SKETCH2BREP: BOUNDARY-FIRST DIFFUSION FOR SKETCH-TO-BREP CAD GENERATION

Anonymous authors
Paper under double-blind review

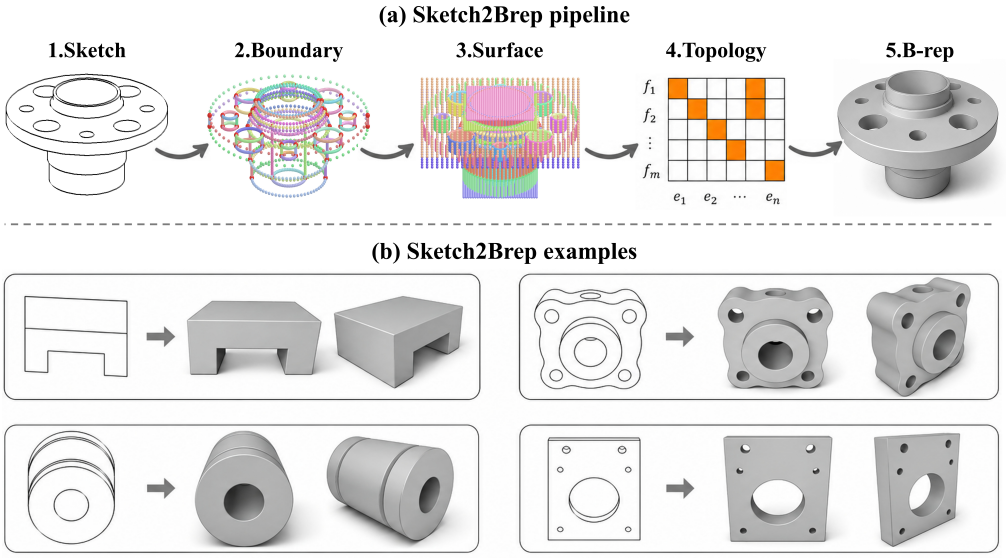

Figure 1: Sketch2BRep pipeline and examples. (a) Boundary-first B-Rep generation from a 2D sketch to boundary, surface, topology, and a B-Rep. (b) Representative sketch-conditioned results shown from two views.

ABSTRACT

Generating B-Rep CAD models from sketches is challenging because sparse 2D drawings provide only partial boundary cues, while B-Reps require both continuous geometric primitives and discrete topological relations. Existing methods typically infer coupled geometry and topology from sparse sketches, often failing to produce B-Reps that are both geometrically precise and topologically valid. We observe that boundary curves naturally bridge sketches and B-Reps: sketches provide intuitive cues for object boundaries, while boundaries reveal faces of a B-Rep. Motivated by this, we propose Sketch2BRep, a boundary-first hierarchical latent diffusion framework for sketch-conditioned B-Rep generation. Sketch2BRep decomposes the generation process into three progressive stages: sketch-conditioned boundary generation, boundary-conditioned surface generation, and explicit boundary-surface relation modeling. Specifically, a Shape-Layout Coupled Sketch Diffusion employs interacting curve-shape and spatial-layout branches to infer a plausible structured 3D boundary conditioned on the partial boundary cues in the sketch. A Boundary-Conditioned Joint Surface Diffusion enables direct interaction between face and edge tokens to synthesize surfaces consistent with the recovered boundaries. A Geometry-Aware RelationNet models pairwise boundary-surface relationships to explicitly assemble the B-Rep topology. Experiments demonstrate that the proposed method, by following the intrinsic boundary–surface–topology construction logic of B-Reps, outperforms state-of-the-art methods in geometric fidelity, topological validity, and sketch consistency.

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

1 INTRODUCTION

Sketches efficiently communicate shape intent (Yang et al., 2026), yet translating a sparse drawing into a CAD model remains difficult. Modern CAD systems generally adopt Boundary Representation (B-Rep) (Liu et al., 2026b), in which parametric faces, edges, and topological relations jointly define a solid (Koch et al., 2019; Willis et al., 2021; Jayaraman et al., 2021; Lambourne et al., 2021). B-Reps support precise editing and manufacturing, but their generation requires both continuous geometry and discrete topology to remain mutually consistent (Hao et al., 2026). This coupling makes sparse-sketch conversion particularly challenging.

Previous B-Rep generators are mainly unconditional (Jayaraman et al., 2023; Xu et al., 2024) and generate reasonable shapes without a specified target, limiting their ability to follow a particular design intent. Conditional approaches instead use point clouds or multi-view images to provide explicit geometric evidence and reduce target ambiguity (Wu et al., 2025; Yu et al., 2026). However, dense point clouds typically require 3D scanning, while consistent multi-view observations require images captured from multiple viewpoints, making these inputs costly and inconvenient for early-stage design. In contrast, sketches provide a lightweight and intuitive way to show shape intent. Recently, HoLa introduces sketch-conditioned B-Rep generation through a holistic surface latent (Liu et al., 2025). However, this is the only research available, so sketch–to–BRep generation remains insufficiently explored.

