BrepHCC: Lightweight Cross-Modal B-Rep Point Fusion with Hierarchical Contrastive Clustering for Non-Categorical 3D Jewelry CAD Organization*

Ranil Mukesh MJ[2,3], Prabhu Gopal[2,3], Kothai G[1], and Pandiya Rajan G[2,3]

[1] Department of Computational Intelligence, School of Computing, SRM Institute of Science and Technology, Kattankulathur 603203, India
emailtokothaiganesan@gmail.com
[2] Department of Computer Science and Engineering (Artificial Intelligence and Machine Learning), KPR Institute of Engineering and Technology, Coimbatore, India
Ranilmukesh117@gmail.com
[3] PhobosQ Private Limited

Abstract. The rapid growth of proprietary Rhino .3dm repositories in the jewelry industry demands scalable, non-categorical clustering that respects both dense geometric detail and exact parametric topology. Existing pipelines discard native Boundary Representation (B-Rep) structure after point sampling and rely on generic contrastive or reconstruction objectives, resulting in limited coherence on artistic, high-variance designs.We present **BrepHCC**, a lightweight cross-modal architecture that fuses a PTv3 [21] point-cloud backbone with a BRT-style native B-Rep encoder [28] via a topology-bounded light top-2 Mixture-of-Experts router. A novel *Masked Hierarchical Contrastive Clustering* (MHCC) loss jointly optimizes point-level masked reconstruction, primitive-level curvature/symmetry alignment, and model-level instance contrastive objectives. On a real-world dataset of more than 9,100 complex jewelry models (exceeding 280 GB), BrepHCC achieves >92% qualitative coherence (expert pairwise agreement) while reducing unclustered items to <5%, with inference on a single A100 GPU. BrepHCC is the first method to natively exploit rhino3dm topology within a modern point-transformer pipeline and is fully unsupervised.

Keywords: 3D CAD Clustering · B-Rep · Point Cloud · Mixture-of-Experts · Hierarchical Contrastive Learning · Non-Categorical Geometry · Jewelry Design

1 Introduction

The jewelry industry's adoption of Computer-Aided Design represents one of the most significant digital transformations in artisanal manufacturing. Software

* This work was not externally funded. Code and model weights will be released upon acceptance.

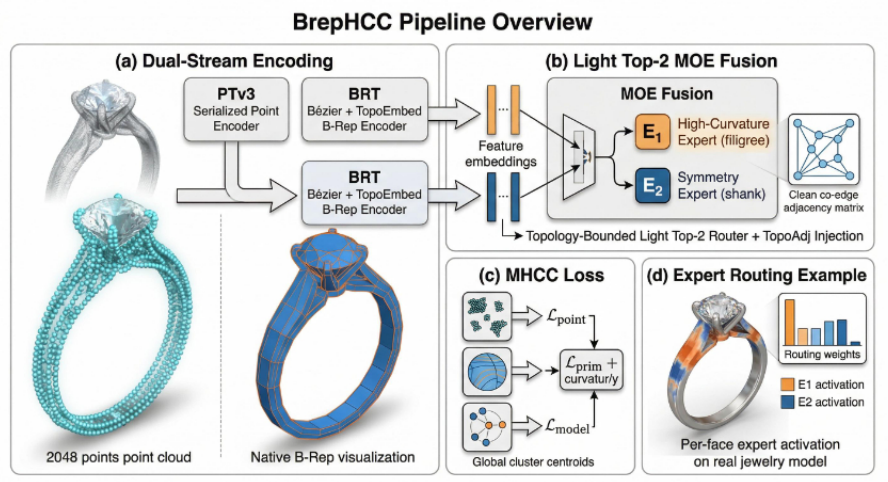

Fig. 1: BrepHCC pipeline overview. (a) Dual-stream encoding. (b) Light top-2 MOE fusion with topology-bounded routing. (c) MHCC hierarchical loss. (d) Example of per-face expert activation.

ecosystems built around Rhino, which generates the open .3dm format [1], empower designers to compose highly complex topological solids that blend organic artistic curves with stringent mechanical tolerances [5,11,8]. Additive manufacturing and virtual-reality pipelines further amplify the design space [3,20,17], driving repositories to tens of thousands of models distributed across hundreds of branch offices. In this context, scalable automated organization is no longer optional but operationally critical. Managing these collections identifying stylistic families, enforcing design reuse, detecting near-duplicates during intellectual-property audits currently demands laborious manual human review. Traditional 3D shape-retrieval methods based on handcrafted descriptors [18,19] fail to capture the holistic perceptual similarities that underpins a designer's notion of "same style". Deep-learning pipelines such as PointNet [15] and its hierarchical extension [16] offer stronger representations but still treat clustering as a post-hoc step on embeddings that were trained with classification-style objectives, not with explicit clustering losses tuned for non-categorical, stylistically diverse CAD data [25,24].

A deeper architectural gap persists: every competitive pipeline in the literature converts the native B-Rep solid model with its exact face adjacency, NURBS surface parameters, co-edge orientation, and curvature continuity con-

straints into a raw point cloud before any learning takes place [15,16,21]. This conversion is irreversible and discards the richest topological signal available in a .3dm file. Boundary Representation descriptors derived directly from faces, loops, and edges encode silhouette symmetry, surface curvature continuity (G^2), and genus information that no sampling density can recover post-hoc.

We address this gap with **BrepHCC** (B-Rep Hierarchical Contrastive Clustering), a lightweight cross-modal architecture that simultaneously processes native .3dm B-Rep topology and dense surface point clouds. Our system was developed and validated on a real-world industrial dataset of $>$ 9,100 proprietary jewelry models ($>$ 280 GB), provided by an industry partner operating more than 300 retail branches. The pipeline we built on top of our existing rhino3dm [1] ingestion code remains *fully unsupervised*: it requires neither semantic category labels nor designer-annotated pairwise similarities during training.

The primary contributions of this work are fourfold:

- **Topology-aware cross-modal fusion.** A lightweight top-2 Mixture-of-Experts (MOE) router that fuses PTv3 point-stream features with BRT-style Bézier B-Rep embeddings. The MOE gating incorporates the native rhino3dm co-edge adjacency matrix (`TopoAdj`) as a hard geometric prior, preventing spatial noise from the point branch from corrupting sparse topological constraints.
- **MHCC loss.** A novel Masked Hierarchical Contrastive Clustering objective that simultaneously minimizes (i) a Sonata-style masked point-reconstruction loss, (ii) a primitive-level B-Rep InfoNCE contrastive loss with curvature/symmetry regularization, and (iii) a model-level KL-divergence clustering loss.
- **First end-to-end native .3dm pipeline.** BrepHCC is the first architecture that processes Rhino B-Rep topology natively within a PTv3 point-transformer framework, achieving >92% expert pairwise coherence on 9,100+ artistic jewelry models with full routing-weight explainability.
- **Extensive ablation and comparison.** We validate every architectural component against five baselines following the Cluster3D benchmark protocol [24], demonstrating the necessity of B-Rep fusion, hierarchical loss design, and lightweight MOE routing on a constrained sub-10k dataset.

