DualBrep: A Dual-Field Continuous Representation for B-rep Modelling

ANONYMOUS AUTHOR(S)
SUBMISSION ID: 295

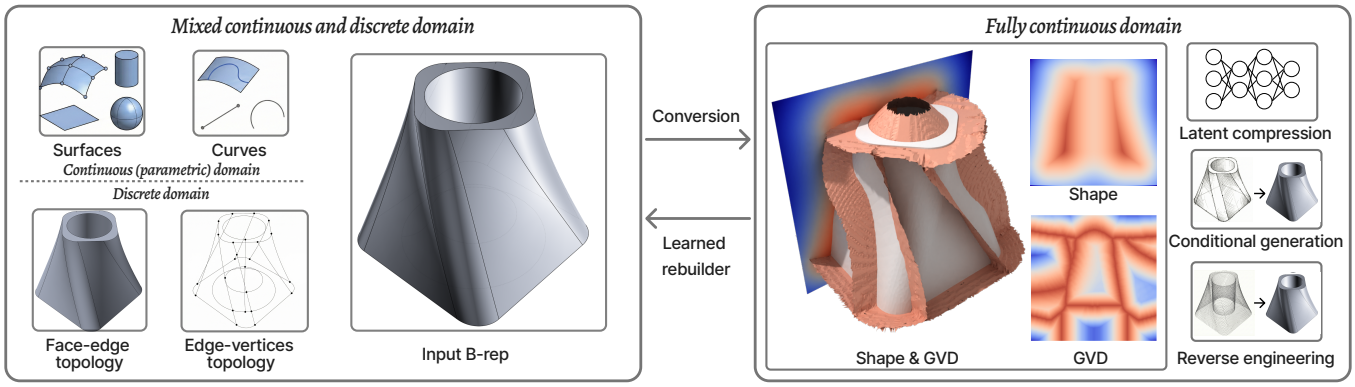


Fig. 1. **DualBrep bridges the discrete-continuous gap in B-rep learning.** Standard B-reps (left) define shapes by explicitly stitching together disjoint parametric surfaces and curves via a discrete connectivity graph, a representation that is difficult to optimize by gradient-based methods. DualBrep reformulates this into a fully continuous domain (right) by encoding geometry as a Shape field and topology as a Generalized Voronoi Diagram (GVD) field. These dual fields are compressed into a single, unified latent representation, enabling robust downstream tasks such as conditional generation and deterministic reverse engineering. A learned rebuilder extracts explicit, watertight B-rep models directly from these continuous signals.

Boundary Representation (B-rep) is the most commonly used data format in Computer-Aided Design (CAD) due to its analytical precision and direct support for parametric editing. However, its heterogeneous data structure—continuous parametric geometry with discrete topological graphs—poses fundamental challenges for deep learning models. Existing methods often *directly* predict the heterogeneous B-rep graph, relying on fixed-size padding or sequential tokenization to handle the varying cardinality of the geometric primitives. These approaches struggle with the combinatorial complexity of CAD models. The discrete, non-differentiable nature of graph data structure prevents end-to-end optimization of the geometry and watertightness. In this work we introduce DualBrep, a novel *continuous* representation that unifies B-rep geometry and topology within a fully structured Euclidean domain. DualBrep encodes a CAD model using dual scalar fields: a Signed Distance Function (SDF) to represent the global shape geometry, and an Unsigned Distance Field (UDF) that implicitly encode the topological structure via a Voronoi partitioning of the surface elements. Crucially, rather than processing these fields independently, we compress them into a single, unified latent representation. This compact embedding forces the model to learn the joint distribution of geometry and topology, ensuring that the predicted topological boundaries align strictly with the underlying geometric surface. Finally, we use a neural rebuilder to extract explicit B-rep models—comprising both prismatic and free-form primitives—directly from our continuous dual scalar fields. We demonstrate that DualBrep serves as a robust, unified backbone for CAD B-rep learning, achieving strong performance in both reverse engineering from raw point clouds and generative modeling via latent diffusion. Code will be released on acceptance.

CCS Concepts: • **Computing methodologies** → **Shape modeling**; **Neural networks**; • **Applied computing** → **Computer-aided design**.

Additional Key Words and Phrases: Boundary representation; Generative models; Diffusion models; Representation learning

1 Introduction

Boundary Representation (B-rep) stands as the de-facto standard format in Computer-Aided Design (CAD), forming the backbone of modern manufacturing, physical simulation, and creative design pipelines. Unlike lightweight representations such as triangle meshes or point clouds, B-reps offer analytical precision and rich editability by encoding solid shapes as a collection of parametric surfaces, trimmed by explicit topological boundaries. With the advent of large-scale CAD datasets [Koch et al. 2019a; Willis et al. 2021] and foundational geometric deep learning, there is a growing interest in modeling the distribution of B-rep data for tasks such as text-to-CAD generation [Khan et al. 2024; Li et al. 2025d; Liu et al. 2025], CAD autocompletion [Xu et al. 2025, 2024], and constrained synthesis [Casey et al. 2025]. However, the inherent structure of B-reps poses a fundamental "compatibility paradox" for deep learning. While neural networks excel at modeling continuous signals in Euclidean space, B-rep data is intrinsically heterogeneous and combinatorial, intertwining various continuous parametric equations (geometry) with discrete graph structures (topology). This structural mismatch makes learning a robust generative B-rep distribution exponentially difficult, especially as shape complexity increases.

Most existing approaches address this heterogeneity by modeling the discrete B-rep either as construction history [Wu et al. 2021; Xu et al. 2023] or raw boundary components [Jayaraman et al. 2023; Xu et al. 2024]. Even recent unified generative models [Li et al. 2025d; Liu et al. 2025; Xu et al. 2025] rely on discrete tokenization or fixed-size padding. This paradigm faces a critical limitation rooted in the complexity of B-reps: modeling the combinatorial topology distribution using gradient-based optimization creates a disjoint optimization landscape. Since the discrete topological choices are non-differentiable, the network cannot be optimized end-to-end

to minimize geometric inconsistencies or to improve watertightness. This optimization challenge is further exacerbated by the long sequences required to represent complex shapes; as the chain of dependencies grows, the generative process suffers from severe error accumulation. Consequently, solid watertightness diminishes rapidly [Liu et al. 2025; Xu et al. 2025], frequently yielding broken shapes that are unusable for engineering.

In contrast, the field of mesh and surface generation have undergone a dramatic transformation driven by continuous implicit representations [Chen et al. 2025; Li et al. 2025c; Wu et al. 2016; Xiang et al. 2025; Zhang et al. 2023]. State-of-the-art methods like Hunyuan3D [Zhao et al. 2025] and TripoSR [Tochilkin et al. 2024] generate highly detailed, watertight shapes with remarkable prompt fidelity. Their success stems from a pivotal shift: rather than predicting discrete mesh directly, they model the underlying shape as a continuous field (e.g., Signed Distance Functions (SDF) or occupancy fields). Since continuous fields are differentiable and resolution-independent, they are naturally aligned with gradient-based optimization, allowing networks to learn complex shape distributions without worrying about combinatorial validity.

Inspired by this dichotomy, we propose to rethink B-rep learning by shifting its representational domain from discrete graphs to continuous fields. Towards this end, we introduce DualBrep, a framework that models B-rep data as a fully continuous signal. Instead of manipulating discrete graph structures, we encode the CAD model into two spatially aligned scalar fields: i) A Signed Distance Function (SDF) that defines the global, watertight shape geometry, ensuring differentiability and resolution independence, and ii) a novel application of the Generalized Voronoi Diagram (GVD), as shown in Fig. 1. Treating the B-rep faces as geometric "sites," the GVD forms a continuous sheet in 3D space that partitions the volume into regions nearest to each B-rep surface. We encode this structure as a *single* Unsigned Distance Field (UDF) representing the distance from any point in space to this GVD boundary. Rather than processing these signals in isolation, we compress them into a single, unified latent representation. This compact embedding forces the model to learn the strict coupling between geometry and topology, ensuring that topological cuts strictly align with the underlying surface. A neural rebuilder then discretizes this consistent signal, recovering explicit surface patches and stitching them into a valid watertight B-rep.

