

TABLE III. AVERAGE OF GAZE RATE IN HMD CONDITION.

	left to right	up to down	shrink in the 4 corners
Subject 1	23%	43%	40%
Subject 2	4%	2%	15%
Subject 3	42%	36%	37%
Subject 4	12%	30%	38%
Subject 5	29%	24%	46%
Subject 6	41%	43%	45%
Subject 7	26%	30%	40%
Subject 8	28%	22%	41%
Subject 9	13%	35%	39%
Subject 10	25%	32%	51%
Subject 11	9%	8%	12%

A two-factor analysis of variance was performed on the gaze rate of the measurement results. There was a significant difference in the image factor (F (1,54)=76.2290, p<.001) and the device factor (F(2,108)=4.3603, .p<.05). There was no significant interaction effect (F(2,108)=0.3063 p>.10). As a result of a post hoc test (Shaffer's modified sequentially rejective Bonferroni procedure), in the image factor, the gaze rate was higher in the order of shrink in the 4 corners, bottom to top , and left and right (shrink in the 4 corners > bottom to top, p<.001; bottom to top > left and right, p<.001; shrink in the 4 corners > left to right, p<.001).

There were differences of gaze rate between images used in the experiment. This result suggested the possibility of the presence or absence of target movement and the anisotropy of vertical and horizontal motion perception. The gaze rate of the HMD was lower than that of the theater screen. It showed that the gaze time of the HMD's target lower than the gaze time of the screen's target. Why were the gaze rate of HMD lower?

One possible cause was fatigue caused by wearing the HMD. When wearing the HMD, the weight of the HMD itself was supported by subjects' head, which can lead to fatigue in the head. In this experiment, all subjects initially observed the video displayed on the theater screen. Later, they observed the video displayed on the HMD.

VI. CONCLUSION

In this study, we simulated a movie theater screen using an HMD and conducted experiments to measure the characteristics of gaze during visual observation of the screen. The results showed that the target had a longer gaze duration on the theater screen compared to the HMD. It is possible that the fatigue caused by the HMD resulted in shorter gaze duration. On the other hand, regardless of the display method, gaze duration was longer for target's left to right movement, bottom to top movement, and shrink in the 4 corners. Similar trends were observed between the HMD and theater screen, regardless of the displayed images.

VII. FUTURE PROBLEM

In this experiment, we did not measure the movement of the subjects' heads using an HMD. In the future, we would like to focus not only on comparing eye gaze data but also on changes in head position.

VIII. REFERENCES

[1] Meta(2023). Meta Quest 3: New Mixed Reality VR Headset. https://www.meta.com/quest/quest-3/ (accessed 2023-9-30)

[2] VIVE. VIVE Focus 3 - VR Headset for Metaverse Solutions. https://www.vive.com/us/product/vive-focus3/overview/ (accessed 2023-9-30)

[3] PICO. PICO 4 is the latest virtual reality headset from PICO. https://www.picoxr.com/global/products/pico4 (accessed 2023-9-30)

[4] Bigscreen. Bigscreen -Software. https://www.bigscreenvr.com/software (accessed 2023-9-30)

[5] Meta. Netflix on Meta Quest. https://www.oculus.com/deeplink/?action=view&path=app/2184912004923042&ref=oculus_desktop (accessed 2023-9-30)

[6] Meta. Youtube VR on Meta Quest. https://www.oculus.com/deeplink/?action=view&path=app/2002317119880945&ref=oculus_desktop (accessed 2023-9-30)

[7] Meta. Prime Video VR on Meta Quest. https://www.oculus.com/deeplink/?action=view&path=app/2274617532624269&ref=oculus_desktop (accessed 2023-9-30)

[8] Koudai ITO.Mitsunori TADA(2018). Proceedings of The23rd Annual Conference of the Virtual Reality Society of Japan. Influence of weight and balance of head-mounted displays on physical load .14E-4