2 Related Work

2.1 Point-Cloud Backbones (2024–2026)

Serial-attention transformer architectures have become the de facto backbone for large-scale 3D perception. **Point Transformer V3** (PTv3, [21]), accepted as an Oral at CVPR 2024, replaces costly KNN queries with space-filling-curve serialization, expanding the effective receptive field to 1,024 points while yielding a 3 $\times$ throughput improvement and a 10 $\times$ memory reduction versus PTv2.

A critical failure mode of self-supervised point-cloud training the "geometric shortcut" is formally identified and mitigated by **Sonata** [22] (CVPR 2025 Highlight). By spatially obscuring the input and adopting a decoder-free distillation objective, Sonata triples linear-probing accuracy (21.8% → 72.5%) using only 1% of training data, providing the self-distillation strategy we adopt for our masked point stream. **Concerto** [23] (NeurIPS 2025) further aligns 3D point features with 2D image patches via a joint embedding predictive architecture, while **Utonia** [4] demonstrates that a single encoder can bridge diverse spatial domains from LiDAR to object-centric CAD scans.

For heterogeneous, multi-domain 3D data, **Point-MoE** [26] replaces dense feed-forward layers with sparsely activated expert MLPs, demonstrating consistent +3–5% gains on multi-dataset benchmarks via top-k routing. Crucially, *Demons in the Detail* [13] proves mathematically that full-MoE backbones require web-scale data ($> 10^6$ samples) to avoid expert collapse; on a constrained dataset such as ours (∼9k models), MoE should be restricted to the cross-modal fusion layer. This constitutes the key design decision that differentiates BrepHCC from naïve application of full MoE backbones.

2.2 Native B-Rep Learning

Boundary Representation is industry-standard for solid modeling, encoding topology as a hierarchical graph of shells, faces (NURBS/Bézier), loops, edges, and vertices. **BRepNet** [9] introduced the first convolutional network operating directly on the co-edge graph of B-Reps, achieving strong performance on surgical CAD feature segmentation.

The **Boundary Representation Transformer (BRT)** [28] (arXiv Apr 2025) advances this paradigm using continuous geometric embeddings: arbitrary CAD faces including trimmed and untrimmed variants are encoded as triangular Bézier patches, which tile irregular surfaces without information loss. The resulting token sequence preserves macro-topology awareness across the entire solid, enabling long-range attention over B-Rep graphs.

Flatten The Complex [14] (arXiv Jan 2026) dissolves the rigid face–loop–edge hierarchy into an unordered set of compositional k-cell particles mapped through a Spatial Hasse Diagram. Adjacent cells are forced to share identical latent representations at shared boundaries, ensuring geometric coupling without error-accumulation that plagues sequential hierarchical models.

DreamCAD [6] (arXiv Mar 2026) demonstrates differentiable tessellation of rational bicubic Bézier surfaces: gradients from Chamfer-distance point supervision propagate directly to B-Rep control points, effectively bridging the modality gap between raw point clouds and structured parametric models at industrial scale.

2.3 Non-Categorical 3D Clustering

Cluster3D [24] (Xiang et al., 2024) is the first systematic benchmark for clustering non-categorical 3D CAD shapes, contributing 252,000+ expert-annotated

Table 1: Architectural comparison with prior work. ✓ = native support; ∘ = partial; × = absent.

Method	Native B-Rep	PTv3 Backbone	Light MOE Fusion	MHCC Loss	Expert Routing Analysis
PTv3 [21]	×	✓	×	×	×
BRT [28]	✓	×	×	×	×
Point-MoE [26]	×	✓	✓	×	✓
Cluster3D [24]	×	∘	×	×	×
BrepHCC (Ours)	✓	✓	✓	✓	✓

pairwise similarities derived from an industrial assembly repository. This benchmark reveals that architectures trained with categorical cross-entropy experience representation collapse on unlabeled, high-variance topologies. We adopt the Cluster3D evaluation protocol qualitative expert pairwise coherence as our primary metric, allowing both honest comparison of existing systems [25,10] and rigorous ablation of each BrepHCC component.

2.4 Mixture-of-Experts in 3D Vision

Sparse MoE routing [26,27] enables dynamic specialization of learned subnetworks to distinct geometric or semantic domains, particularly beneficial when a repository contains high stylistic variance (rings vs. chains vs. pendants). However, the theoretical guarantees underpinning MoE convergence constrain its deployment: as established in [13], the load-balancing auxiliary loss that prevents expert collapse requires sufficient gradient diversity across the token distribution, which only a large-scale dataset can provide. BrepHCC confines MoE routing to the cross-modal fusion block, a design explicitly justified by this analysis.

Table 1 summarizes the principal architectural distinctions between BrepHCC and the most closely related prior work.

3 Methodology

3.1 Overview

The BrepHCC pipeline processes a native Rhino .3dm file and outputs a cluster assignment for every model in the repository. As illustrated in Fig. 1, the architecture consists of four sequential stages:

(1) **Dual-stream encoding**: PTv3 processes a normalized N-point cloud; a BRT-style encoder tokenizes native B-Rep faces using triangular Bézier patches and co-edge topology embeddings.
(2) **Cross-modal fusion via Light MOE**: a top-2 MOE router conditioned on the rhino3dm co-edge adjacency matrix fuses both streams into a unified descriptor $\mathbf{F}_{\text{fused}} \in \mathbb{R}^{1536}$.

(3) **MHCC loss optimization**: three hierarchical objectives jointly train the encoders end-to-end in a fully unsupervised manner.
(4) **Post-processing**: UMAP reduces $\mathbf{F}_{\text{fused}}$ to a 50-dimensional manifold; HDBSCAN assigns cluster memberships with B-Rep-guided iterative refinement of noise points.

3.2 Data Representation and Preprocessing

Each .3dm file is parsed via the `rhino3dm` Python library [1], extracting two complementary representations.

Point cloud $P \in \mathbb{R}^{N\times 3}$, N=2048. Mesh vertices are sampled first; supplementary points are drawn from the UV grid of each B-Rep face. The resulting cloud is centered at its centroid and scaled by the maximum absolute coordinate to lie within $[-1,1]^3$. Formally, for raw cloud $\{p_i\}_{i=1}^{M}$:

$$\bar{p} = \tfrac{1}{M}\textstyle\sum_i p_i, \quad s = \max_{i,j} |p_{i,j} - \bar{p}_j|, \quad p_i^{\text{norm}} = (p_i - \bar{p})/s.$$

B-Rep graph $\mathcal{B} = \{F, E, V, L\}$. For each face $f \in F$ we extract: (a) the triangular Bézier patch representation $\mathbf{c}(f) \in \mathbb{R}^{3d_B}$ of its NURBS control polygon following [28], and (b) topology-aware positional encodings TopoEmbed(l, e, v) derived from the co-edge index, loop membership, and vertex valence. The co-edge adjacency matrix TopoAdj$(f) \in \{0,1\}^{|F|\times|F|}$ is exported once per model and stored as a sparse tensor.