Generating a complete B-Rep from a sparse sketch is challenging because the sketch provides only partial visible boundaries (Zang et al., 2026; Chen et al., 2026), whereas a B-Rep requires both continuous geometric primitives and discrete topological relations. The gap between these two types of information makes it difficult for a network to learn the mapping from incomplete inputs solely through data-driven supervision. A natural strategy is to decouple the task into several subproblems. We observe that boundary curves serve as a natural bridge between sketches and B-Reps. Sketches provide intuitive cues about object boundaries, while boundaries explicitly reveal the faces of a B-Rep. Motivated by this insight, we propose to factorize sketch-to-B-Rep generation into a sequential process, namely sketch $\rightarrow$ boundary $\rightarrow$ surface $\rightarrow$ topology.

In this paper, we introduce Sketch2BRep, a boundary-first hierarchical latent diffusion framework, as shown in Fig. 1(a). First, Shape-Layout Coupled Sketch Diffusion employs interacting curve-shape and spatial-layout branches to infer a plausible structured 3D boundary conditioned on the partial boundary cues in the sketch. Next, Boundary-Conditioned Joint Surface Diffusion enables direct interaction between face and edge tokens to synthesize surfaces consistent with the recovered boundaries. Finally, Geometry-Aware RelationNet models pairwise boundary–surface relationships to explicitly assemble the B-Rep topology. With the boundary-to-surface-to-topology pipeline, Sketch2BRep models sketch-aligned geometric structures alongside their topological relationships and generates valid B-Rep solids, as shown in Fig. 1(b).

The main contributions are summarized as follows:

- We propose a boundary-first hierarchical latent diffusion framework that progressively generates boundaries, surfaces, and explicit boundary–surface relations from sketches.
- Shape-Layout Coupled Sketch Diffusion employs interacting curve-shape and spatial-layout branches to infer a plausible structured 3D boundary from the sketch.
- Boundary-Conditioned Joint Surface Diffusion enables direct interaction between face and edge tokens to synthesize surfaces consistent with the recovered boundaries.
- Geometry-Aware RelationNet models boundary–surface relationships to explicitly assemble the B-Rep topology.

2 RELATED WORK

2.1 UNCONDITIONAL CAD GENERATION

Unconditional CAD generation synthesizes valid B-Rep models by learning geometric and topological distributions without target-specific guidance. UV-Net (Jayaraman et al., 2021) and BRepNet (Lambourne et al., 2021) establish neural representations of native B-Rep geometry and connec-

tivity, demonstrating that parametric entities and their incidence can be encoded without conversion to meshes or point clouds. SolidGen autoregressively generates vertices, edges, and faces (Jayaraman et al., 2023), while BrepGen applies cascaded diffusion to structured B-Rep latents (Xu et al., 2024). These methods establish the feasibility of complete B-Rep generation, but errors in geometry can propagate to topology. Later methods improve consistency through different generation orders and representations. DTGBrepGen predicts adjacency before B-spline geometry (Li et al., 2025a). Stitch-A-Shape alternates geometric and topological inference (Li et al., 2025b). AutoBrep interleaves geometry with topological references (Xu et al., 2025). GraphBRep predicts surface adjacency before shared boundary curves (Lai et al., 2026), BrepGiff incorporates topology-aware face graphs (Guo et al., 2025), and BrepVGAE encodes the complete B-Rep in a global graph latent (Guo et al., 2026). HiFi-BRep further adds differentiable manifold constraints during decoding (Hou et al., 2026). Recent methods explore volumetric, compositional-cell, and topology-guided diffusion formulations (Zhang et al., 2025; Lu et al., 2026; Liu et al., 2026a). BrepDiff reconstructs topology from surface grids (Lee et al., 2025), BrepForge generates wireframes before surfaces (Li et al., 2026). Across these methods, geometry–topology consistency remains central, and most methods target unconditional generation.