Our continuous backbone unifies two distinct CAD capabilities. First, for *deterministic reconstruction*, the single latent space acts as a robust geometric prior; it imposes structural regularity on unstructured inputs (e.g., raw point clouds), allowing the rebuilder to extract valid models without the "hallucination" artifacts common in graph prediction. Second, the regularized Euclidean manifold enables *generative modeling*, allowing us to train Latent Flow Matching Models that synthesize complex watertight topologies without the error accumulation inherent to autoregressive generation.

To the best of our knowledge, DualBrep is the first framework to leverage a unified continuous domain for B-rep representation learning. While the final conversion back to discrete B-rep is not strictly lossless and may still introduce invalid geometry, this design allows us to defer discretization until the global geometry and topology are well established. Consequently, we avoid forcing the model to hallucinate topology from noise, effectively reducing the final B-rep extraction to a well-defined reconstruction task.

Through extensive experiments, we demonstrate that this paradigm effectively shifts the bottleneck of B-rep modeling. By leveraging our continuous dual-field backbone, DualBrep achieves strong performance across various B-rep learning tasks, including reverse engineering from raw point clouds and conditional generation (point cloud, image) via latent flow matching models. More importantly, we show that DualBrep scales gracefully with shape complexity, maintaining high validity rates and low geometric and topological error as the number of B-rep primitives increases.

2 Related Work

2.1 B-rep Reconstruction

Reconstructing B-rep models from raw point clouds is a longstanding challenge in reverse engineering. Early learning-based approaches, such as ParseNet [Sharma et al. 2020], HPNet [Yan et al. 2021], SED-Net [Li et al. 2023], and Point2CAD [Liu et al. 2024b], typically rely on segmented point sets as their core representation. These methods treat reconstruction as a primitive fitting problem, where the point cloud is first segmented into patches, and each patch is then fitted to a parametric surface. However, this paradigm is usually limited by the expressiveness of the input point cloud and the segmentation accuracy, often failing to capture clean boundaries or global topological consistency.

Distinct from these segmentation-based approaches, Complex-Gen [Guo et al. 2022] formulates reconstruction as a detection task, using a primitive-based representation to regress surfaces, edges, and vertices directly. The lack of explicit topological constraints often leads to models with high rates of invalid topology, such as missing or overlapping primitives. Most relevant to our work is NVDNet [Liu et al. 2024a], which introduces a structure-aware Voronoi partitioning to represent the segmentation of B-rep models. However, NVDNet focuses on local geometric priors and only performs the reconstruction of the Voronoi diagram. Our method extends the Voronoi diagram concept from NVDNet into a dual-distance field formulation. We employ a Variational Autoencoder (VAE) to encode *both* the shape geometry and its Voronoi-based segmentation into a unified global latent space. This global prior not only enables more robust and accurate B-rep reconstruction but also paves the way for conditional generative modeling. Furthermore, our learned rebuilder significantly enhances the capability to recover free-form surfaces, a limitation in many prior works.

2.2 B-rep Generation

Generative modeling of B-rep data has evolved through several paradigms. Command-based methods, such as DeepCAD [Wu et al. 2021], SkexGen [Xu et al. 2022], and HNC-CAD [Xu et al. 2023], represent B-reps as a sequence of construction operations (e.g., sketch and extrude). While these methods guarantee valid CAD models by design, they are limited by the availability of construction history data and struggle to represent complex, free-form surfaces that cannot be easily described by simple operations.

Direct B-rep modeling approaches attempt to generate the B-rep data structure itself, but employ different sequential strategies. SolidGen [Jayaraman et al. 2023] tokenizes the B-rep into a sequence of geometric and topological elements, employing three Transformer models to autoregressively predict them. BrepGen [Xu et al. 2024] leverages a structured latent space to represent the B-rep hierarchy, using diffusion models to generate the geometric and topological features. DTGBrepGen [Li et al. 2025a] explicitly decouples the generation of topology and geometry to reduce complexity. However, these sequential pipelines suffer from severe error accumulation, especially when modeling complex topology graphs, frequently resulting in invalid or broken B-reps. Recent single-stage unified predictors, including HoLa [Liu et al. 2025], AutoBrep [Xu et al. 2025], and BrepGPT [Li et al. 2025d], aim to mitigate this by predicting the entire B-rep structure in a more holistic manner. Nevertheless, they still face the fundamental challenge of modeling variable-length sequences to accommodate the varying cardinality of B-rep elements, which complicates the optimization landscape and fails to guarantee topological validity, often resulting in non-watertight models.

2.3 Mesh Generation

In contrast to the discrete struggles of B-rep generation, mesh generation has seen rapid progress driven by continuous representations. Various continuous representations have been explored to facilitate robust modeling. Latent-based approaches like VecSet [Zhang et al. 2023] and TRELLIS [Xiang et al. 2025] encode shapes into structured latent codes for scalable generation. Meanwhile, methods such as CraftsMan [Li et al. 2025b], DoRA [Chen et al. 2025], Hunyuan3D [Zhao et al. 2025], and TripoSR [Tochilkin et al. 2024] leverage powerful neural fields or VAEs to achieve high-fidelity generation. Sparse voxel is also utilized by XCube [Ren et al. 2024] and Sparc3D [Li et al. 2025c] for efficient high-resolution modeling. These representations are resolution-independent and naturally differentiable, making them exceptionally easy to model with neural networks. Inspired by the success of these continuous approaches, we propose to bridge the gap between B-rep and mesh learning. We introduce a dual-field representation that captures the precision of B-rep topology within a continuous Euclidean domain, allowing for robust learning and generation of complex CAD models.

3 Method

3.1 Dual-Field Representation

B-rep Preliminaries. A standard Boundary Representation (B-rep) model $\mathcal{B} = (\mathcal{V}, \mathcal{E}, \mathcal{F}, \mathcal{T})$ defines a solid via a hierarchy of geometric and topological elements: vertices $\mathcal{V}$, edges $\mathcal{E}$ (parametric curves), faces $\mathcal{F}$ (parametric surfaces), and a topology graph $\mathcal{T}$. The graph $\mathcal{T}$ encodes connectivity, specifically the face-to-edge incidences $\mathcal{T}_{f\to e}$ (defining trimming loops) and edge-to-vertex incidences $\mathcal{T}_{e\to v}$. For the purpose of neural generation, explicitly modeling the full tuple is redundant. Our framework focuses on recovering the minimal sufficient set: the surfaces $\mathcal{F}$, the curves $\mathcal{E}$, and the face-to-edge connectivity $\mathcal{T}_{f\to e}$. Derivative elements, such as vertices $\mathcal{V}$ and edge-to-vertex connectivity $\mathcal{T}_{e\to v}$, can be inferred deterministically from curve intersections once the primary structure is established.

To unify the heterogeneous analytical primitives (e.g., B-splines, planes, cylinders) typically found in $\mathcal{F}$ and $\mathcal{E}$, we adopt a consistent structured discretization [Jayaraman et al. 2021; Lambourne et al. 2021; Xu et al. 2024]. Each face $f_k \in \mathcal{F}$ is represented as a regular geometry grid $\mathbf{G}_k \in \mathbb{R}^{M\times M\times 3}$, generated by uniformly sampling the surface within its parametric bounds $[u_{\min}, u_{\max}] \times [v_{\min}, v_{\max}]$. Similarly, each edge $e_j \in \mathcal{E}$ is sampled into a linear grid $\mathbf{C}_j \in \mathbb{R}^{M\times 3}$, where M is the number of sample points (set to 16 in all our experiments).

Continuous Dual-Field Formulation. As shown in Fig. 1, to circumvent the optimization difficulties inherent to discrete topology graphs, we map the B-rep entirely into the continuous Euclidean domain. Our key insight is that a B-rep can be fundamentally viewed as a watertight geometric hull that is partitioned into distinct regions (faces). We capture this duality by encoding the model into two complementary scalar fields—one representing the geometry and the other representing the structural boundaries (topology).