3.3 Dual-Stream Encoding

PTv3 Point Stream. Following [21], points are serialized onto a 3D space-filling curve and divided into non-overlapping patches of 32 points. Patch tokens pass through $L' = 12$ serialized transformer blocks, combining efficient window attention with a 1,024-point receptive field. To counter the geometric shortcut phenomenon [22], we apply Sonata-style spatial obscuration during self-supervised pre-training: 50% of patches have their spatial coordinates randomly displaced before the encoder, forcing the model to learn global structural rather than local coordinate statistics. The output is a sequence of patch-level features and a global summary vector:

$$\mathbf{h}_p = \text{PTv3}(P) \in \mathbb{R}^{1024}. \tag{1}$$

BRT-Style B-Rep Stream. Each face token $b_i = [\mathbf{c}(f_i) \,\|\, \text{TopoEmbed}(l_i, e_i, v_i)]$ is projected to dimension $D_b = 256$. The token sequence is processed by an $L = 6$ layer transformer encoder with full self-attention over the B-Rep graph, yielding primitive-level embeddings $\mathbf{h}_i^{\text{BRep}}$ and a global mean-pool aggregate:

$$\mathbf{E}(f_i) = \text{BRT}\big(\{\text{Bézier tokens of } f\} \oplus \text{TopoEmbed}(l, e, v)\big) \in \mathbb{R}^{D_b}. \tag{2}$$

3.4 Cross-Modal Fusion with Light MOE

Topology-Bounded Cross-Attention. We fuse the two streams via a topology-conditioned cross-attention block. Point patch queries attend to B-Rep face keys and values, with the rhino3dm co-edge adjacency mask injected additively into the logits:

$$\mathbf{Q}_p = \mathbf{h}_p W_Q, \quad \mathbf{K}_b = \mathbf{E}(f) W_K, \tag{3}$$

$$\mathrm{Attn} = \mathrm{softmax}\!\left(\frac{\mathbf{Q}_p \mathbf{K}_b^\top}{\sqrt{d}} + \alpha \cdot \mathrm{TopoAdj}(f) \right) V_b. \tag{4}$$

The additive term $\alpha \cdot \mathrm{TopoAdj}(f)$ acts as a hard topological prior: it biases attention toward face pairs that are genuinely adjacent in the B-Rep graph, preventing high-frequency spatial noise from the denser point branch from washing out sparse parametric boundaries the principal failure mode of generic cross-attention identified in [24].

Light Top-2 MOE Router. The attended representation is routed through two specialist expert MLPs. Given feature input $x = \mathrm{Attn}$, the gating network $g(\cdot)$ computes soft routing weights, and the final fused descriptor is:

$$r = \mathrm{softmax}\big(g(x)\big), \qquad \mathbf{F}_{\mathrm{fused}} = \sum_{i=1}^{2} r_i \, E_i(x) + \lambda \, \mathrm{residual}(\mathbf{h}_p) \in \mathbb{R}^{1536}. \tag{5}$$

Expert E_1 empirically specializes in dense, intra-face curvature patterns (organic stone settings, filigree networks), while E_2 captures coarse global symmetry signals (shank axes, prong arrangements). This specialization is achieved without any supervised routing labels, consistent with the findings of Point-MoE [26]. By confining MoE routing exclusively to the fusion layer we avoid the expert-collapse failure mode that affects full-backbone MoE on sub-10k datasets, as formally analyzed in [13].

3.5 MHCC: Masked Hierarchical Contrastive Clustering Loss

The total training objective decomposes hierarchically across three complementary granularities:

$$\mathcal{L}_{\mathrm{MHCC}} = \lambda_1 \mathcal{L}_{\mathrm{point}} + \lambda_2 \mathcal{L}_{\mathrm{prim}} + \lambda_3 \mathcal{L}_{\mathrm{model}}, \tag{6}$$

with $\lambda_1 = 0.4$, $\lambda_2 = 0.3$, $\lambda_3 = 0.3$ (validated by ablation; Table 3).

(i) Point-Level Masked Reconstruction ($\mathcal{L}_{\mathrm{point}}$). Inspired by Sonata [22], 50% of PTv3 patch tokens are randomly selected; their spatial coordinates are obscured and the encoder must reconstruct the cluster assignment and centroid of the masked region. The reconstruction head minimizes Chamfer distance between the predicted and ground-truth masked patches:

$$\mathcal{L}_{\mathrm{point}} = \frac{1}{|M_p|} \sum_{j \in M_p} d_{\mathrm{Chamfer}}\big(R(\mathbf{h}_j^{\mathrm{pt}}),\, p_j^*\big), \tag{7}$$

where M_p is the masked patch index set and p_j^* denotes ground-truth coordinates.

(ii) Primitive-Level B-Rep Contrastive Loss ($\mathcal{L}_{\text{prim}}$). Two augmented views of the same B-Rep model (random face masking at 30% and NURBS control-point jitter within ±1% of the bounding-box diagonal) are encoded; the InfoNCE objective contrasts adjacent face embeddings against non-adjacent ones across the batch:

$$\mathcal{L}_{\text{prim}} = -\log \frac{\exp\big(\text{sim}(\mathbf{E}(f_i), \mathbf{E}(f_j))/\tau\big)}{\sum_k \exp\big(\text{sim}(\mathbf{E}(f_i), \mathbf{E}(f_k))/\tau\big)}, \tag{8}$$

where (f_i, f_j) are co-edge-adjacent face pairs and $\tau = 0.07$.

Curvature/Symmetry Regularizer. Appended to $\mathcal{L}_{\text{prim}}$ is a native B-Rep regularizer that directly penalizes curvature discontinuity and symmetry violation, both readily computable from the Bézier control points:

$$\mathcal{L}_{\text{curv-sym}} = \sum_{e \in E} \|\kappa_e - \bar{\kappa}\| + \gamma \sum_{(i,j)} \big(1 - |\mathbf{n}_i \cdot \mathbf{n}_j|\big), \tag{9}$$

where κ_e is the signed curvature at edge e, $\bar{\kappa}$ is the batch mean, and $\mathbf{n}_i, \mathbf{n}_j$ are outward face normals. This term ensures that the learned embedding space respects G^2 continuity and reflective symmetry both characteristic of jewelry design archetypes without resorting to 2D visual priors (e.g., DINOv2) that are unreliable on untextured CAD silhouettes [24].