2.2 Conditional CAD Generation

Conditional CAD generation reconstructs valid models from single images, point clouds, text, or multi-view images. SketchGraphs organizes 2D entities and constraints as graphs, while SketchGen and Vitruvion generate constrained sketch primitives (Seff et al., 2020; Para et al., 2021; Seff et al., 2021). Other procedural methods represent construction histories as sketch-and-extrude command sequences (Ganin et al., 2021; Wu et al., 2021; Xu et al., 2022; 2023; Fan et al., 2025). Although editable, their predefined primitives and operations limit the range of representable geometry, and a valid 2D parametric sketch does not determine the hidden geometry and topology of a complete 3D B-Rep. Geometry-based methods infer extrusion structures from point clouds (Uy et al., 2022; Ren et al., 2022; Li et al., 2023), assuming explicit 3D measurements and extrusion-compatible shapes. CADDreamer optimizes image-derived primitives into a B-Rep (Li et al., 2025c), CADCrafter predicts CAD command sequences from images (Chen et al., 2025), and BrepGaussian recovers B-Rep entities using multi-view consistency (Yu et al., 2026). Although effective for their respective inputs, these methods benefit from depth, surface appearance, or cross-view correspondences that sparse line sketches do not provide. More directly related, HoLa supports sketch conditioning through a holistic surface latent but recovers boundary curves only during surface decoding, leaving sketch boundaries implicit during generation (Liu et al., 2025). CMT combines cascaded masked autoregression with direct topology prediction from compact tokens, enabling B-Rep generation from text, point clouds, and multi-view images (Wu et al., 2025).

Overall, few studies have explored B-Rep CAD generation from sparse sketches. The existing methods typically infer coupled geometry and topology directly from sketch cues, often failing to produce B-Reps that are both geometrically precise and topologically valid. In this paper, Sketch2BRep models sketch-aligned geometric structures alongside their topological relationships through a boundary-to-surface-to-topology pipeline, thereby generating valid B-Rep solids consistent with the input sketch.

3 Method

3.1 Overview

Given a sketch I_s, Sketch2BRep generates a B-Rep in the order of boundary–surface–topology, as shown in Fig. 2. Shape-Layout Coupled Sketch Diffusion maps noisy edge layout and edge point latents to $\mathbf{B}_0$ and refines them into boundary tokens $\mathbf{B}$ under sketch guidance. Conditioned on $\mathbf{B}$, Boundary-Conditioned Joint Surface Diffusion maps noisy face layout and face point latents to $\mathbf{S}_0$ and produces surface tokens $\mathbf{S}$. Geometry-Aware RelationNet then predicts the adjacency of every face–edge pair. These pairwise predictions form $\mathcal{A}_{fe}$, from which the final B-Rep $\mathcal{M}$ is reconstructed:

$$I_s \rightarrow \mathbf{B} \rightarrow \mathbf{S} \rightarrow \mathcal{A}_{fe} \rightarrow \mathcal{M}. \tag{1}$$

Here, $\mathbf{B}$ and $\mathbf{S}$ retain explicit boundary and surface geometry, respectively, so sketch cues are expressed before topology is assembled.

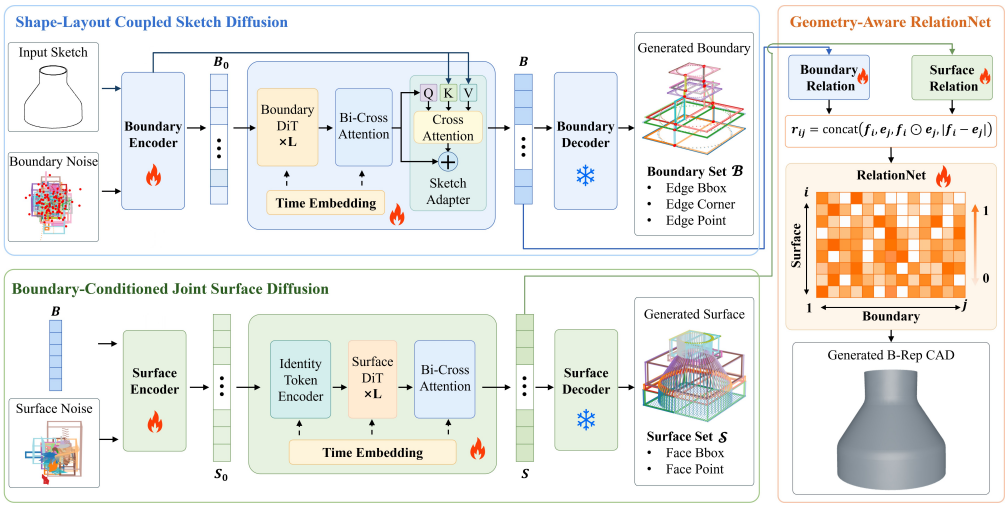

Figure 2: The framework of Sketch2BRep. Flame and snowflake symbols denote trainable modules and pretrained modules.