[9] Toshiaki SUGIHARA,Tsutomu MIYASATO(1998). JVRSJ.Human Factors Issues with Head Mounted Displays(HMD).3(2) 126-131

Development of a Computer Vision-Based Medical Display Monitoring System

Hyeon Hoo Hwang
School of Computer Science and Engineering
Kyungpook National University
Republic of Korea
hhh2022@knu.ac.kr

Jin Ho Lee
School of Computer Science and Engineering
Kyungpook National University
Republic of Korea
ghfkdrjal777@knu.ac.kr

Jun Hyeok Jang
School of Computer Science and Engineering
Kyungpook National University
Republic of Korea
wkdwnsgur@knu.ac.kr

Min Jang
Department of Veterinary Surgery,
College of Veterinary Medicine,
Kyungpook National University
Republic of Korea
jangmin@knu.ac.kr

Soon Ki Jung*
School of Computer Science and Engineering
Kyunpook National University,
Republic of Korea
skjung@knu.ac.kr

Abstract − Monitoring patients' health status in real-time using medical monitors is of utmost importance. This research applies computer vision technology to automate the medical monitoring system. We propose a method to precisely monitor the patient's health status and improve the quality of medical services by precisely detecting the inner area of the display and classifying the output data through computer vision technology. This monitoring system aims to facilitate continuous verification and recording, thereby aiding healthcare professionals in checking the patient's condition more efficiently.

Keywords— Computer Vision, Monitoring System, Medical Monitor Equipment, Canny Edge, Hough Transform, Sub Pixel, Homography, Histogram Equalization

I. INTRODUCTION

In modern healthcare, medical monitors are pivotal in real-time surveillance and diagnosis of patients' health conditions. Furthermore, these devices significantly improve the quality of healthcare services by providing essential information to medical experts [1-4]. Therefore, continuous monitoring and meticulous record-keeping are vital aspects of medical monitoring operations and cannot be ignored [5-7]. Despite the importance of continuous monitoring and record-keeping, it can be challenging to maintain these tasks for medical equipment. In order to solve this problem, it is necessary to establish an efficient management system by automating the collection procedure of medical data and organizing it in a consistent format.

The computer vision used in this study is a technology that processes and interprets image and video data and can play an important role in rapidly recognizing and analyzing information related to patient biometric data. Additionally, computer vision technology is widely recognized for making monitoring tasks more convenient and efficient, rather than relying on more staff members [8, 9]. Therefore, such technology is considered to help collect detailed patient information. The primary purpose of this research is to automate and optimize the medical display monitoring system using computer vision technology. The main development of our work can be summarized as follows:

Fig. 1. Medical Monitor Equipment

- We utilize Corner Detection technology to find precise corner coordinates of the display area for capturing the display region.
- We apply geometric transformations to obtain images from the interior of the monitor's display. This ensures that the location of each data point remains consistent.
- Based on the recorded coordinate regions, we predict the class of each data point using a model trained on image data.