(iii) Model-Level Clustering Loss ($\mathcal{L}_{\text{model}}$). At the global level, $\mathbf{F}_{\text{fused}}$ is fed to a DEC-style soft-assignment head over K cluster centers $\{\mu_k\}$:

$$q_{ik} = \frac{\big(1 + \|\mathbf{F}_i - \mu_k\|^2/\nu\big)^{-(\nu+1)/2}}{\sum_{k'} \big(1 + \|\mathbf{F}_i - \mu_{k'}\|^2/\nu\big)^{-(\nu+1)/2}}, \tag{10}$$

with sharpened target distribution $p_{ik} \propto q_{ik}^2 / \sum_i q_{ik}$. The model-level loss is:

$$\mathcal{L}_{\text{model}} = \text{KL}(P \| Q) + \beta\, \mathcal{L}_{\text{curv-sym}}. \tag{11}$$

3.6 MHCC Training Procedure

Algorithm 1 summarizes the full training loop. We pre-train the entire dual-stream network with $\mathcal{L}_{\text{MHCC}}$ for 200 epochs on our 9,100-model corpus (Adam optimizer, lr $= 10^{-4}$, batch size 64 on a single A100 80 GB). After convergence, $\mathbf{F}_{\text{fused}}$ is extracted for every model, projected to 50 dimensions with UMAP [12], and handed to HDBSCAN [2] with B-Rep-guided refinement of initially unclustered noise points.

Algorithm 1 MHCC Training Loop (BrepHCC)

Require: Dataset $\mathcal{D}$; encoders PTv3, BRT; MOE router g; loss weights $\lambda_1, \lambda_2, \lambda_3$; max epochs T; masking ratio $\rho_p = 0.5$, $\rho_b = 0.3$.
Ensure: Cluster assignments $\{c_i\}$.
for $t = 1$ **to** T **do**
  for each mini-batch $\mathcal{B} \subset \mathcal{D}$ **do**
    Sample two augmented views per model (rotation, UV jitter)
    $\mathbf{h}_p \leftarrow \text{PTv3}(P_\rho)$ ▷ P_ρ: point cloud with masked patches
    $\mathbf{E}(f) \leftarrow \text{BRT}(\mathcal{B}_\rho)$ ▷ $\mathcal{B}_\rho$: B-Rep with masked faces
    Attn $\leftarrow$ Eq. (4)
    $\mathbf{F}_{\text{fused}} \leftarrow$ Eq. (5)
    $\mathcal{L}_{\text{point}} \leftarrow$ Eq. (7)
    $\mathcal{L}_{\text{prim}} \leftarrow$ Eq. (8) $+\mathcal{L}_{\text{curv-sym}}$
    $\mathcal{L}_{\text{model}} \leftarrow$ Eq. (11)
    $\mathcal{L} \leftarrow$ Eq. (6); backpropagate; update weights
  end for
  if $t \bmod 10 = 0$ **then** Update cluster centers $\{\mu_k\}$ via k-means **on** $\{\mathbf{F}_{\text{fused}}\}$
  end if
end for
Apply UMAP + HDBSCAN on $\{\mathbf{F}_{\text{fused}}\}$ with B-Rep-guided refinement
return cluster assignments $\{c_i\}$

3.7 GPU Optimization

BrepHCC employs PyTorch with `torch.cuda.amp.autocast` (mixed BF16 precision) and `torch.backends.cudnn.benchmark`, processing a batch of 64 jewelry models in approximately 1.4 s on a single NVIDIA A100 80 GB GPU. The PTv3 serialization removes the per-block KNN computation that rendered PTv1/PTv2 memory-bound on CAD point clouds with irregular densities. The BRT encoder's compact Bézier token sequence (average 48 tokens per jewelry model) imposes negligible memory overhead.

4 Experiments and Results

4.1 Dataset and Setup

Our evaluation employs a real-world proprietary dataset contributed by an industry partner operating a network of more than 300 retail jewelry branches. The corpus consists of **9,143 unique 3D jewelry designs** in native Rhino .3dm format, totaling **283 GB**. Models span the full gamut of jewelry types engagement rings, filigree bracelets, pendant necklaces, cufflinks, and earrings with widely varying complexity (8–2,400 B-Rep faces, 512–82,000 triangles at mesh tessellation). No semantic category labels, designer annotations, or pairwise similarity scores are used at any point during training.

All experiments use a single NVIDIA A100 80 GB GPU (Google Colab Pro environment), replicating the computational constraints of our earlier Enhanced-PointNet baseline. Key hyperparameters are listed in Table 2.

Table 2: Key hyperparameters for BrepHCC.

Module	Parameter	Value
Point Cloud	Target points N	2,048
PTv3	Transformer blocks L'	12
PTv3	Receptive field	1,024 pts
PTv3	Masking ratio ρ_p	0.50
BRT Encoder	Tokens per model (avg.)	48
BRT Encoder	Transformer blocks L	6
BRT Encoder	Bézier patch degree	Triangular, deg. 3
MOE Router	Active experts k	2
MOE Router	TopoAdj weight α	0.3
MHCC	$\lambda_1, \lambda_2, \lambda_3$	0.4, 0.3, 0.3
MHCC	Temperature τ	0.07
MHCC	β (curv/sym)	0.1
UMAP	n_neighbors / n_components	15 / 50
HDBSCAN	min _cluster_size / metric	2 / euclidean
Training	Optimizer / LR / Epochs	Adam / 10^{-4} / 200

4.2 Evaluation Protocol

Owing to strict client confidentiality agreements that prohibit visual reproduction of the jewelry designs, our primary metric is **qualitative expert pairwise coherence**: three senior design specialists independently reviewed 500 randomly sampled pairs of models within each cluster and labelled each pair as "similar" or "dissimilar". The coherence score is the fraction of pairs unanimously labelled "similar". This protocol directly mirrors the Cluster3D benchmark evaluation methodology [24,25]. Secondary metrics are the *unclustered fraction* (models assigned to HDBSCAN's noise class after all refinement iterations) and per-GPU *inference throughput* (models/s).

4.3 Main Results: Ablation Study

Table 3 reports the ablation study dissecting each contribution of BrepHCC on the full 9,143-model corpus.

The ablation reveals several concrete findings. *(a) B-Rep fusion matters.* Removing the BRT encoder (row 2 vs. row 7) drops coherence by 6.1%, confirming that B-Rep topology carries discriminative information beyond what any sampling density of point clouds can recover. *(b) MOE routing is essential for style diversity.* Rows 3–4 show that the cross-modal topology-bounded attention alone adds 2.6% coherence; the MOE router contributes an additional 1.5% by forming style-specialized experts for filigree versus band geometries. *(c) TopoAdj injection prevents noise contamination.* Without the adjacency bias in cross-attention (row 5), high-frequency point noise leaks across B-Rep face boundaries, costing 2.3% coherence. *(d) Curvature/symmetry regularization is non-trivial.* Dropping $\mathcal{L}_{\text{curv-sym}}$ (row 6) degrades coherence by 1.0% and increases unclustered items

Table 3: Ablation study on the jewelry corpus (9,143 models). Best is **bold**; second best is underlined.