3.2 Structured Geometry Representation

To make CAD geometry amenable to latent diffusion, each boundary curve and surface patch is represented by a compact structured token. The j-th edge token contains the edge point latent of a curve, its two corner coordinates, and a bounding box:

$$e_j = [z_j^e, v_j^e, b_j^e], \tag{2}$$

where z_j^e is obtained by encoding uniformly sampled edge points with a pretrained 1D VAE, v_j^e denotes the corners (endpoints), and b_j^e is the edge bounding box. Following Fig. 2, the ordered boundary set is denoted by $\mathcal{B} = \{e_j\}_{j=1}^{N_e}$.

Similarly, the i-th face token contains the face point latents of a face and a bounding box:

$$f_i = [z_i^f, p_i^f], \tag{3}$$

where z_i^f is obtained by encoding a regularly sampled face point grid with a pretrained 2D VAE and p_i^f is the face bounding box. The generated surface set is $\mathcal{S} = \{f_i\}_{i=1}^{N_f}$. Separating point-based geometry from bounding-box layout keeps the diffusion variables compact while preserving the information needed for reconstruction.

The edge and face VAEs are pretrained and frozen (Xu et al., 2024). Their decoders recover edge point sequences and face point grids from generated latents. The Boundary and Surface Encoders in Fig. 2 project these structured variables into model-space sequences $\mathbf{B}_0$ and $\mathbf{S}_0$, whereas the corresponding decoders map $\mathbf{B}$ and $\mathbf{S}$ back to explicit geometry.

3.3 Shape-Layout Coupled Sketch Diffusion

The first stage recovers complete 3D boundary curves from the partial edge clue in I_s. A frozen DINOv2 (Oquab et al., 2024) extracts sketch features, and a trainable sketch projection produces the keys and values used by the Sketch Adapter:

$$[K_s, V_s] = P_s(\Phi_s(I_s)). \tag{4}$$

Given a clean boundary set $\mathcal{B}$, Gaussian noise produces $\widetilde{\mathcal{B}}_t$ at timestep t (Ho et al., 2020). The Boundary Encoder separately projects its edge bounding boxes, corners, and point latents. The two layout projections are combined and concatenated with the point-latent projection to form the initial model tokens $\mathbf{B}_0$, as detailed in Fig. 3(a).

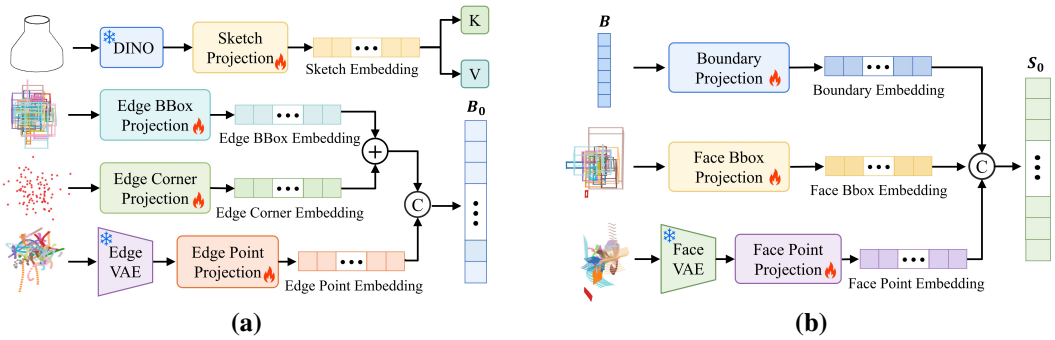

Figure 3: Boundary and surface encoders. (a) The Boundary Encoder fuses projected edge boxes, corners, point latents, and DINO sketch features into $\mathbf{B}_0$. (b) The Surface Encoder fuses boundary tokens, face boxes, and face point latents into $\mathbf{S}_0$.

The Boundary DiT contains two transformer branches. A curve-shape branch models edge point embedding, and a spatial-layout branch processes edge bounding box and corner embedding. Let $\mathbf{H}^z$ and $\mathbf{H}^p$ denote their respective outputs. A bidirectional cross-attention layer updates each branch using the keys and values of the other:

$$\begin{aligned}\mathcal{A}(Q,K,V,M) &= \mathrm{softmax}\Big(\tfrac{QK^\top}{\sqrt{d_h}} + M\Big)\,V,\\ \widetilde{\mathbf{H}}^p &= \mathbf{H}^p + W_o^p\,\mathcal{A}(Q^p,K^z,V^z,M),\\ \widetilde{\mathbf{H}}^z &= \mathbf{H}^z + W_o^z\,\mathcal{A}(Q^z,K^p,V^p,M).\end{aligned} \tag{5}$$

Each branch is first normalized and linearly projected into queries, keys, and values. In each direction, one branch queries the other under the padding mask M. The attended features are then projected and added back through a residual connection. Two Sketch Adapters subsequently inject K_s and V_s into the updated branches. The output $\mathbf{B}$ therefore couples curve shape and spatial extent while retaining stroke-level conditioning. The boundary denoiser is trained with

$$\mathcal{L}_{bound} = \mathbb{E}_{\mathcal{B},t,\epsilon}\left[\left\|\epsilon - \epsilon_\theta^e(\widetilde{\mathcal{B}}_t, t, K_s, V_s)\right\|_2^2\right]. \tag{6}$$

At inference, the stage starts from Gaussian boundary variables and iteratively predicts $\mathbf{B}$. The Boundary Decoder maps the denoised edge bounding boxes, edge corners, and edge point latents to the explicit boundary set $\mathcal{B}$.