Geometry field. We use a standard Signed Distance Function $\mathcal{S} : \mathbb{R}^3 \to \mathbb{R}$ to capture the global shape. Its zero-level set, $\mathcal{S}(\mathbf{p}) = 0$, defines the continuous, watertight hull of the object, ignoring the internal segmentation of faces.

Topology Field (GVD). This is the critical component that recovers the B-rep structure. In a B-rep, topological edges exist exactly where two distinct surfaces meet. We capture this relationship volumetrically using the Generalized Voronoi Diagram (GVD). Conceptually, the GVD partitions the ambient 3D space into "Voronoi cells," where every point in a cell is uniquely closest to a single B-rep face f_i. The boundaries between these cells form continuous "medial sheets" in 3D space. We encode these sheets as an Unsigned Distance Field (UDF) $\mathcal{U} : \mathbb{R}^3 \to \mathbb{R}$, where $\mathcal{U}(\mathbf{p})$ gives the distance from point $\mathbf{p}$ to the nearest point on the GVD surface.

Why this representation. The power of this formulation lies in the intersection of these fields. The SDF defines *where the surface exists*, while the GVD defines *where the surface identity changes*.

- $\mathcal{S}(\mathbf{p}) \approx 0$: The point lies on the object surface.
- $\mathcal{U}(\mathbf{p}) \approx 0$: The point is equidistant to multiple faces (i.e., it lies on a medial sheet or edge).

Therefore, the superposition of $\mathcal{S}$ and $\mathcal{U}$ provides a complete definition of the B-rep: the SDF recovers the geometry, and the GVD implicitly "cuts" this geometry into the correct topological patches. This allows us to learn complex topologies without needing to predict discrete adjacency matrices or handle the variable cardinality of faces, as the GVD naturally adapts to any number of partitions.

3.2 Dual-Field Variational Autoencoder

We employ a Variational Autoencoder (VAE) to compress this dual-field representation into a single, compact latent space. Inspired by recent neural field architectures [Chen et al. 2025; Zhang et al. 2023], we design a Perceiver-style encoder that uses cross-attention to aggregate multi-modal inputs, as illustrated in Fig. 2.

Input Representation. To capture high-frequency details, we sample three point sets as input: (i) Surface points $\mathcal{P}_s$ from the shape geometry ($N_s = 32,768$); (ii) Edge points $\mathcal{P}_e$ along B-rep curves

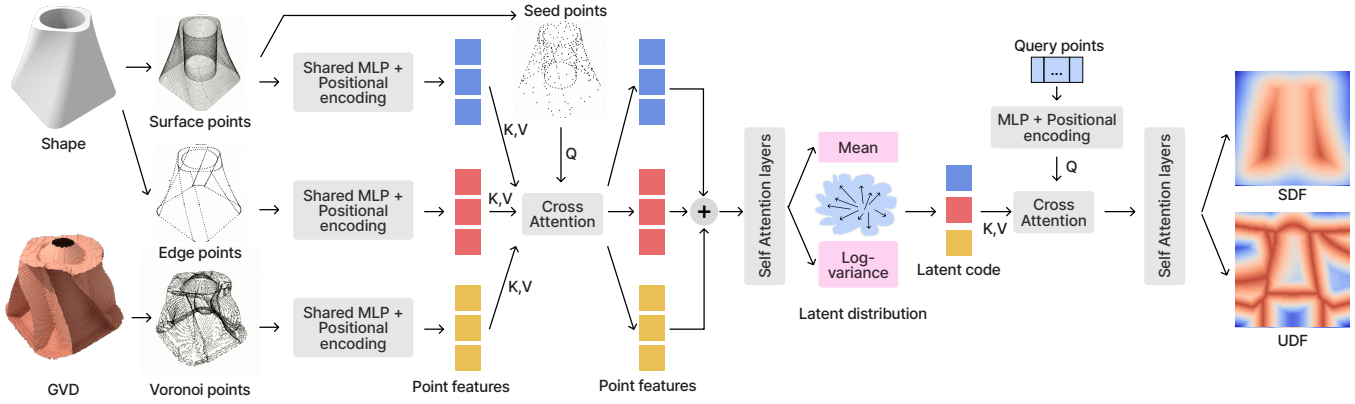


Fig. 2. *Architecture of the VAE. A Perceiver-style encoder fuses surface, edge, and Voronoi point features into a shared latent representation using cross-attention operation. Then cross-attention–based decoder queries this latent code to reconstruct the shape (SDF) and GVD of the B-rep faces (UDF), providing a unified and differentiable representation of B-rep geometry and topology.*

($N_e = 32,768$); and (iii) Voronoi points $\mathcal{P}_v$ from the GVD sheets ($N_v = 32,768$). These sets collectively describe the surface, the sharp features, and the topological boundaries.

Encoder Architecture. The encoder maps these inputs to a latent embedding $\mathbf{Z} \in \mathbb{R}^{K \times D}$, with sequence length $K = 2048$ and dimension $D = 32$ (split into mean and log-variance). We first embed the spatial coordinates of each set $\mathcal{P}_m$ ($m \in \{s, e, v\}$) using a shared MLP with a shared sinusoidal positional encodings $\gamma(\cdot)$:

$$\mathbf{F}_m = \mathrm{MLP}_{\mathrm{in}}(\gamma(\mathbf{p})) \mid \mathbf{p} \in \mathcal{P}_m. \tag{1}$$

To form the latent bottleneck, we select K anchor points from $\mathcal{P}_s$ via Farthest Point Sampling to serve as queries $\mathbf{Q}$. To disentangle the contributions of geometry and topology, the queries attend to each input modality *individually* via cross-attention:

$$\mathbf{H}_m = \mathrm{CrossAttn}(\mathbf{Q}, \mathbf{F}_m), \quad m \in \{s, e, v\}. \tag{2}$$

These modality-specific features are fused via a weighted sum and refined via self-attention layers to enable global information exchange, before being projected to the variational parameters $\boldsymbol{\mu}, \log \boldsymbol{\sigma}^2$ as the latent distribution $Z \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\sigma})$.

Decoder Architecture. The decoder acts as an implicit neural function. Given a query coordinate $\mathbf{x} \in \mathbb{R}^3$, we extract local context by attending to the sampled latent codes $\mathbf{z} \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\sigma})$:

$$\mathbf{h}_x = \mathrm{CrossAttn}(\mathrm{MLP}_q(\gamma(\mathbf{x})), \mathbf{z}). \tag{3}$$

After a set of self-attention layers, two separate MLP heads then regress the field values: $\hat{s}(\mathbf{x}) = \mathrm{MLP}_{\mathrm{SDF}}(\mathbf{h}_x)$ and $\hat{u}(\mathbf{x}) = \mathrm{MLP}_{\mathrm{UDF}}(\mathbf{h}_x)$.

Training Objectives. We train the VAE end-to-end using a hybrid sampling strategy for query points $\mathbf{x}$, drawing from the bounding volume, near-surface regions, and near-edge regions in a 1:1:2 ratio [Chen et al. 2025]. The loss function is the sum of L_1 reconstruction loss and KL regularization terms with a weight of 0.001. The VAE is ≈250M parameters, consisting of 16 and 8 self-attention layers in the encoder and decoder respectively. Each layer has 8 attention heads and a hidden dimension of 1024.

Deterministic Mode for Reverse Engineering. For the task of reverse engineering (reconstructing B-reps from raw point clouds), we adapt the VAE into a *deterministic* autoencoder and train the model from scratch. We mask out the edge and Voronoi inputs ($\mathcal{P}_e, \mathcal{P}_v$) and set the weight of KL loss to 0. This forces the encoder to infer the complete dual-field structure—including topological segmentation—solely from the surface geometry $\mathcal{P}_s$.