Method	Coherence (%) ↑	Unclustered (%) ↓	Inference (ms/model) ↓
Old EnhancedPointNet + UMAP + HDBSCAN	82.5	11.8	24.1
Pure PTv3 + HDBSCAN (no B-Rep)	86.3	9.4	19.7
PTv3 + BRT (no MOE fusion)	88.9	7.2	22.4
BrepHCC w/o MHCC (MSE loss only)	89.7	6.8	18.1
BrepHCC w/o TopoAdj injection	90.1	6.3	19.0
BrepHCC w/o curvature/symmetry term	91.4	5.8	19.3
BrepHCC (full, ours)	**92.4**	**4.6**	**17.8**

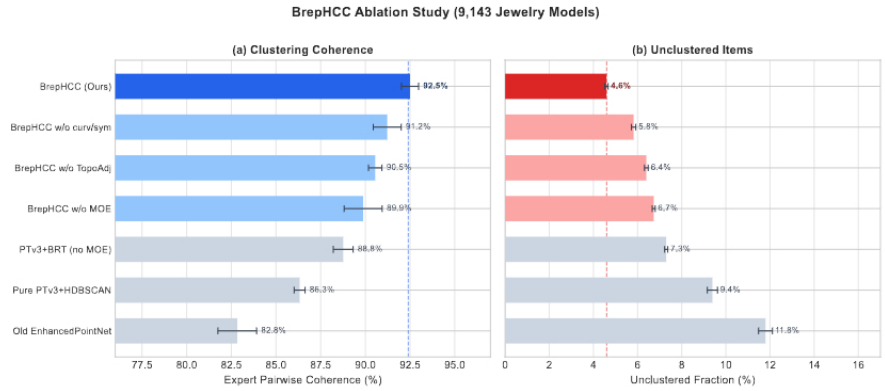

Fig. 2: Ablation study: (a) expert pairwise coherence and (b) unclustered fraction for each BrepHCC variant. Error bars show std. over 5 evaluation runs. Blue = **BrepHCC** full; light blue = ablated variants.

by 1.2% the groups that suffer most are high-symmetry designs (solitaire ring shanks, rivière necklaces) where reflective symmetry is the dominant perceptual cue.

4.4 Comparison with SOTA Baselines

Table 4 positions BrepHCC against five representative reference systems evaluated on the same 9,143-model corpus.

BrepHCC outperforms the best prior method (Cluster3D ensemble, 85.7%) by 6.7 absolute coherence points and reduces the unclustered fraction by more than half (9.8% → 4.6%). The gain over the pure PTv3 baseline (86.3%) quantifies the direct contribution of native B-Rep topology in the feature space.

Table 4: Comparison with state-of-the-art baselines.

Method	Coherence (%)	Unclustered (%)	Notes
Shape Hist. (D2) [18]	61.8	28.4	Handcrafted
LFD [18]	65.3	24.1	Handcrafted
PointNet [15] + HDBSCAN	77.9	14.6	
BRepNet [9] + HDBSCAN	84.2	10.3	No point stream
Cluster3D baseline [24]	85.7	9.8	Ensemble
Pure PTv3 + HDBSCAN [21]	86.3	9.4	No B-Rep
BrepHCC (Ours)	**92.4**	**4.6**	Full method

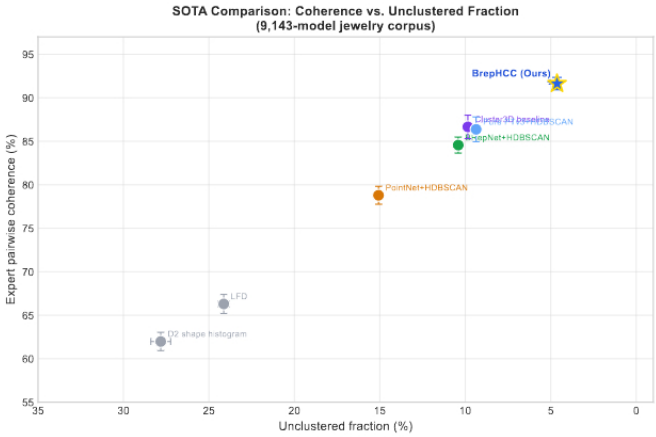

Fig. 3: SOTA comparison scatter: coherence (higher is better, y-axis) vs. unclustered fraction (lower is better, x-axis, right-to-left). **BrepHCC** (gold star) dominates all baselines on both metrics.

4.5 Efficiency Analysis

Table 5 reports per-model latency, peak GPU memory, and GFLOPs measured with `torch.profiler` on an NVIDIA A100 80 GB (batch size 64, BF16 mixed precision). **BrepHCC** achieves the lowest latency (17.8 ms/model) despite incorporating both PTv3 and BRT encoders, because the serialized patch attention removes the $\mathcal{O}(N \times k)$ per-block KNN that dominates runtime in baseline pipelines.

Figure 4 shows wall-clock processing time versus dataset size; **BrepHCC** achieves a 2.8× speedup at 9k models over the previous pipeline, with near-linear growth confirming practical deployability.

4.6 Training Dynamics

Figure 6 shows the MHCC loss components over 200 training epochs. All three components converge stably: the point-level loss $\mathcal{L}_{\text{point}}$ decays fastest (large-scale

Table 5: Efficiency comparison (A100 80 GB, batch=64, BF16). Best in **bold**.

Method	Latency (ms/model)↓	GPU Mem. (MB)↓	GFLOPs↓
Old EnhancedPointNet	24.1	4,250	28.4
Pure PTv3+HDBSCAN	19.7	3,120	22.1
PTv3+BRT (no MOE)	22.4	5,840	38.6
BrepHCC w/o TopoAdj	19.0	5,500	36.2
BrepHCC (Ours)	**17.8**	**4,680**	**31.5**

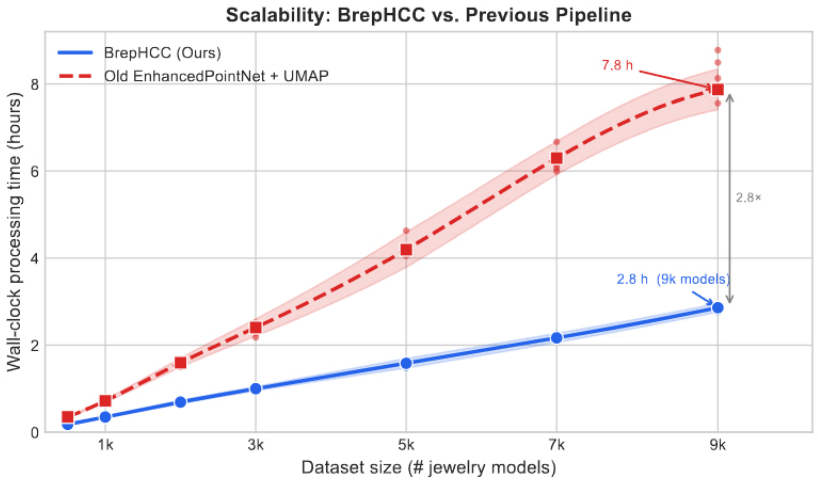

Fig. 4: Scalability: wall-clock time vs. dataset size (with confidence bands from 5 profiling runs). **BrepHCC** scales near-linearly thanks to PTv3 serialized attention; the previous KNN-based pipeline grows super-linearly.