3.4 Boundary-Conditioned Joint Surface Diffusion

The second stage generates surface patches under the structural guidance of $\mathbf{B}$. Given a clean surface set $\mathcal{S}$, Gaussian noise produces $\widetilde{\mathcal{S}}_t$. As shown in Fig. 3(b), the Surface Encoder projects the boundary tokens, noisy face bounding boxes, and noisy face points into a shared feature space. The projected boundary and face bbox embeddings are fused first and then concatenated with the to face point embeddings form $\mathbf{S}_0$.

An Identity Token Encoder assigns a learnable identity embedding to each token slot:

$$\mathbf{S}_0 = \mathrm{Concat}_{r\in\{z,p,b\},\,k}\left(P_r(\mathbf{u}_k^r) + \mathbf{e}_k^r\right). \tag{7}$$

Here, $\mathbf{u}_k^r$ denotes face point latents, face bounding boxes, or boundary tokens, P_r is its input projection, and $\mathbf{e}_k^r$ is the corresponding learnable identity embedding. The resulting identity-aware sequence lets the Surface DiT distinguish candidate entities while modeling their dependencies. Then bidirectional cross-attention exchanges information between boundary and surface features, allowing each face to query the edges that may delimit it throughout denoising. The training objective is

$$\mathcal{L}_{surf} = \mathbb{E}_{\mathcal{S},\mathbf{B},t,\epsilon}\left[\left\|\epsilon - \epsilon_\theta^f(\widetilde{\mathcal{S}}_t, \mathbf{B}, t)\right\|_2^2\right]. \tag{8}$$

The output $\mathbf{S}$ retains the generated face point latents and face bounding boxes. The Surface Decoder reconstructs normalized surface point grids with face point latents and maps them to world coordinates with the predicted face bounding boxes, yielding $\mathcal{S}$.

3.5 GEOMETRY-AWARE RELATIONNET

The first two stages generate geometry but do not specify which edges bound each face. Geometry-Aware RelationNet therefore formulates topology recovery as pairwise face–edge adjacency prediction. Given $\mathbf{B}$ and $\mathbf{S}$, its Boundary Relation Encoder and Surface Relation Encoder produce context-aware edge and face features:

$$h_j^e = \Phi_B(\mathbf{B})_j, \qquad h_i^f = \Phi_S(\mathbf{S})_i. \tag{9}$$

For every face–edge pair, their embeddings, element-wise product, and absolute difference are concatenated:

$$r_{ij} = \left[h_i^f,\, h_j^e,\, h_i^f \odot h_j^e,\, \left|h_i^f - h_j^e\right|\right], \qquad A_{ij} = \sigma\left(g(r_{ij})\right). \tag{10}$$

where A_{ij} denotes the predicted probability that edge j bounds face i, $g(\cdot)$ is the pair classifier, and $\sigma(\cdot)$ is the sigmoid function. All pairwise predictions A_{ij} yields the face–edge adjacency matrix $\mathcal{A}_{fe}$. Geometry-Aware RelationNet is trained with binary cross-entropy:

$$\mathcal{L}_{\text{rel}} = -\sum_{i,j}\left[A_{ij}^{\text{gt}} \log A_{ij} + \left(1 - A_{ij}^{\text{gt}}\right) \log\left(1 - A_{ij}\right)\right], \tag{11}$$

where $A_{ij}^{\text{gt}} \in \{0, 1\}$ denotes the ground-truth face–edge adjacency.

The explicit relation encoders let Geometry-Aware RelationNet evaluate compatibility between independently generated edges and faces instead of deriving topology through duplicated entities or surface intersections.

3.6 B-REP RECONSTRUCTION

The preceding stages produce geometric primitives and face–edge adjacency estimates rather than an explicit B-Rep. Reconstruction converts these outputs into a solid by completing connectivity, recovering shared vertices, and assembling trimmed faces.

