



A Retrieval-Augmented Architecture for Engineering Knowledge and CAD Models: An Industrial Case Study

Sueon Kim¹⁾, Kyong Hoon Kim¹⁾, Yundo Choi¹⁾, Taekyeong Lee¹⁾ and Chan-il Park²⁾

¹⁾School of Computer Science and Engineering, Kyungpook National University, South Korea

²⁾Daedong Industrial, South Korea

Motivation

Why CAD-aware retrieval?

- Engineering design reuse requires access to both **design documents and 3D CAD models**.
- Text-based retrieval cannot directly capture geometric features such as **hole arrangements and cavity structures**.
- Geometry-based CAD retrieval complements document-based retrieval by considering **structural differences between parts**.

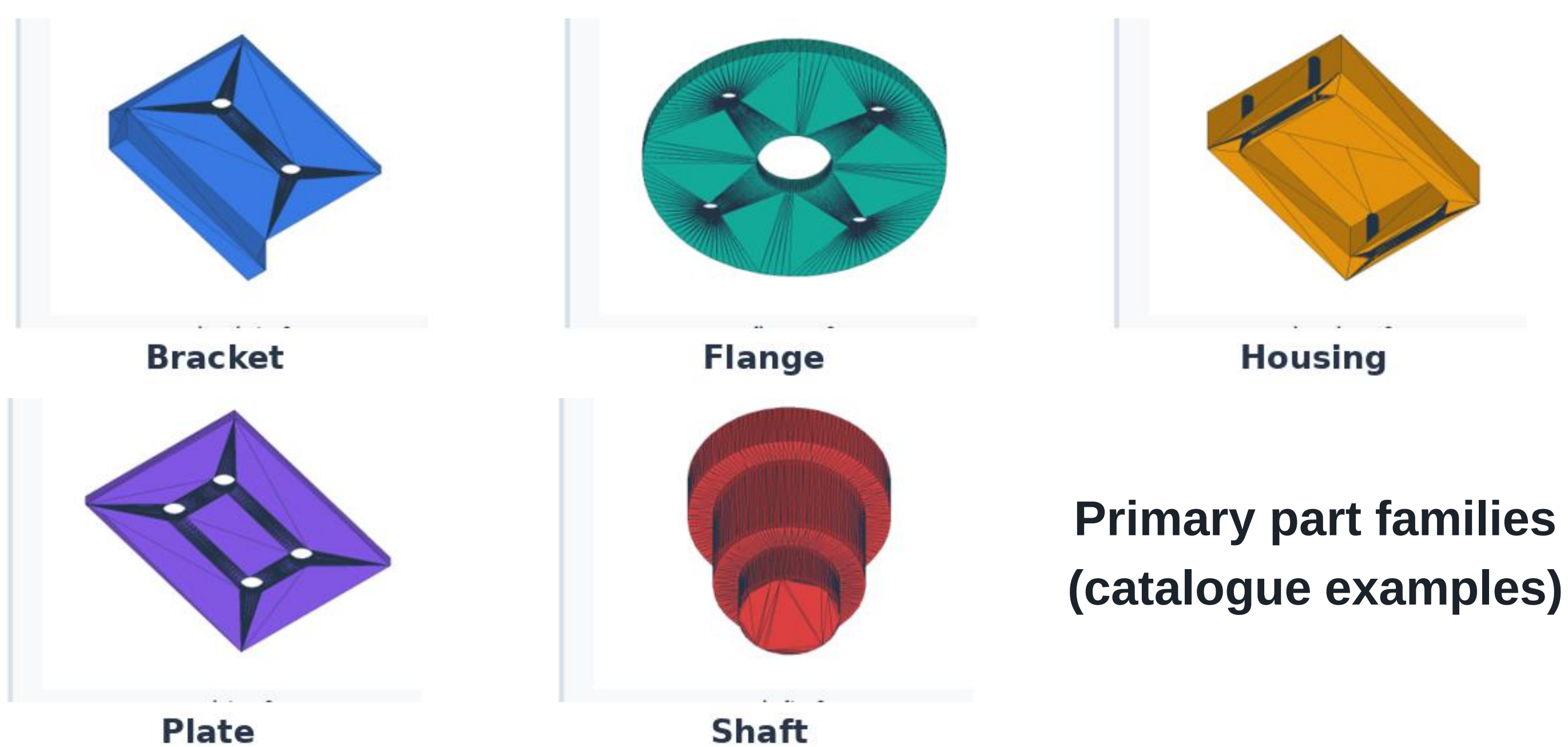
Main

Baseline | 8 geometric features

- Each CAD model is represented by **8 features**: volume, surface area, three bounding-box dimensions, and face/edge/vertex counts.
- Volume, area, and dimensions are normalized as V/L^3 , A/L^2 , and D/L , respectively, where L is the longest bounding-box dimension. Topological counts use $\log(1 + \text{count})$.

UV-Net | learned B-Rep embedding

- UV-Net learns geometric representations from **B-Rep face and edge information** using self-supervised contrastive learning.
- Trained on **24,203 training and 6,053 validation STEP models** from the Fusion 360 Gallery Segmentation Dataset for **100 epochs**.
- The selected model produces **128-dimensional embeddings**, using L2 normalization and cosine distance for retrieval.