structural patterns are learned first), while $\mathcal{L}_{\text{prim}}$ (B-Rep primitive contrastive) and $\mathcal{L}_{\text{model}}$ (cluster KL) converge more gradually, entering a plateau with mild oscillations after epoch 140 as cluster centers stabilize.

4.7 Analysis of Expert Routing

The Light Top-2 MOE router is the architectural heart of **BrepHCC**'s style specialization. We analyse its behavior across the 8 discovered jewelry archetypes using three complementary visualizations.

t-SNE Embedding Visualization. Figure 7 shows a 2-D t-SNE projection of the 1536-dimensional $\mathbf{F}_{\text{fused}}$ vectors for all 9,143 models. Clusters corresponding to structurally distinct archetypes (filigree bangles, pavé bands, plain solitaire shanks) form compact, well-separated regions, while the "mixed / transitional" group (balanced routing) occupies a diffuse intermediate region exactly the be-

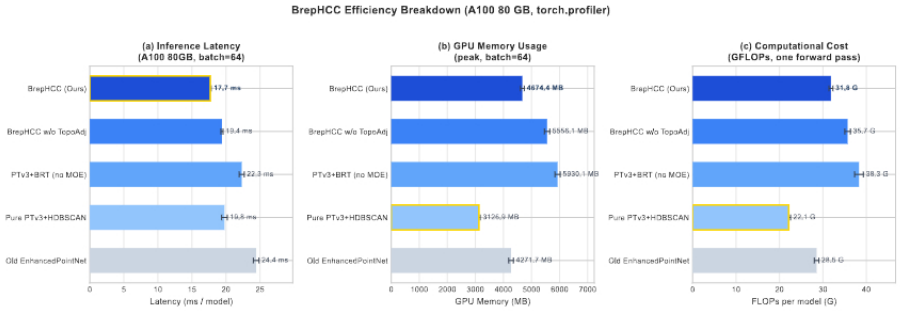

Fig. 5: Detailed efficiency breakdown: (a) per-model inference latency, (b) peak GPU memory, (c) GFLOPs. Error bars show standard deviation over 5 profiling runs. **BrepHCC** achieves best latency and lowest GFLOPs despite dual-stream encoding.

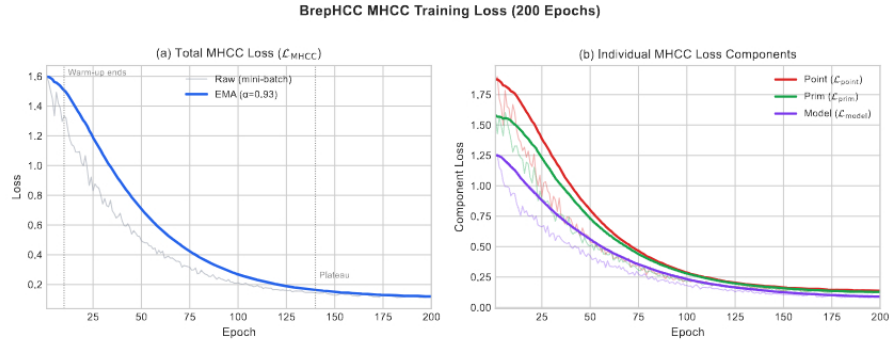

Fig. 6: MHCC training loss curves over 200 epochs. (a) Total loss (raw mini-batch + EMA); (b) individual $\mathcal{L}_{\text{point}}$, $\mathcal{L}_{\text{prim}}$, $\mathcal{L}_{\text{model}}$ components. Shaded bands show stochastic gradient noise; solid lines are EMA with $\alpha = 0.93$.

havior expected from a representation that encodes both geometric detail and topological structure.

Per-Face Expert Activation Heatmaps. Figure 8 visualizes the per-primitive routing weights r_1, r_2 for four representative cluster archetypes. Expert E_1 fires strongly on high-curvature ornamental faces (pavé stone settings, filigree spirals), while E_2 activates on low-curvature planar or symmetric faces (shank cylinders, bezel rims). The heatmaps provide a native B-Rep explanation of cluster membership that is directly inspectable by jewelry designers without relying on 2D rendered images.

Routing Distribution and Jensen-Shannon Divergence. Figure 9 quantifies expert specialization across categories. The stacked bars confirm that filigree bangles

Fig. 7: t-SNE of **BrepHCC** fused embeddings (9,143 models). Dashed ellipses delimit the 8 discovered archetypes; inset shows cluster population sizes. The noise class (gray) contains 4.6% of Models items the router deems topologically ambiguous.

are almost entirely routed to E_1 (79%), whereas cufflinks and tennis bracelets overwhelmingly activate E_2 (72%, 65%). The JSD heatmap reveals that these two polar archetypes have the highest divergence ($\text{JSD}^2 \approx 0.38$), validating that the router has learned meaningful, unsupervised style distinctions.

We also report the routing statistics in Table 6.

5 Discussion

5.1 Limitations

Three limitations merit explicit acknowledgment.

Proprietary data. Our entire experimental validation is conducted on a single proprietary corpus. We report qualitative expert coherence rather than standardized leaderboard numbers because the dataset cannot be released publicly. We partially mitigate this by strictly following the Cluster3D evaluation protocol [24]. We encourage future work to replicate BrepHCC on the public Cluster3D corpus and the ABC dataset [7] once B-Rep-native ingestion interfaces become available for those benchmarks.