First, if thresholding leaves a face with too few incident edges to form a closed boundary, the highest-ranked remaining edges are added as supplementary candidates, reducing open trimming loops caused by isolated false negatives. Next, edge–vertex adjacency is recovered by clustering spatially proximate corners decoded from $\mathbf{B}$ into shared vertices. Finally, the decoded edge sequences and surface grids are fitted with parametric curves and surfaces. The selected edges are ordered into boundary loops to trim each surface, and the resulting faces are sewn through shared edges and clustered vertices to produce the final B-Rep $\mathcal{M}$.

4 EXPERIMENTS

4.1 EXPERIMENTAL SETUP

Datasets. Experiments are conducted on the DeepCAD (Wu et al., 2021) and ABC (Koch et al., 2019) datasets. Following BrepGen (Xu et al., 2024) and HoLa (Liu et al., 2025), the same preprocessing procedure is adopted for both datasets. Specifically, duplicated and multi-body models are removed, closed periodic faces are split at their seams, and samples with nearly duplicated face or edge bounding boxes are discarded using a threshold of 0.05 after scaling the normalized coordinates by 3. Since neither dataset provides sketch–B-Rep pairs, paired sketches are generated from STEP geometry by applying hidden-line removal under a fixed orthographic view and rendering visible edges and silhouettes as black strokes on a white background.

Implementation Details. We implement Sketch2BRep in PyTorch and train its components sequentially with mixed precision (Micikevicius et al., 2018). Input sketches are resized to 656×528, and a frozen DINOv2-Base (Oquab et al., 2024) extracts the conditioning features. The edge and surface VAEs (Xu et al., 2024) are pretrained and frozen, followed by noise-prediction training of the two diffusion stages and binary cross-entropy training of Geometry-Aware RelationNet. For unconditional generation, we remove the sketch-conditioning branch and train a separate model without sketch inputs. Experiments are conducted on four NVIDIA RTX 4090 GPUs. More implementation details are provided in Appendix A.1.

Table 1: Unified quantitative evaluation of sketch-conditioned and unconditional B-Rep generation. The best result for each dataset and metric is shown in bold.

(a) Sketch-conditioned generation

Setting		Chamfer ↓ $\times 10^{-1}$			Geometry ↑			Topology ↑		Sketch ↑		Val. ↑
Dataset	Method	Vertex	Edge	Face	Vertex	Edge	Face	FE	EV	CLIP	SSIM	
DeepCAD	HoLa	0.77	0.22	0.13	0.77	0.79	**0.84**	0.79	0.78	0.81	0.73	86.28%
	Ours	**0.76**	**0.18**	**0.11**	**0.79**	**0.82**	0.83	**0.81**	**0.80**	**0.86**	**0.81**	**88.71%**
ABC	Ours	**0.83**	**0.25**	**0.14**	**0.75**	**0.79**	**0.80**	**0.76**	**0.73**	**0.83**	**0.79**	**66.84%**

(b) Unconditional generation

Dataset	Method	COV ↑	MMD ↓	JSD ↓	CC ↑	MC ↑	Val. ↑
DeepCAD	DeepCAD	69.27%	0.0125	0.0116	13.53	0.62	50.82%
	BrepGen	70.61%	0.0112	0.0064	11.23	1.24	47.74%
	HoLa	78.87%	0.0106	**0.0053**	14.22	2.07	82.68%
	Ours	**79.91%**	**0.0097**	0.0056	**15.71**	**2.97**	**86.37%**
ABC	BrepGen	67.35%	0.0112	0.0098	13.14	1.77	32.68%
	HoLa	73.83%	0.0109	0.0069	19.78	2.93	60.46%
	Ours	**75.31%**	**0.0103**	**0.0057**	**20.62**	**3.58**	**65.83%**

Table 2: Ablation study on DeepCAD dataset.

Setting			Chamfer ↓ $\times 10^{-1}$			Geometry ↑			Topology ↑		Sketch ↑		Val. ↑
A1	A2	A3	Vertex	Edge	Face	Vertex	Edge	Face	FE	EV	CLIP	SSIM	
✓	✓	✓	**0.76**	**0.18**	**0.11**	**0.79**	**0.82**	**0.83**	**0.81**	**0.80**	**0.86**	**0.81**	**88.71%**
✗	✓	✓	0.91	0.45	0.48	0.64	0.65	0.70	0.72	0.70	0.73	0.70	82.27%
✓	✓	✗	0.83	0.32	0.30	0.68	0.71	0.78	0.76	0.73	0.74	0.72	79.51%
✓	✗	✓	0.80	0.27	0.25	0.70	0.74	0.79	0.77	0.75	0.76	0.73	81.13%
✓	✗	✗	0.95	0.51	0.56	0.58	0.55	0.61	0.63	0.59	0.60	0.59	74.60%
✗	✓	✗	0.97	0.58	0.63	0.52	0.50	0.55	0.57	0.58	0.56	0.52	72.16%
✗	✗	✓	0.95	0.53	0.59	0.60	0.53	0.64	0.59	0.62	0.63	0.60	76.25%
✗	✗	✗	1.32	0.88	0.71	0.47	0.40	0.45	0.41	0.46	0.43	0.45	68.41%