Case Study

Primary set | broad family matching

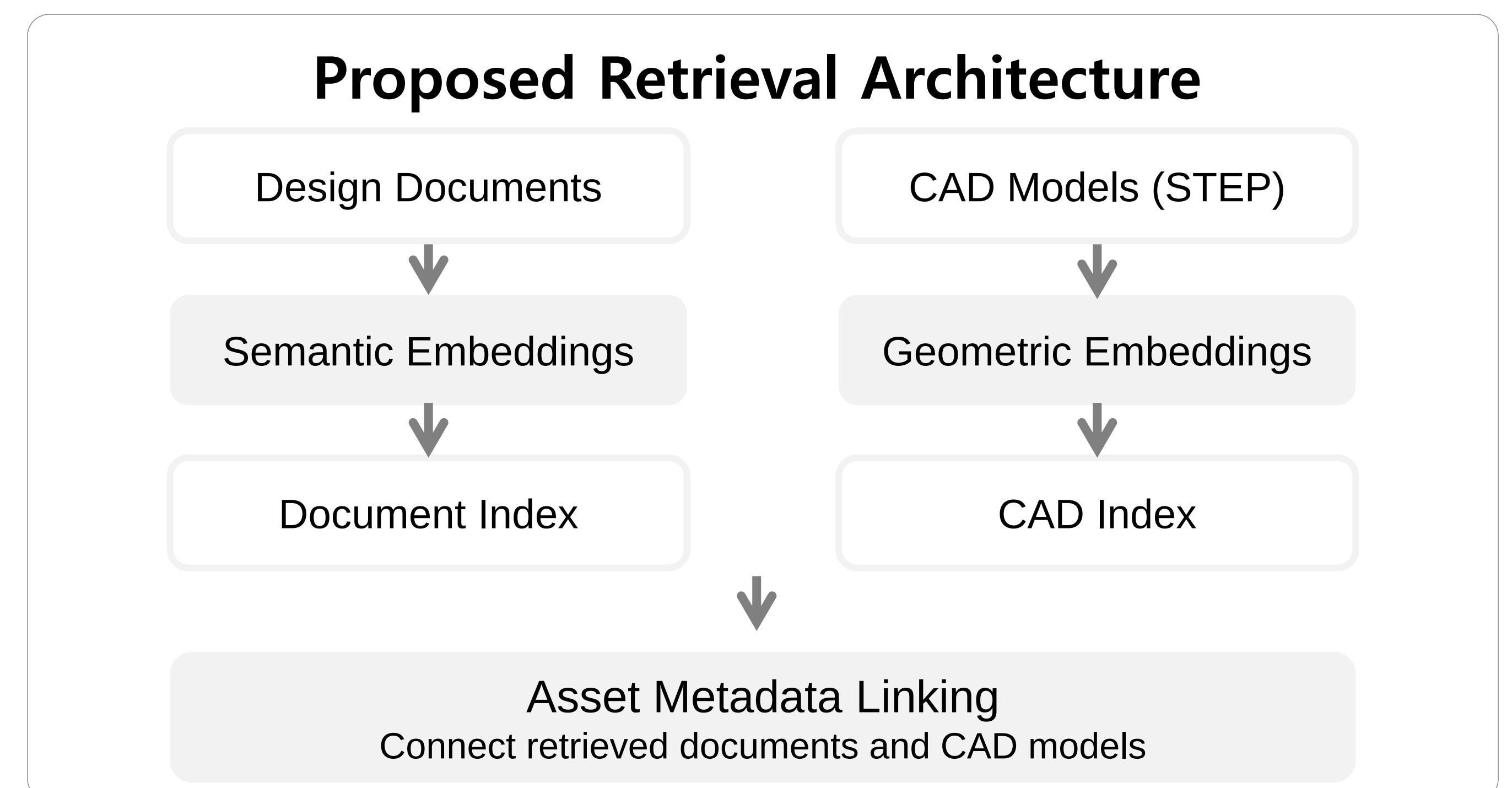
- **30 CAD parts**: 5 families × 6 variants.
- The other five parts in the same family are considered relevant, despite dimensional and local feature variations.

Retrieval results (mean Precision@K, %)

Set / method	P@1	P@3	P@5
Primary / Baseline	100.00	100.00	100.00
Primary / UV-Net	96.67	96.67	95.33
Stress / Baseline	11.11	25.93	33.33
Stress / UV-Net	98.15	91.98	85.93

From STEP to B-Rep

- **B-Rep** represents CAD solids using faces, edges, vertices, and their connectivity.
- **UV-Net** learns geometric features from B-Rep face and edge information.



Controlled CAD evaluation

- Evaluation uses **controlled single-solid STEP models**, separate from UV-Net training data.
- Each model serves as a query (excluding itself), with labels used **only for evaluation**, not training or model selection.

Conclusion

Stress set | hard negatives

- **54 CAD parts**: 3 families × 3 structural layouts × 6 variants.
- Parts are relevant only if they share both the same family and structural layout. Other layouts within the same family are hard negatives.



Separation margin

$$m(q) = \min_{i \in I} d(q, i) - \max_{r \in R} d(q, r)$$

A negative margin indicates overlap between relevant and irrelevant parts; **48 of 54 baseline stress queries** had negative margins (median: -0.53875).

UV-Net better distinguishes structural differences, but is not universally superior. Future work will validate real Siemens NX models and extend retrieval to assemblies and revision-aware documents.