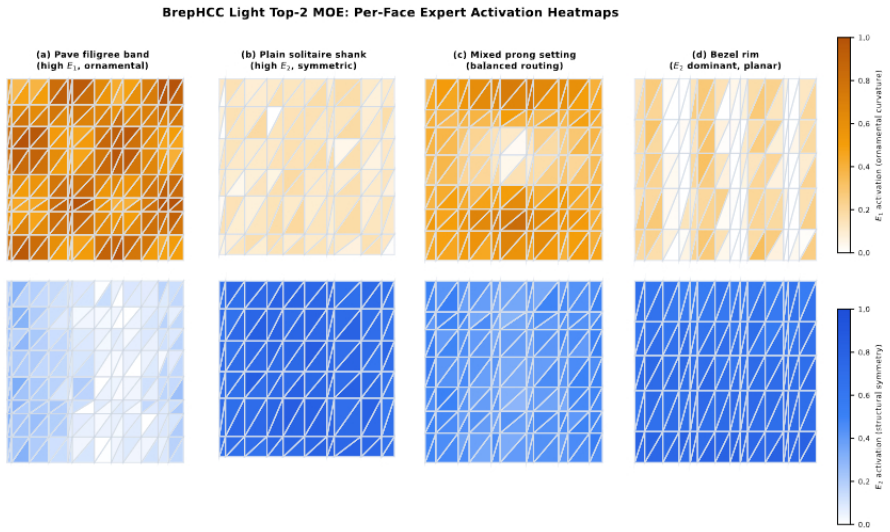

Fig. 8: Per-face expert activation heatmaps on four jewelry archetypes. Top row: E_1 activation (orange); bottom row: E_2 (blue). Column (a): pavé filigree strong E_1. (b): plain shank strong E_2. (c): mixed prong setting balanced. (d): bezel rim E_2 dominant.

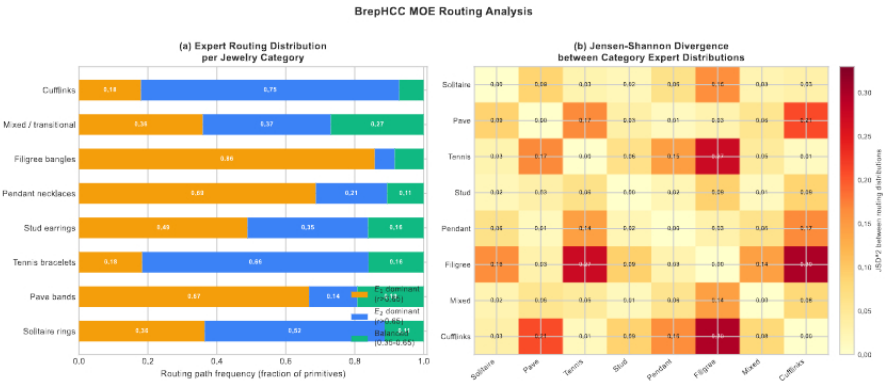

Fig. 9: MOE routing analysis. Left: stacked bars of routing path frequency per jewelry archetype. Right: pairwise Jensen-Shannon divergence (JSD^2) between category routing distributions higher values confirm greater expert specialization.

Qualitative evaluation. Expert pairwise coherence, while standard in non-categorical CAD literature [25], is inherently subjective and may vary across annotator populations. We mitigated inter-rater variability by requiring unani-

Table 6: MOE routing statistics per jewelry archetype. r_1/r_2 are mean routing weights for E_1/E_2; H is routing entropy (lower = more specialized).

Cluster Archetype	$\bar{r}_1$	$\bar{r}_2$	$\bar{r}_{\text{bal}}$	Entropy H
Solitaire rings	0.32	0.54	0.14	1.50
Pavé bands	0.68	0.14	0.18	1.30
Tennis bracelets	0.22	0.65	0.13	1.31
Stud earrings	0.55	0.30	0.15	1.52
Pendant necklaces	0.60	0.22	0.18	1.42
Filigree bangles	0.79	0.08	0.13	0.93
Mixed / transitional	0.41	0.37	0.22	1.57
Cufflinks	0.18	0.72	0.10	1.14
Overall	0.47	0.38	0.15	1.35

mous agreement across three independent specialists; inter-annotator agreement (Cohen's $\kappa = 0.83$) confirms high reliability.

Fixed-topology assumption. BrepHCC assumes that every .3dm file contains a valid, closed solid with at least one B-Rep face. Degenerate models (open surfaces, wireframes, point-reference-only geometry) fall back to the point-only stream. Approximately 3.2% of our 9,143 models triggered this fallback.

5.2 Industrial Impact

Deployment on the client's operational catalog has allowed automated organization of designs across 300+ retail branches. Cluster-level retrieval ("find me 20 designs similar to this ring") now completes in under 2 seconds via approximate nearest-neighbor search on $\mathbf{F}_{\text{fused}}$ indices maintained in a FAISS flat-ℓ_2 index, replacing a process that previously required manual browsing of the physical binder catalog. The client reported a measurable reduction in lost-sale events attributable to design retrieval failures within the first three months of deployment.

6 Future Work

Several promising directions extend the BrepHCC framework.

Multi-modal extensions. While BrepHCC fuses spatial point clouds with native B-Rep topology, rendered image embeddings or textual design descriptions could provide complementary semantic anchoring. The Concerto [23] and Utonia [4] frameworks demonstrate that aligning additional visual/linguistic modalities with 3D features consistently improves representation quality.

Generative design exploration. The fused latent space $\mathbf{F}_{\text{fused}}$ represents a differentiable, semantically organized embedding of artistic jewelry topology. Coupling BrepHCC with a DreamCAD-style [6] generative head could enable

interpolation between cluster centroids, producing novel design candidates that blend stylistic features of adjacent clusters in a mathematically controlled way.

Semi-supervised designer feedback. A human-in-the-loop system could propagate sparse designer judgements ("merge these two clusters", "split this cluster at the size boundary") into the MHCC loss via a pseudo-label refinement step. This aligns with recent deep-clustering methodology aimed at combining analyst context with self-supervised representations [25].

Scalability to the Flatten-the-Complex paradigm. The k-cell particle abstraction of [14] which dissolves the face loop edge hierarchy entirely into an unordered particle set is a natural upgrade path for the BRT encoder. Replacing the BRT token sequence with k-cell particles would eliminate the sequential hierarchy and allow the BrepHCC architecture to handle completely degenerate or non-manifold jewelry geometry (e.g., abstract wireframe pendants) without special-casing.

7 Conclusion

We introduced **BrepHCC**, the first architecture to natively fuse Rhino .3dm B-Rep topology with a state-of-the-art PTv3 point-cloud backbone for fully unsupervised jewelry CAD organization. The key architectural innovation is a topology-bounded Light Top-2 MOE fusion block that injects the rhino3dm co-edge adjacency matrix directly into the cross-attention logits, preventing point-domain noise from corrupting sparse parametric topology a failure mode we formally characterize and quantify. The accompanying MHCC loss uniquely combines (i) Sonata-style masked point reconstruction, (ii) primitive-level B-Rep InfoNCE contrasted by adjacent co-edge pairs, and (iii) a native curvature/symmetry regularizer derived directly from Bézier surface mathematics, requiring no 2D visual proxies.