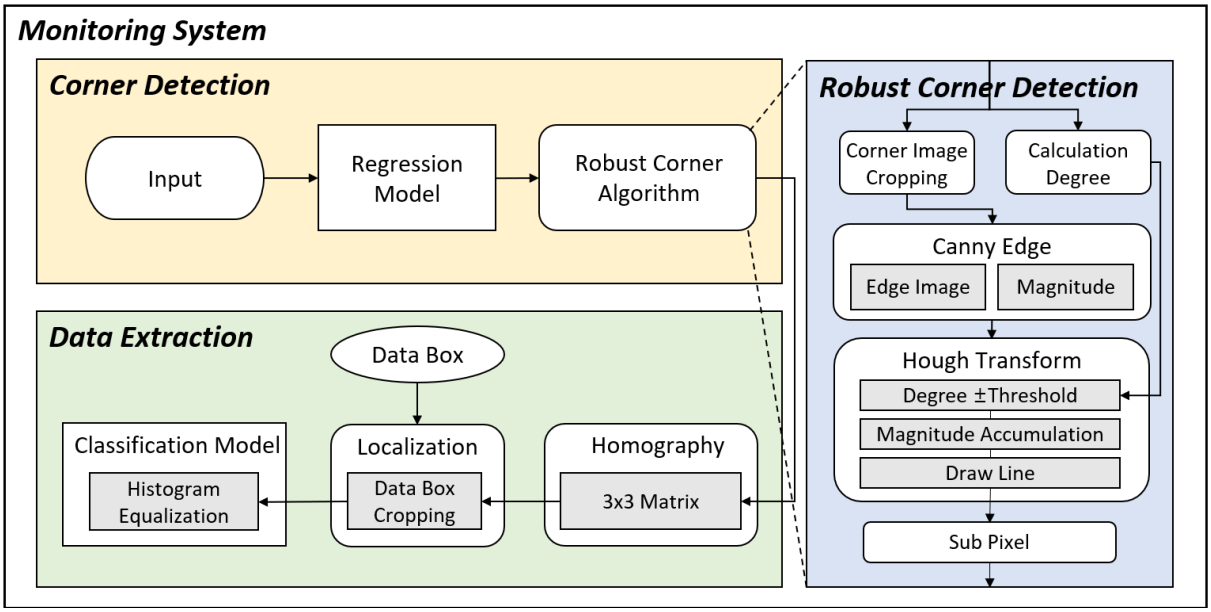


Fig. 2. Monitoring System Architecture

II. Related Work

A. Canny Edge

Canny Edge is one of the critical edge detection algorithms in the fields of digital image processing and computer vision. It is utilized to detect object boundaries within an image accurately. This algorithm involves identifying and extracting edge pixels in an image, allowing for the understanding of an object's shape, contours, and structure [10].

1) Gaussian Smoothing: Gaussian smoothing is a technique used to reduce noise in an image, enhance edge detection, and transform the image from high-frequency to low-frequency components. This process helps eliminate fine details and emphasizes essential edges.

2) Gradient Calculation: Calculating the gradient of an image is a crucial step in identifying edge locations. Gradient operators such as Sobel and Prewitt compute edge magnitude and direction at each pixel.

3) Non-Maximum Suppression: This step involves the selection of edge candidates from the image. It examines the gradient strength of each pixel based on the gradient direction and selects the most substantial edges. This process eliminates duplicate edges, resulting in obtaining thin and precise edges.

4) Hysteresis Thresholding: The high threshold defines firm edges, while the low threshold defines weak edges. Pixels identified as firm edges are selected as actual edges, and those identified as weak edge candidates are chosen as edges only if connected to firm edges. This method enhances edge connectivity and enables precise edge detection within the image.

Canny edge detection is effective in extracting crucial features within images and accurately detecting object boundaries. It finds extensive applications in various fields, such as autonomous vehicles, facial recognition, image segmentation, and object recognition, where edge detection is essential [11-14].

B. Hough Transform

The Hough transform is a technique used in image processing to generate a candidate set of all possible lines for each point within the given image space and to find specific patterns among these candidate lines. It detects circles, lines, and other shapes in images [15]. The most basic Hough transform line detection generally follows the following steps.

1) Edge Detection: Firstly, edges (lines or contours) are detected in the input image. Typically, techniques like Canny edge detection are employed to extract edges.