3.3 Latent Flow Matching for Generation

To enable scalable and controllable generation, we train a conditional Flow Matching [Lipman et al. 2023] model over the latent space $\mathcal{Z}$.

Unified Geometric Conditioning. While our framework is agnostic to the conditioning modality $\mathbf{c}$, we prioritize *point clouds* as a primary modality due to their widespread availability and adoption. We encode point cloud conditions using a Perceiver-style encoder similar to our VAE, aggregating features from $N_c = 1024$ seed points. Motivated by its potential for future editing applications, we also include results for native single-view image generation, where the conditional signal is obtained by a pre-trained DINOv2 [Oquab et al. 2023] ViT. Other modalities (e.g., text embeddings from LLMs) can be similarly integrated.

Architecture and Inference. We implement the flow matching model using a Diffusion Transformer (DiT) [Peebles and Xie 2022] adapted for set-structured latents (16 layers, 1024 dim, ≈300M parameters). Conditioning is injected via cross-attention, and time t modulates features via Adaptive Layer Norm (AdaLN). During inference, we solve the ODE using the Euler method (50 steps) to obtain $\hat{\mathbf{z}}$. This code is decoded into dual fields $\mathcal{S}$ and $\mathcal{U}$. We extract the surface mesh via Marching Cubes on $\mathcal{S}$ and segment it into a segmented mesh $\mathcal{M}_{\mathrm{seg}}$ via region growing constrained by the boundaries in $\mathcal{U}$.

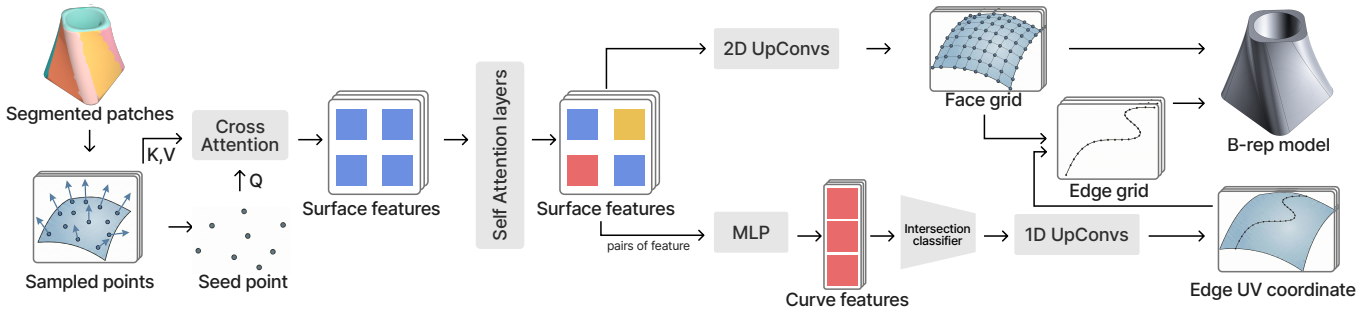


Fig. 3. *Architecture of the learned rebuilder. Segmented face patches are encoded into patch-level features and processed by global self-attention to capture topological context. For each patch, the network predicts parametric surface grids, adjacency relationships, and UV-space trimming curves, which are then assembled into a watertight B-rep model.*

3.4 Learned B-rep Rebuilder

The final component is a learned rebuilder that converts the explicit segmented mesh $\mathcal{M}_{\text{seg}}$ into an analytical B-rep. Unlike heuristic fitting methods [Liu et al. 2024a], our neural rebuilder directly predicts parametric structures, offering robustness to noise and support for free-form surfaces.

Problem Formulation. Given N disjoint face patches $\{\mathcal{P}_1, \ldots, \mathcal{P}_N\}$ from the segmented mesh $\mathcal{M}_{\text{seg}}$, the rebuilder predicts: (i) *Surface Geometry:* A structured grid $\mathbf{G}_k$ for each patch; (ii) *Topology:* An adjacency matrix $\mathbf{A} \in \{0,1\}^{N\times N}$; (iii) *Trim Curves:* For each connected pair (i, j), a parametric curve $\mathbf{C}_{ij}$ defining the shared boundary.

Feature Encoding. We represent each patch $\mathcal{P}_k$ by sampling $N_p = 100$ points (positions and normals). A lightweight Perceiver-style encoder maps these points to a patch feature vector $\mathbf{h}_k^{(0)} \in \mathbb{R}^{4\times 256}$, where 4 is the number of subsampled seed points per patch and 256 is the feature dimension. To capture topological context, these features are processed by a global self-attention module:

$$[\mathbf{h}_1, \ldots, \mathbf{h}_N] = \text{SelfAttn}([\mathbf{h}_1^{(0)}, \ldots, \mathbf{h}_N^{(0)}]), \tag{4}$$

where $\mathbf{h}_k \in \mathbb{R}^{4\times 256}$ is the refined feature for patch $\mathcal{P}_k$.

Surface Geometry Head. This head reconstructs the parametric surface geometry. For each patch, we first reshape the learned feature tokens $\mathbf{h}_k \in \mathbb{R}^{4\times 256}$ into a spatial feature map of size $2 \times 2 \times 256$. This feature map is progressively upsampled via a sequence of 2D transposed convolution blocks (ConvTranspose → ReLU → BatchNorm) to resolutions 4^2, 8^2, and finally 16^2:

$$\mathbf{G}_k = \text{ConvTranspose}_{2\text{D}}(\text{Reshape}(\mathbf{h}_k, [2, 2, 256])) \in \mathbb{R}^{16\times 16\times 3}. \tag{5}$$

The output grid $\mathbf{G}_k$ represents 3D control points sampled uniformly in the patch's UV parameter domain, which are subsequently used to fit the final B-spline surface.

Edge and Trim Prediction. To determine connectivity and boundaries between two patches i and j, we first construct an edge feature by concatenating their token sequences: $\mathbf{e}_{ij} = [\mathbf{h}_i; \mathbf{h}_j] \in \mathbb{R}^{8\times 256}$. This feature is flattened to a vector of size 2048 and processed by an MLP to predict the adjacency probability $\hat{a}_{ij}$. For active edges ($\hat{a}_{ij} > 0.5$), we regress the trim curve. Unlike prior works [Liu et al. 2025; Xu et al. 2024] that predict curves in 3D Euclidean space, we predict the curve in the 2D UV domain of its corresponding surface to ensure consistency with the surface parameterization. The flattened feature is projected and reshaped into a 1D sequence, then progressively upsampled via 1D transposed convolutions to produce the final curve $\mathbf{C}_{ij} \in \mathbb{R}^{16\times 2}$:

$$\mathbf{C}_{ij} = \text{ConvTranspose}_{1\text{D}}(\text{MLP}_{\text{proj}}(\text{Flatten}(\mathbf{e}_{ij}))). \tag{6}$$

Canonicalization Strategy. A key challenge is the ambiguity of UV mappings (e.g., arbitrary rotation). We enforce a *canonical pose* during training: for every ground-truth surface, we sort the sampled points lexicographically (ZYX) and define the point with the minimum value as the UV origin. This ensures the network learns a pose-invariant parameterization.

B-rep Rebuilding. Similar to previous works [Liu et al. 2025; Xu et al. 2024], the final B-rep is assembled deterministically: (i) B-spline surfaces are fitted to the predicted grids $\mathbf{G}_k$; (ii) Predicted 2D trim points $\mathbf{C}_{ij}$ are projected onto these surfaces to form 3D edges; (iii) The topology graph is built from the predicted adjacency matrix $\hat{\mathbf{A}}$, and the model is stitched using a standard CAD kernel.

4 Results

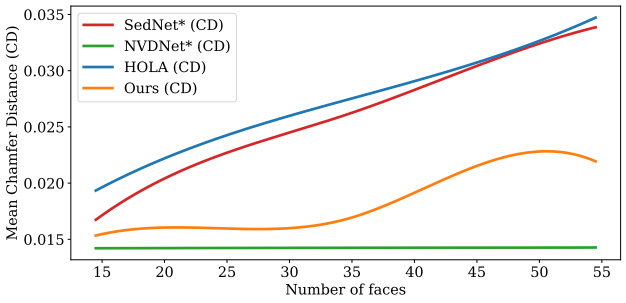
We validate DualBrep primarily as a robust backbone for deterministic reverse engineering—converting raw, noisy point clouds into watertight, editable CAD models. Additionally, we demonstrate that our unified latent representation constructs a high-quality manifold for generative tasks, enabling multi-modal synthesis via flow matching.