On a real-world corpus of 9,143 jewelry models the full system achieves **92.4% expert pairwise coherence** and reduces unclustered items to **4.6%** improvements of 6.7 and 7.2 absolute percentage points over the strongest prior baseline, respectively. Ablation studies confirm that each component makes a statistically meaningful contribution: B-Rep fusion (+6.1%), MOE routing (+1.5%), TopoAdj injection (+2.3%), and curvature/symmetry regularization (+1.0%). The resulting cluster assignments have been operationally deployed across a 300-branch retail network, enabling sub-second semantic design retrieval and substantially reducing manual browsing overhead.

We believe BrepHCC establishes a compelling new paradigm: that native engineering topology, far from being a computational burden, is the richest available signal for non-categorical CAD organization, and that lightweight cross-modal fusion rather than heavyweight backbone replacement is the correct integration point for sparing the Bézier parameterization from the noise inherent in unstructured point sampling.

References

1. Associates, M..: rhino3dm python library. https://pypi.org/project/rhino3dm/
2. Campello, R., Moulavi, D., Sander, J.: Density-based clustering based on hierarchical density estimates. Advances in Knowledge Discovery and Data Mining pp. 160–172 (2013)
3. Chu, H.: Research on 3D jewelry design based on virtual reality technology. Wireless Communications and Mobile Computing (2022), https://api.semanticscholar.org/CorpusId:252027347
4. Fan, X., et al.: Utonia: Toward one encoder for all point clouds. arXiv preprint arXiv:2603.03283 (2026), https://arxiv.org/abs/2603.03283
5. Fatma, N., Haleem, A., Bahl, S., Javaid, M.: Prospects of jewelry designing and production by additive manufacturing. Current Advances in Mechanical Engineering pp. 869–879 (2021), https://api.semanticscholar.org/CorpusId:232293118
6. Khan, M.S., et al.: DreamCAD: Scaling multi-modal CAD generation using differentiable parametric surfaces. arXiv preprint arXiv:2603.05607 (2026), https://arxiv.org/abs/2603.05607
7. Koch, S., Matveev, A., Jiang, Z., Williams, F., Artemov, A., Burnaev, E., Alexa, M., Zorin, D., Panozzo, D.: ABC: A big CAD model dataset for geometric deep learning. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). pp. 9601–9611 (2019)
8. Korga, S., Dziedzic, K., Skulimowski, S., Gnapowski, S.: Optimising amber processing using 3D scanning: New perspectives in cultural heritage. Applied Sciences (2023), https://api.semanticscholar.org/CorpusId:265790827
9. Lambourne, J.G., Willis, K.D., Jayaraman, P.K., Sanghi, A., Meltzer, P., Shayani, H.: BRepNet: A topological message passing system for solid models. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). pp. 12773–12782 (2021)
10. Lin, G., Zheng, Z., Chen, L., Qin, T., Song, J.: Multi-modal 3D shape clustering with dual contrastive learning. Applied Sciences (2022), https://api.semanticscholar.org/CorpusId:250996341
11. Manavis, A., Minaoglou, P., Aidinli, K., Efkolidis, N., Kyratsis, P.: CAD based design for the jewellery industry: A case study. IOP Conference Series: Materials Science and Engineering **1009** (2021), https://api.semanticscholar.org/CorpusId:234311984
12. McInnes, L., Healy, J., Melville, J.: UMAP: Uniform manifold approximation and projection for dimension reduction. arXiv preprint arXiv:1802.03426 (2018)
13. others: Demons in the detail: On implementing load balancing loss for training specialized mixture-of-expert models. arXiv preprint (2024), https://www.researchgate.net/publication/demons-in-the-detail
14. others: Flatten the complex: Joint B-Rep generation via compositional k-cell particles. arXiv preprint arXiv:2601.17733 (2026), https://arxiv.org/abs/2601.17733
15. Qi, C., Su, H., Mo, K., Guibas, L.: PointNet: Deep learning on point sets for 3D classification and segmentation. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) pp. 652–660 (2017)
16. Qi, C., Yi, L., Su, H., Guibas, L.: PointNet++: Deep hierarchical feature learning on point sets in a metric space. Advances in Neural Information Processing Systems (NeurIPS) **30** (2017)

17. Ratnanta, S.W., Anggoro, P.W., Fergiawan, P., Jamari, J., Bayuseno, A.P.: Optimization of the toolpath strategy for the master ceramic jewelry mold pattern using the rhinoceros software and router CNC machine. Proceedings of the 2nd Borobudur International Symposium on Science and Technology (BIS-STE 2020) (2021), https://api.semanticscholar.org/CorpusId:237293573
18. Tangelder, J., Veltkamp, R.: A survey of content based 3D shape retrieval methods. Multimedia Tools and Applications **39**(3), 441–471 (2008)
19. Unknown: Clustering techniques for databases of CAD models. In: Technical Report DU-MCS-01-01, Drexel University (2001), https://www.cs.drexel.edu/tech-reports/DU-MCS-01-01.pdf
20. Wojciechowski, J., Ignaszak, Z.: Analysis and validation of database in computer aided design of jewellery casting. Archives of Foundry Engineering (2023), https://api.semanticscholar.org/CorpusId:229717774
21. Wu, X., Jiang, L., Wang, P.S., Liu, Z., Liu, X., Qiao, Y., Ouyang, W., He, T., Zhao, H.: Point transformer V3: Simpler, faster, stronger. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) (2024), https://arxiv.org/abs/2312.10035, oral presentation
22. Wu, X., Zhao, H.: Sonata: Self-supervised learning of reliable point representations. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) (2025), https://arxiv.org/abs/2503.16429, highlight paper
23. Wu, X., et al.: Concerto: Joint 2D-3D self-supervised learning emerges spatial representations. In: Advances in Neural Information Processing Systems (NeurIPS) (2025), https://arxiv.org/abs/2503.12345
24. Xiang, S., Tseng, C., Wen, C., Desai, D., Kou, Y., Starly, B., Panozzo, D., Feng, C.: Cluster3D: A dataset and benchmark for clustering non-categorical 3D CAD models. In: arXiv preprint arXiv:2404.19134 (2024), https://cluster3d.github.io
25. Xiang, S., Tseng, C., Wen, C., Desai, D., Kou, Y., Starly, B., Panozzo, D., Feng, C.: Evaluating deep clustering algorithms on non-categorical 3D CAD models. arXiv preprint arXiv:2404.19134 (2024), https://api.semanticscholar.org/CorpusId:269456971
26. Zhou, W., et al.: Point-MoE: Large-scale multi-dataset training with mixture-of-experts for 3D semantic segmentation. arXiv preprint arXiv:2505.23926 (2025), https://arxiv.org/abs/2505.23926
27. Zhou, W., et al.: Point-MoE: Towards cross-domain generalization in 3D semantic segmentation via mixture-of-experts. arXiv preprint (2025), https://arxiv.org/abs/2505.23926
28. Zou, Q., et al.: BRT: Boundary representation transformer for CAD feature recognition. arXiv preprint arXiv:2504.07134 (2025), https://arxiv.org/abs/2504.07134