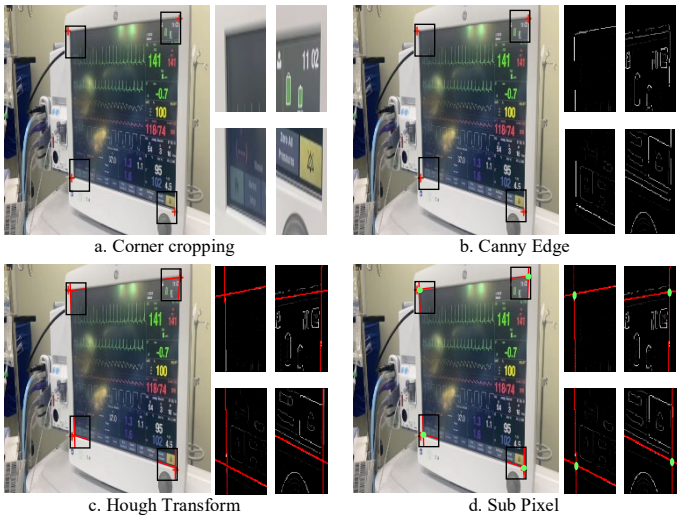
2) Hough Space Generation: Hough transform creates a parameter space known as the Hough space. For instance, when detecting lines, the Hough space consists of parameters used to represent the equation of a line. Each pixel corresponds to a specific line in the Hough space.

3) Hough Transformation: Edge pixel positions are transformed into the Hough space as line parameters. Here, edge pixels are mapped to points in the Hough space representing possible lines.

4) Accumulation and Thresholding: Points of intersection in the Hough space, where parameter lines cross, are identified and accumulated. Only lines with accumulated values above a certain threshold are selected for the final result.

5) Result Presentation: The chosen parameters draw lines or shapes corresponding to the original image. This visually displays the detected lines or shapes in the image.

The Hough transform is used in various applications, such as line, circle, and shape detection. This technique is utilized in autonomous vehicle road recognition, medical image analysis, industrial automation, and many other fields. It holds a



a. Corner cropping b. Canny Edge

c. Hough Transform d. Sub Pixel

Fig. 3. Robust Corner Detection

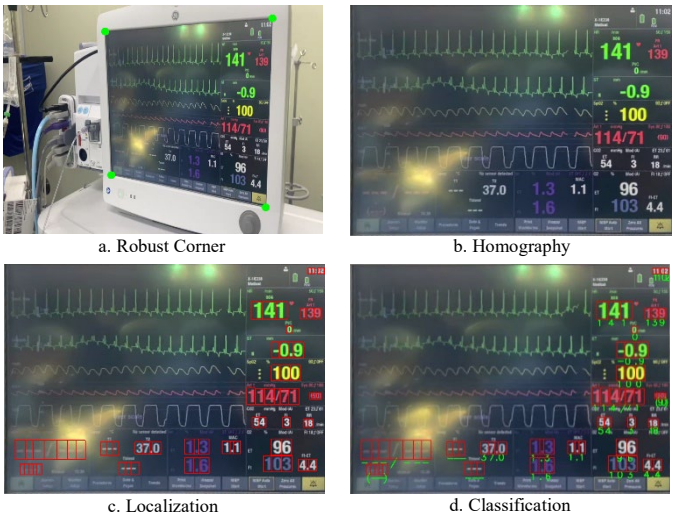
significant position in image processing and computer vision research [16-18].

III. Methodology

The proposed medical display monitoring system has three parts: A. Corner Detection, B. Robust corner detection, and C. Data extraction. The entire system architecture is visualized in Fig 2.

A. Corner Detection

When the monitor's display is analyzed based on the image processing technology, the input image does not have a fixed position and size. In addition, the patient's medical data is only displayed in the inner area of the medical monitor. Therefore, computer vision technology is used to limit the analysis image range to the monitor's display area, and a robust corner coordinate of the display edge line is required. For this purpose, we predict four corner approximate coordinates in the display image of the monitor using the pre-trained Regression Model. However, it is difficult to reliably find the precise corners of the monitor display with these predictions alone, so a robust corner detection process is required.



a. Robust Corner b. Homography

c. Localization d. Classification

Fig. 4. Data Extraction and Result

B. Robust Corner Detection

Robust corner coordinates are essential for extracting and analyzing medical data on monitor displays. In order to obtain robust corner coordinates, each corner image area is cropped and divided based on the approximation of four square corner coordinates using the Regression Model. At the same time, the calculation degree of four straight lines connected based on four coordinates is calculated. This value is then used in Hough Transform. The next Canny Edge task is to proceed, and the girth cropped of the corner area image of 4 is detected. In this process, the edge is detected by calculating the gradient and magnitude of the edge. Next, the edge image and the magnitude value are used in the Hough transform, which performs only the region of the angle's critical value and accumulates the edge's magnitude value calculated in Canny Edge. Two border lines are drawn on the monitor's display found through the previous process. Finally, Sub Pixel is used to find the coordinates of the intersection of the two drawn lines, and the coordinates found are the robust corner coordinates of the display of the medical monitor. Acquisition of robust coordinates can expect better results in Data Extraction.