4.1 Experimental Setup

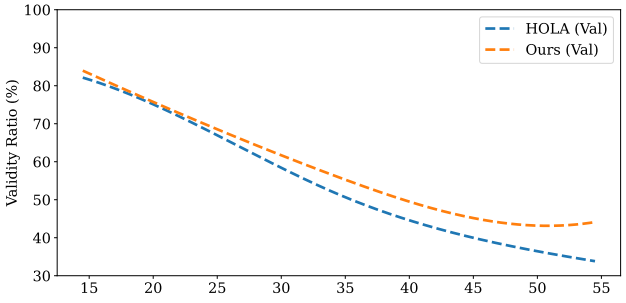
To evaluate the proposed DualBrep framework, we treat it as the backbone model on point cloud to B-rep reverse engineering tasks. We conduct experiments on the ABC dataset [Koch et al. 2019b], which contains a large collection of CAD models with corresponding B-rep representations. To include more diverse and complex CAD models, we use the complete ABC dataset instead of the DeepCAD [Wu et al. 2021] subset, filtering out models that are too simple (i.e., consisting of less than 10 faces) or too complex (consisting of more than 100 faces or multiple solids) and duplicated samples. After

Table 1. Quantitative comparison of point cloud to B-rep reconstruction on ABC dataset. We report the number of predicted primitives versus GT primitives, Chamfer Distance (CD, lower is better) for geometric accuracy, F1-score (%) for primitive detection accuracy and topological correctness, and overall validity rate (%). Best results are in **bold**, second best are underlined.

Method	Primitive Count (#)			Geometric Accuracy (CD ↓)			Primitive F1-score (%) ↑			Topology F1-score (%) ↑		Validity
	Surface	Edge	Vertex	Surface	Edge	Vertex	Surface	Edge	Vertex	Face-Edge	Edge-Vertex	Rate (%) ↑
SEDNet+Point2CAD	7.9/19.2	18.2/48	12.7/31.5	0.0259	0.0336	0.1499	48.68%	45.01%	41.21%	40.87%	33.65%	/
NVDNet	22.8/19.2	67.7/48	53.2/31.5	**0.0142**	**0.0059**	0.0301	**83.70%**	78.48%	71.92%	**80.84%**	74.27%	/
HoLa-BRep	16.0/19.2	**44.5/48**	**29.7/31.5**	0.0211	0.0329	0.0750	79.33%	71.65%	65.08%	68.81%	66.41%	73.98%
Ours_{gen}	**19.3/19.2**	55.8/48	40.2/31.5	0.0157	0.0133	0.0247	77.18%	86.68%	81.33%	70.31%	84.91%	69.49%
$\text{Ours}_{\text{recon}}$	**19.3/19.2**	53.9/48	37.6/31.5	0.0156	0.0129	**0.0203**	81.98%	**89.87%**	**84.74%**	76.36%	**88.17%**	**76.34%**



(a) Geometric error vs. shape complexity



(b) Validity vs. shape complexity

Fig. 4. **Reconstruction performance vs. shape complexity.** We analyze how the reconstruction performance of different methods varies with shape complexity, measured by the number of faces in the B-rep model. *Top*: Chamfer Distance (lower is better) vs. number of faces. *Bottom*: Validity rate (higher is better) vs. number of faces. Our DualBrep framework maintains stable performance across different shape complexities, outperforming baseline methods in both geometric accuracy and validity, especially for complex shapes with many faces. Note that curves are interpolating B-splines of degree 3 fit to the data.

filtering, we obtain a dataset of 80k CAD models, where 4k models are used for testing and the rest are for training and validation.

Metrics. Following prior works [Liu et al. 2024a, 2025], we evaluate the performance of our method using the following metrics:

- *Geometric accuracy:* We use Chamfer Distance to measure the geometric accuracy of the reconstructed B-rep surface, edges and vertices against their ground truth counterparts.
- *Primitive accuracy:* In addition to geometric accuracy that measures the global shape similarity, we also evaluate how the parametric primitives are grouped to form the final B-rep structure. We measure the detection-based scores including precision, recall, and F1-score for each primitive type (surfaces, edges, vertices).
- *Topological accuracy:* We evaluate the correctness of the predicted B-rep topology by comparing the predicted and ground truth adjacency matrices of B-rep elements (faces, edges, vertices). We report the accuracy, precision, recall, and F1-score for each adjacency matrix.
- *Validity:* We also report the percentage of valid B-rep models among all reconstructed models, where a valid B-rep model is defined as one that satisfies the topological and geometric constraints of a well-formed B-rep [Xu et al. 2024].

Baselines. We compare our DualBrep framework with several state-of-the-art methods for point cloud to B-rep reconstruction, including segmentation based methods (SEDNet+Point2CAD [Li et al. 2023; Liu et al. 2024b]), voronoi-based method (NVDNet [Liu et al. 2024a]), and generative methods which have point cloud as an input condition [Liu et al. 2025].

4.2 Reverse Engineering

Table 1 summarizes the performance on the point cloud to B-rep reconstruction task. DualBrep outperforms baseline methods across the majority of metrics, achieving a notable validity ratio of 76.34%. This result validates our core hypothesis: compressing geometry and topology into a unified latent space efficiently enforces their alignment, significantly surpassing segmentation-based approaches that treat them separately.

While NVDNet achieves a marginally lower Chamfer Distance, it optimizes for raw point-to-surface distance without guaranteeing structural integrity. Lacking explicit UV parameterization, NVDNet is restricted to fitting standard prismatic primitives and relies on alpha-shapes for trimming—a heuristic that frequently yields non-watertight models incompatible with CAD software. In contrast, DualBrep explicitly models UV grids for both faces and edges. This allows us to recover intricate edge loops and vertex relationships with superior accuracy. Consequently, although our raw Chamfer distance is slightly higher, our framework delivers models that are topologically valid and engineering-ready, rather than just geometrically close clouds of surfaces.

We further investigated how performance degrades as shape complexity increases (Fig. 4). Discrete methods like HoLa struggle due to the combinatorial explosion of the topological search space. Directly predicting discrete adjacency graphs becomes exponentially difficult as the number of faces grows, making it hard to ensure global validity. Conversely, Voronoi-based methods—both NVDNet and DualBrep —maintain stable performance regardless of face count due to their continuous field representation. However, DualBrep distinguishes itself by combining this scalability with global consistency. Unlike NVDNet, which relies solely on local cues and often produces disjointed elements, our global dual-field backbone ensures structural coherence for both B-rep faces and edges. As shown in Fig. 7, our method faithfully reconstructs intricate details—such as free-form surfaces, gears with 20+ teeth, and thin-walled frames—where baselines often fail.

Finally, we report results for the generative variant of DualBrep trained via latent flow matching. Interestingly, this model yields slightly lower reconstruction metrics than the deterministic version. We attribute this to the nature of reverse engineering, which is an inherently one-to-one mapping task; the stochasticity of a generative model can introduce plausible variations that, while valid, strictly deviate from the specific ground truth instance. Nevertheless, its competitive performance confirms the robustness of our latent space, demonstrating its potential for broader CAD generation tasks beyond direct reconstruction.

4.3 Native Image-to-B-rep Generation

Beyond deterministic reconstruction, we demonstrate the versatility of our latent representation by training a Latent Flow Matching model for image-based conditional generation. As shown in Fig. 4.3, our model successfully generates watertight B-rep assemblies directly from single-view RGB images. Unlike deterministic reverse engineering, which optimizes for a single solution, this generative approach learns the conditional distribution, allowing it to hallucinate plausible missing details in occluded regions while respecting the global structure of the input. Crucially, the generated models maintain high topological validity and geometric fidelity, effectively handling complex features such as free-form curvature and intricate mechanical details.