Metrics. Following HoLa (Liu et al., 2025), sketch-conditioned evaluation uses bidirectional **Chamfer Distance (CD)** ↓ and **Geometry F1** ↑ for matched vertices, edges, and faces, with a 0.1 F1 threshold. Tables report CD with a 10^{-1} factor, whereas the text uses unscaled values. **Topology F1** ↑ evaluates Face–Edge and Edge–Vertex incidence after matching. **CLIP** ↑ and **SSIM** ↑ measure semantic and structural agreement with a same-view rendering of the generated B-Rep.

For unconditional generation, **COV** ↑ and **MMD** ↓ assess coverage and fidelity, **JSD** ↓ measures distributional discrepancy, and **CC** ↑ and **MC** ↑ characterize topological cycles and surface curvature. In both settings, **Validity (Val.)** ↑ is the percentage of valid B-Rep outputs. All metrics are macro-averaged. Metric definitions are detailed in Appendix A.2.

4.2 Sketch-Conditioned Comparison

We compare Sketch2BRep with HoLa (Liu et al., 2025), the closest method for sketch-conditioned B-Rep generation. Since HoLa does not provide public code and pretrained weights for ABC, we conduct the direct comparison on DeepCAD and report only Sketch2BRep results on ABC.

Quantitative Comparison. The sketch-conditioned rows of Table 1 show that Sketch2BRep improves over HoLa on most evaluated metrics on DeepCAD dataset. It reduces the vertex, edge, and face Chamfer distances from 0.077, 0.022, and 0.013 to 0.076, 0.018, and 0.011, respectively, while improving vertex and edge Geometry F1 from 0.77 and 0.79 to 0.79 and 0.82 with comparable face

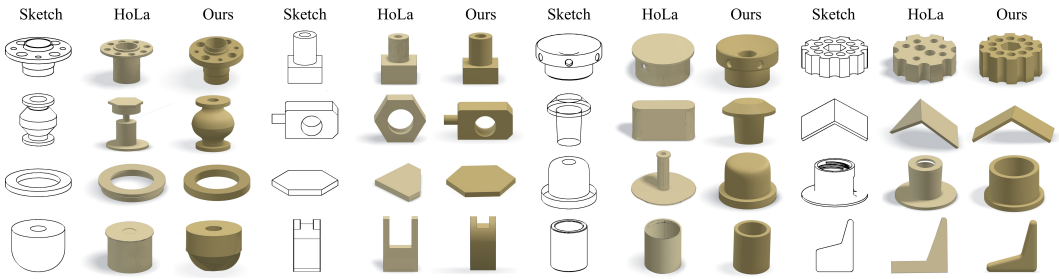

Figure 4: Qualitative comparison with HoLa on DeepCAD dataset. Each triplet shows the input sketch, the HoLa result, and our Sketch2BRep result from left to right.

accuracy. FE and EV Topology F1 both increase by 0.02, CLIP and SSIM rise from 0.81 and 0.73 to 0.86 and 0.81, and validity improves from 86.28% to 88.71%. These results demonstrate that the boundary-first hierarchy and explicit boundary–surface relation modeling jointly improve geometric fidelity, topological coherence, sketch agreement, and validity.

On ABC dataset, Sketch2BRep obtains vertex, edge, and face Chamfer distances of 0.083, 0.025, and 0.014, CLIP and SSIM scores of 0.83 and 0.79, and a validity of 66.84%. These results indicate that Sketch2BRep remains effective across the more geometrically diverse shapes in the ABC dataset.

Qualitative Comparison. Fig. 4 compares the two methods on DeepCAD sketches. HoLa generally captures the coarse object silhouette, but may omit or deform boundary-defined details in complex inputs. Sketch2BRep more faithfully preserves repeated holes, stepped profiles, gear teeth, and nested cylindrical contours, while producing fewer disconnected or floating components. The clearer closed boundaries and component connections show the benefit of explicitly recovering a complete 3D boundary before surface generation. Fig. 5 evaluates Sketch2BRep on the more geometrically diverse ABC dataset. Across two rendered views, the generated models maintain consistent openings, thin-wall layouts, extrusion directions, and junctions between planar and cylindrical surfaces. This cross-view consistency indicates that our method recovers coherent 3D structures rather than matching only the visible sketch silhouette. Additional sketch-conditioned results are provided in Appendix A.3.