C. Data Extraction

Based on the Robust Coner detection derived from the previous work, a geometric transformation technique called Homography converts the image frame in which the medical monitor display is photographed into the display interior area. Homography is an image processing technique that transforms pixel coordinates from one plane to another plane. Therefore, we compute a 3x3 transform matrix containing information for transforming from one plane to another and transform it into the interior region of the display by applying a geometric transformation. Each data is localized with the Data Box coordinate, which previously prepares the image of the display interspace transformed into the Homography. Finally, each patient's state data is cropped based on localization coordinates, and the Classification Model predicts the previously learned class of each data, which is then extracted.

IV. Results

A. pre-Trained Model

The pre-trained models consist of both a Regression Model and a Classification Model. The datasets were created based on monitoring data from animals provided by the College of Veterinary Medicine at Kyungpook National University Hospital. For the purpose of training the Regression Model, we collected a diverse set of monitor images from various angles and the corresponding corner coordinates of the display, creating a dataset for training. The trained model takes monitor images as input and produces four corner coordinates as output. This task served as a preliminary step for conducting Robust Corner Detection, allowing for the calculation of approximate values for the four corner coordinates on monitor images captured at different angles. In the Classification Model, data is extracted from images within the display to construct a classification dataset. However, in some instances, there can be images with imbalanced pixel brightness or contrast due to reflections of

light at the data location. To solve this problem, we applied histogram equalization techniques to construct a dataset by equalizing pixel values that are biased to specific values. Likewise, when inputting data into the trained model, Histogram Equalization is applied to conduct the prediction process of the trained model.

B. Result Analysis

This research aims to automate the data collection of the medical display monitoring system and to accurately classify the information needed to check the patient's health condition. In this experiment, we obtain the robust corner coordinates of the display from the image frame of the medical display monitoring system and extract the data by performing the perspective transformation based on the corner coordinates. Figure 3 visually shows the process of acquiring these robust corner coordinates. In the corner detection, the regression model is used to predict four corner coordinate approximations in the monitor display image frame, and elaborate corners are extracted in the order of Corner Image Cropping, Canny Edge, Hough Transform, and Sub Pixel method. Through this process, the robust corner coordinates of the medical display were successfully obtained. Fig. 4. provides a visual representation of the Data Extraction process. To carry out one of the geometric transformations, Homography, the study used the four coordinates of a robust display corner. In addition, we localize the data using the data box coordinates for the display area converted through the geometric transformation and predict the class of each data using the classification model.

V. Conclusion and Future wokr

In this paper, we propose a system based on computer vision to automate the process of extracting data from medical monitoring equipment. In the future, we will conduct research to develop an application for running this system on smartphones and to record data, enabling rapid diagnosis of potential risks.

We hope that this research contributes to the automation of medical monitoring systems and stimulates research on various ideas.

Acknowledgment

This research was supported by the MSIT(Ministry of Science and ICT), Korea, under the Innovative Human Resource Development for Local Intellectualization support program(IITP-2023-RS-2022-00156389) supervised by the IITP(Institute for Information & communications Technology Planning & Evaluation) and also was supported by the Digital Innovation Hub project supervised by the Daegu Digital Innovation Promotion Agency(DIP) grant funded by the Korea government(MSIT and Daegu Metropolitan City) in 2023(No. DBSD1-04, Smart management system for preventing to lonely deaths of elderly people living alone based on automatic meter reading information and CCTV access information).
※ MSIT: Ministry of Science and ICT

References

[1] Zainuddin, A.A., et al., Patient Monitoring System using Computer Vision for Emotional Recognition and Vital Signs Detection. 2020 18th Ieee Student Conference on Research and Development (Scored), 2020: p. 22-27.