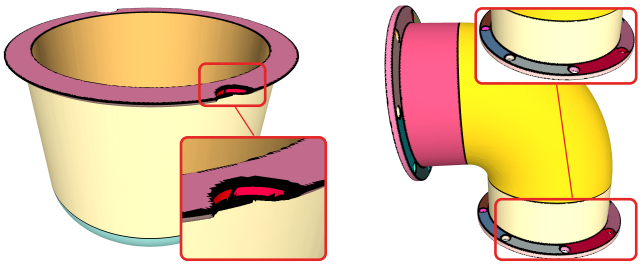


Fig. 5. **Failure cases and limitations.** While we showed cases with narrow or thin structures can be handled well by our DualBrep framework, extremely thin features may still be lost during the segmentation, leading to invalid B-rep models.

4.4 Limitations and Failure Cases

Despite the robustness of our unified continuous representation, several limitations remain. First, as with any volumetric approach, our method is constrained by the grid resolution of the latent fields. Extremely thin geometric features or high-frequency topological details that fall below the sampling rate may be aliased or lost entirely. As illustrated in Fig. 5, this resolution bottleneck can cause thin-walled structures to vanish or inadvertently merge with adjacent regions during the GVD segmentation, leading to topological validity failures in the final model. Second, while the predicted continuous fields are generally consistent, the final extraction of explicit B-reps relies on a neural rebuilder. In rare cases involving complex intersections, this discretization step may still fail to stitch patches perfectly, resulting in minor non-watertight artifacts. Future work could mitigate these issues by incorporating adaptive octree sampling or hybrid explicit-implicit representations [Li et al. 2025c], effectively decoupling fine-scale feature preservation from global grid resolution.

5 Conclusion

We introduce DualBrep, a paradigm shift in B-rep learning that reframes the representation from discrete combinatorial graphs to continuous scalar fields. Our approach compresses geometry and topology into a single latent representation, decoded into dual fields: a Signed Distance Function (SDF) for watertight geometry and an Unsigned Distance Field (UDF) derived from the Generalized Voronoi Diagram for topological segmentation. By deferring discrete decisions until continuous fields are established, DualBrep avoids the error accumulation and optimization challenges inherent to discrete tokenization, enabling the network to learn complex shape distributions without worrying about combinatorial validity. Through extensive experiments across reverse engineering and conditional diffusion, we demonstrate that continuous representations provide a robust backbone for B-rep modeling.

The continuous dual-field paradigm opens promising avenues beyond generative modeling. A natural extension is to incorporate physics-based simulations—such as finite element analysis, computational fluid dynamics, or mechanical stress testing—directly into the learning pipeline, as these applications inherently operate on continuous representations. The differentiable nature of our field formulation enables end-to-end optimization toward simulation objectives, potentially allowing networks to learn manufacturable, structurally sound designs that satisfy both geometric and physical constraints. Similarly, extending to other engineering domains that rely on continuous field representations, such as thermal analysis or electromagnetic simulation, could enable holistic CAD intelligence that reasons about form, function, and performance simultaneously.

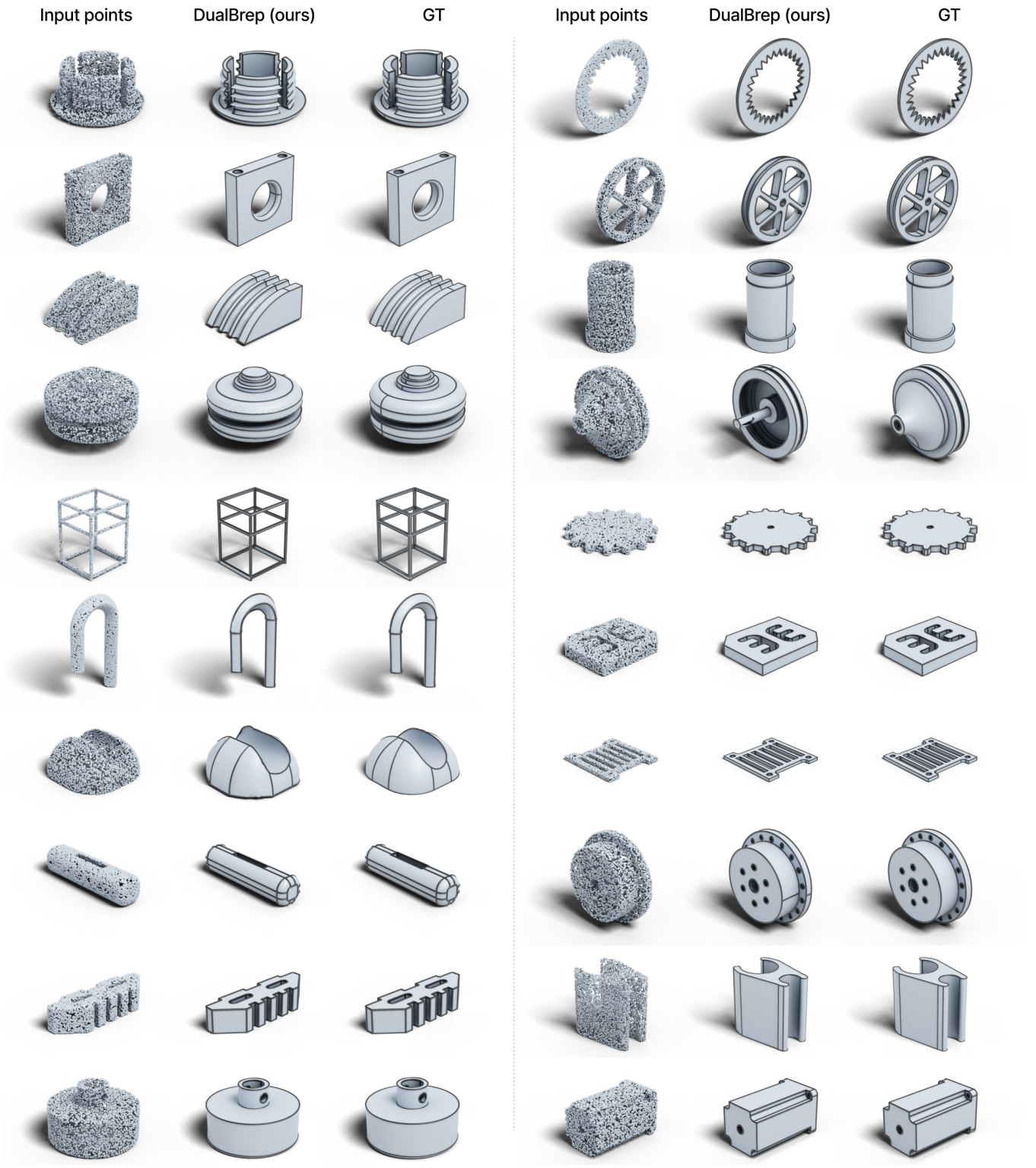


Fig. 6. **Point cloud to B-rep reconstruction gallery.** We showcase diverse reconstruction results from our deterministic $\text{DualBrep}_{\text{recon}}$ across various CAD model categories, including shapes with free-form surfaces, mechanical parts with intricate details and thin-walled structures.