[2] Takla, G., et al., The problem of artifacts in patient monitor data during surgery: a clinical and methodological review. Anesthesia & Analgesia, 2006. 103(5): p. 1196-1204.

[3] Baumgartner, B., K. Rödel, and A. Knoll. A data mining approach to reduce the false alarm rate of patient monitors. in 2012 Annual International Conference of the IEEE Engineering in Medicine and Biology Society. 2012. IEEE.

[4] Rajan Jeyaraj, P. and E.R.S. Nadar, Smart-monitor: Patient monitoring system for IoT-based healthcare system using deep learning. IETE Journal of Research, 2022. 68(2): p. 1435-1442.

[5] Hathaliya, J., et al. Blockchain-based remote patient monitoring in healthcare 4.0. in 2019 IEEE 9th international conference on advanced computing (IACC). 2019. IEEE.

[6] Cheikhrouhou, O., et al., A lightweight blockchain and fog-enabled secure remote patient monitoring system. Internet of Things, 2023. 22: p. 100691.

[7] Várady, P., Z. Benyó, and B. Benyó, An open architecture patient monitoring system using standard technologies. IEEE transactions on information technology in biomedicine, 2002. 6(1): p. 95-98.

[8] Kittipanya-Ngam, P., O.S. Guat, and E.H. Lung. Computer vision applications for patients monitoring system. in 2012 15th International Conference on Information Fusion. 2012. IEEE.

[9] Inoue, M., R. Taguchi, and T. Umezaki. Vision-based bed detection for hospital patient monitoring system. in 2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC). 2018. IEEE.

[10] Canny, J., A computational approach to edge detection. IEEE Transactions on pattern analysis and machine intelligence, 1986(6): p. 679-698.

[11] Farag, W., Real-time detection of road lane-lines for autonomous driving. Recent Advances in Computer Science and Communications (Formerly: Recent Patents on Computer Science), 2020. 13(2): p. 265-274.

[12] Zhao-Yi, P., Z. Yan-Hui, and Z. Yu. Real-time facial expression recognition based on adaptive canny operator edge detection. in 2010 Second International Conference on Multimedia and Information Technology. 2010. IEEE.

[13] Muthukrishnan, R. and M. Radha, Edge detection techniques for image segmentation. International Journal of Computer Science & Information Technology, 2011. 3(6): p. 259.

[14] Gao, T. and Z. Yang. 3D object recognition method based on improved canny edge detection algorithm in augmented reality. in 2020 IEEE 5th International Conference on Image, Vision and Computing (ICIVC). 2020. IEEE.

[15] Hough, P.V., Method and means for recognizing complex patterns. 1962, Google Patents.

[16] Zheng, F., et al., Improved lane line detection algorithm based on Hough transform. Pattern Recognition and Image Analysis, 2018. 28: p. 254-260.

[17] Philip, K.P., et al., The fuzzy hough transform-feature extraction in medical images. IEEE Transactions on Medical Imaging, 1994. 13(2): p. 235-240.

[18] Du, S., et al., Straight-Line Detection Within 1 Millisecond Per Frame for Ultrahigh-Speed Industrial Automation. IEEE transactions on industrial informatics, 2022. 19(4): p. 5965-5975.

Proceedings

International Conference on Culture Technology(ICCT) 2023

Published on 1-4 December 2023

Published by International Association for Convergence Science and Technology

Editor in Chief Taesoo Yun, Ph.D.

Publisher Pyeoungkee Kim, Ph.D.

Room# 536, International Education Building, Silla University,
700-Bunkil 140, BaekyangDaero, Busan 46958, Korea

Tel : +82-51-999-5066