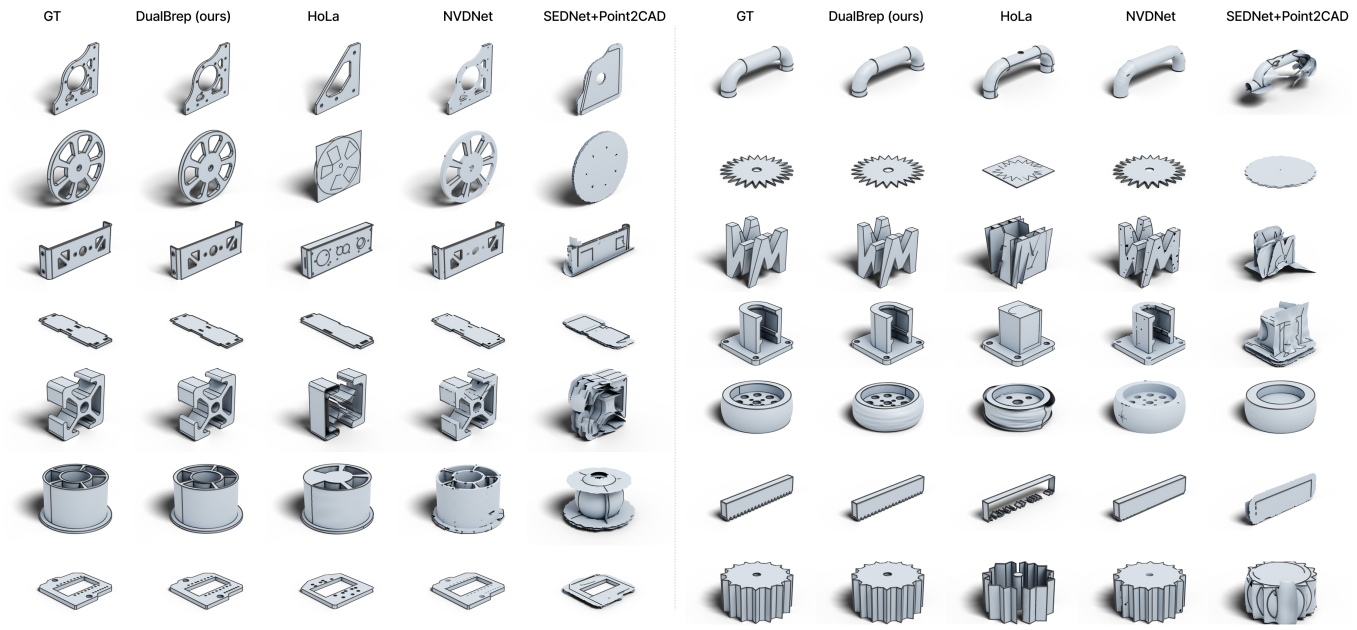


Fig. 7. **Qualitative comparison on point cloud to B-rep reconstruction.** We compare our DualBrep$_{\text{recon}}$ with baseline methods on various CAD models. From left to right: ground truth B-rep, DualBrep (Ours) HoLa, NVDNet and SEDNet+Point2CAD. Our method produces more accurate surface segmentation and better preserves geometric details while maintaining topological validity, even on complex shapes like gears with 20+ teeth or mechanical parts with numerous holes and cutouts.

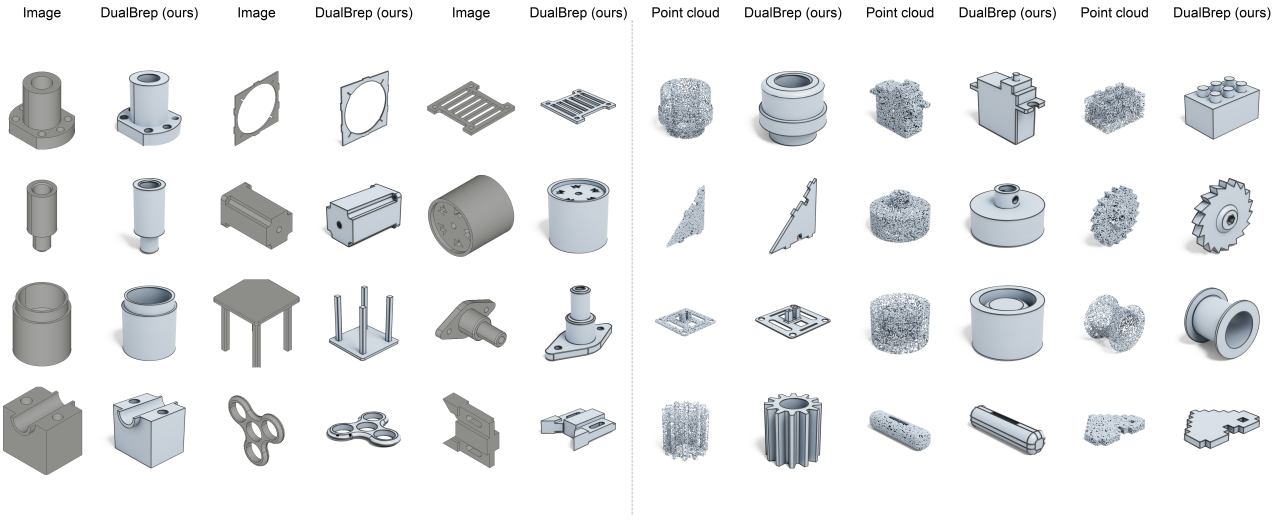


Fig. 8. **Native conditional generation.** Single view image or point cloud can be injected as conditions to our latent flow matching model DualBrep$_{\text{gen}}$ for direct B-rep generation. We show various results of point cloud to B-rep generation (right) and image to B-rep generation (left).

References

Evan Casey, Tianyu Zhang, Shu Ishida, John Roger Thompson, Amir Khasahmadi, Joseph George Lambourne, Pradeep Kumar Jayaraman, and Karl D.D. Willis. 2025. Aligning Constraint Generation with Design Intent in Parametric CAD. In *Proc. Int. Conf. on Computer Vision*. 8613–8622.

Rui Chen, Jianfeng Zhang, Yixun Liang, Guan Luo, Weiyu Li, Jiarui Liu, Xiu Li, Xiaoxiao Long, Jiashi Feng, and Ping Tan. 2025. Dora: Sampling and Benchmarking for 3D Shape Variational Auto-Encoders. In *Proc. IEEE/CVF Conf. on Computer Vision & Pattern Recognition*. 16251–16261.

Haoxiang Guo, Shilin Liu, Hao Pan, Yang Liu, Xin Tong, and Baining Guo. 2022. ComplexGen: CAD reconstruction by B-rep chain complex generation. *ACM Trans. on Graphics (Proc. SIGGRAPH)* 41, 4 (2022), 129:1–129:18.

Pradeep Kumar Jayaraman, Joseph George Lambourne, Nishkrit Desai, Karl D. D. Willis, Aditya Sanghi, and Nigel J. W. Morris. 2023. SolidGen: An Autoregressive Model for Direct B-rep Synthesis. *Trans. on Machine Learning Research* (2023).

Pradeep Kumar Jayaraman, Aditya Sanghi, Joseph G. Lambourne, Karl D. D. Willis, Thomas Davies, Hooman Shayani, and Nigel J. W. Morris. 2021. UV-Net: Learning From Boundary Representations. In *Proc. IEEE/CVF Conf. on Computer Vision & Pattern Recognition*. 11703–11712.

Mohammad Sadil Khan, Sankalp Sinha, Sheikh Talha Uddin, Didier Stricker, Sk Aziz Ali, and Muhammad Zeshan Afzal. 2024. Text2CAD: Generating Sequential CAD Designs from Beginner-to-Expert Level Text Prompts. In *Proc. Conf. on Neural Information Processing Systems*.

Sebastian Koch, Albert Matveev, Zhongshi Jiang, Francis Williams, Alexey Artemov, Evgeny Burnaev, Marc Alexa, Denis Zorin, and Daniele Panozzo. 2019a. ABC: A Big CAD Model Dataset for Geometric Deep Learning. In *Proc. IEEE/CVF Conf. on Computer Vision & Pattern Recognition*. 9601–9611.

Sebastian Koch, Albert Matveev, Zhongshi Jiang, Francis Williams, Alexey Artemov, Evgeny Burnaev, Marc Alexa, Denis Zorin, and Daniele Panozzo. 2019b. ABC: A Big CAD Model Dataset For Geometric Deep Learning. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*.

Joseph G. Lambourne, Karl D.D. Willis, Pradeep Kumar Jayaraman, Aditya Sanghi, Peter Meltzer, and Hooman Shayani. 2021. BRepNet: A topological message passing system for solid models. In *Proc. IEEE/CVF Conf. on Computer Vision & Pattern Recognition*. 12773–12782.

Jing Li, Yihang Fu, and Falai Chen. 2025a. DTGBrepGen: A Novel B-rep Generative Model through Decoupling Topology and Geometry. In *Proc. IEEE/CVF Conf. on Computer Vision & Pattern Recognition*.

Pu Li, Wenhao Zhang, Weize Quan, Biao Zhang, Peter Wonka, and Dongming Yan. 2025d. BrepGPT: Autoregressive B-rep Generation with Voronoi Half-Patch. *ACM Trans. on Graphics (Proc. SIGGRAPH Asia)* 44, 6 (2025), 226:1–226:18 pages.

Weiyu Li, Jiarui Liu, Hongyu Yan, Rui Chen, Yixun Liang, Xuelin Chen, Ping Tan, and Xiaoxiao Long. 2025b. CraftsMan3D: High-fidelity Mesh Generation with 3D Native Diffusion and Interactive Geometry Refiner. In *Proc. IEEE/CVF Conf. on Computer Vision & Pattern Recognition*. 5307–5317.

Yuanqi Li, Shun Liu, Xinran Yang, Jianwei Guo, Jie Guo, and Yanwen Guo. 2023. Surface and Edge Detection for Primitive Fitting of Point Clouds. In *Proc. SIGGRAPH*. 44:1–44:10.

Zhihao Li, Yufei Wang, Heliang Zheng, Yihao Luo, and Bihan Wen. 2025c. Sparc3D: Sparse Representation and Construction for High-Resolution 3D Shapes Modeling. In *Proc. Conf. on Neural Information Processing Systems*.

Yaron Lipman, Ricky T. Q. Chen, Heli Ben-Hamu, Maximilian Nickel, and Matthew Le. 2023. Flow Matching for Generative Modeling. In *Proc. Int. Conf. on Learning Representations*.

Yilin Liu, Jiale Chen, Shanshan Pan, Daniel Cohen-Or, Hao Zhang, and Hui Huang. 2024a. Split-and-Fit: Learning B-Reps via Structure-Aware Voronoi Partitioning. *ACM Trans. on Graphics (Proc. SIGGRAPH)* 43, 4 (2024), 108:1–108:13.

Yujia Liu, Anton Obukhov, Jan Dirk Wegner, and Konrad Schindler. 2024b. Point2CAD: Reverse Engineering CAD Models from 3D Point Clouds. In *Proc. IEEE/CVF Conf. on Computer Vision & Pattern Recognition*. 3763–3772.

Yilin Liu, Duoteng Xu, Xinyao Yu, Xiang Xu, Daniel Cohen-Or, Hao Zhang, and Hui Huang. 2025. HoLa: B-Rep Generation using a Holistic Latent Representation. *ACM Transactions on Graphics (SIGGRAPH)* 44, 4 (2025).

Maxime Oquab, Timothée Darcet, Théo Moutakanni, Huy Vo, Marc Szafraniec, Vasil Khalidov, Pierre Fernandez, Daniel Haziza, Francisco Massa, Alaaeldin El-Nouby, et al. 2023. Dinov2: Learning robust visual features without supervision. *arXiv preprint arXiv:2304.07193* (2023).

William Peebles and Saining Xie. 2022. Scalable Diffusion Models with Transformers. *arXiv preprint arXiv:2212.09748* (2022).

Xuanchi Ren, Jiahui Huang, Xiaohui Zeng, Ken Museth, Sanja Fidler, and Francis Williams. 2024. XCube: Large-Scale 3D Generative Modeling using Sparse Voxel Hierarchies. In *Proc. IEEE/CVF Conf. on Computer Vision & Pattern Recognition*. 4209–4219.

Gopal Sharma, Difan Liu, Subhransu Maji, Evangelos Kalogerakis, Siddhartha Chaudhuri, and Radomír Měch. 2020. ParSeNet: A Parametric Surface Fitting Network for 3D Point Clouds. In *Proc. Euro. Conf. on Computer Vision*, Vol. 12352. 261–276.

Dmitry Tochilkin, David Pankratz, Zexiang Liu, Zixuan Huang, , Adam Letts, Yangguang Li, Ding Liang, Christian Laforte, Varun Jampani, and Yan-Pei Cao. 2024. TripoSR: Fast 3D Object Reconstruction from a Single Image. *arXiv preprint arXiv:2403.02151* (2024).

Karl DD Willis, Yewen Pu, Jieliang Luo, Hang Chu, Tao Du, Joseph G Lambourne, Armando Solar-Lezama, and Wojciech Matusik. 2021. Fusion 360 gallery: A dataset and environment for programmatic cad construction from human design sequences. *ACM Trans. on Graphics* 40, 4 (2021), 1–24.

Jiajun Wu, Chengkai Zhang, Tianfan Xue, Bill Freeman, and Josh Tenenbaum. 2016. Learning a probabilistic latent space of object shapes via 3d generative-adversarial modeling. In *Proc. Conf. on Neural Information Processing Systems*, Vol. 29.

Rundi Wu, Chang Xiao, and Changxi Zheng. 2021. DeepCAD: A Deep Generative Network for Computer-Aided Design Models. In *Proc. IEEE/CVF Conf. on Computer Vision & Pattern Recognition*. 6752–6762.

Jianfeng Xiang, Zelong Lv, Sicheng Xu, Yu Deng, Ruicheng Wang, Bowen Zhang, Dong Chen, Xin Tong, and Jiaolong Yang. 2025. TRELLIS: Structured 3D Latents for Scalable and Versatile 3D Generation. In *Proc. IEEE/CVF Conf. on Computer Vision & Pattern Recognition*. 21469–21480.

Xiang Xu, Pradeep Jayaraman, Joseph Lambourne, Yilin Liu, Durvesh Malpure, and Pete Meltzer. 2025. AutoBrep: Autoregressive B-Rep Generation with Unified Topology and Geometry. In *Proc. SIGGRAPH Asia*. 94:1–94:12 pages.

Xiang Xu, Pradeep Kumar Jayaraman, Joseph George Lambourne, Karl D. D. Willis, and Yasutaka Furukawa. 2023. Hierarchical Neural Coding for Controllable CAD Model Generation. In *Proc. Int. Conf. on Machine Learning*, Vol. 202. 38443–38461.

Xiang Xu, Joseph G. Lambourne, Pradeep Kumar Jayaraman, Zhengqing Wang, Karl D. D. Willis, and Yasutaka Furukawa. 2024. BrepGen: A B-rep Generative Diffusion Model with Structured Latent Geometry. *ACM Trans. on Graphics (Proc. SIGGRAPH)* 43, 4 (2024), 119:1–119:14.

Xiang Xu, Karl DD Willis, Joseph G Lambourne, Chin-Yi Cheng, Pradeep Kumar Jayaraman, and Yasutaka Furukawa. 2022. SkexGen: Autoregressive Generation of CAD Construction Sequences with Disentangled Codebooks. In *International Conference on Machine Learning*. PMLR, 24698–24724.

Siming Yan, Zhenpei Yang, Chongyang Ma, Haibin Huang, Etienne Vouga, and Qixing Huang. 2021. HPNet: Deep Primitive Segmentation Using Hybrid Representations. In *Proc. Int. Conf. on Computer Vision*. 2733–2742.

Biao Zhang, Jiapeng Tang, Matthias Niessner, and Peter Wonka. 2023. 3DShape2VecSet: A 3D shape representation for neural fields and generative diffusion models. *ACM Trans. on Graphics (Proc. SIGGRAPH)* 42, 4 (2023), 92:1–92:16.

Zibo Zhao, Zeqiang Lai, Qingxiang Lin, Yunfei Zhao, et al. 2025. Hunyuan3D 2.0: Scaling Diffusion Models for High Resolution Textured 3D Assets Generation. arXiv:2501.12202 [cs.CV]