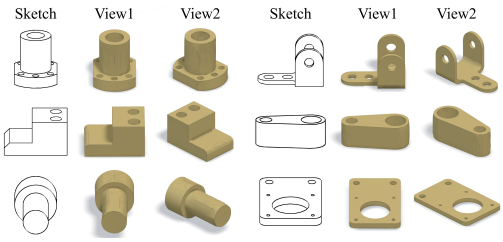

Figure 5: Conditional results on ABC.

4.3 Unconditional Comparison

We evaluate unconditional B-Rep generation on DeepCAD and ABC datasets using the six unconditional metrics defined in the experimental setup. We compare Sketch2BRep against DeepCAD (Wu et al., 2021), BrepGen (Xu et al., 2024), and HoLa (Liu et al., 2025).

Quantitative Comparison. As shown in Table 1, Sketch2BRep achieves the highest coverage on both datasets. On DeepCAD dataset, it obtains a COV of 79.91% and the lowest MMD of 0.0097, improving coverage by 1.04 percentage points and reducing MMD by 0.0009 relative to HoLa. Although its JSD of 0.0056 is slightly higher than HoLa's 0.0053, it remains lower than those of DeepCAD and BrepGen. Sketch2BRep also achieves the highest CC of 15.71, MC of 2.97, and validity of 86.37%, exceeding HoLa in validity by 3.69 percentage points. These results indicate broader coverage, greater structural richness, and more reliable generation.

On ABC dataset, Sketch2BRep outperforms HoLa across all six metrics, reaching a COV of 75.31%, MMD of 0.0103, JSD of 0.0057, CC of 20.62, MC of 3.58, and validity of 65.83%. The improvements on the more geometrically diverse ABC further demonstrate a strong balance of coverage, distributional fidelity, structural complexity, and validity.

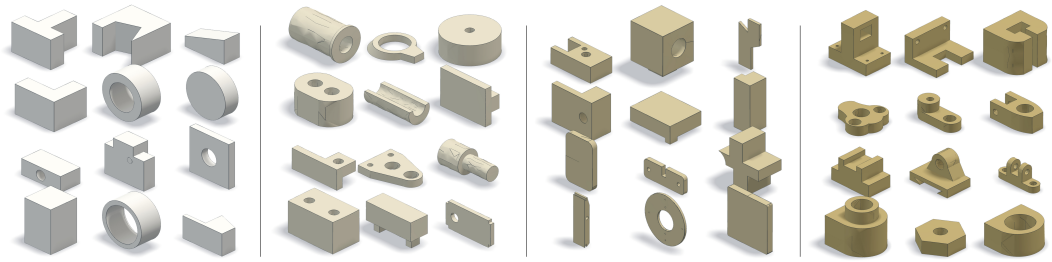

Figure 6: Qualitative comparison of unconditional B-Rep generation on DeepCAD dataset. Samples are generated by (a) DeepCAD, (b) BrepGen, (c) HoLa, and (d) Sketch2BRep (Ours).

Qualitative Comparison. Fig. 6 compares unconditional generation on DeepCAD dataset. Compared with previous methods, Sketch2BRep generates more structurally diverse and complete B-Reps, including repeated openings, stepped profiles, curved components, and nested features, while maintaining clean boundaries and coherent connectivity. Fig. 7 presents our unconditional results on ABC dataset. The generated models include complex flanges, perforated plates, recessed cavities, and rotational parts while preserving clean boundaries and coherent connectivity. These examples show that the boundary-first hierarchy increases geometric and topological richness without sacrificing solid-level coherence. Additional unconditional results are provided in Appendix A.4.

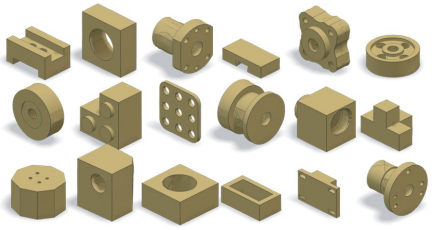

Figure 7: Unconditional results on ABC.

4.4 Ablation Studies

We conduct ablations on the DeepCAD dataset for Bi-Cross Attention (A1), the Identity Token Encoder (A2), and Geometry-Aware RelationNet (A3). As shown in Table 2, the full model performs best on every metric. Removing A1 causes the largest overall degradation, increasing the edge and face Chamfer distances from 0.018 and 0.011 to 0.045 and 0.048, which highlights the importance of bidirectional interaction for geometric consistency. CLIP and SSIM also fall from 0.86 and 0.81 to 0.73 and 0.70, indicating weaker sketch alignment. Removing A2 consistently impairs geometry, topology, and sketch agreement, demonstrating the benefit of identity-aware token representations. Without A3, FE and EV Topology F1 decrease from 0.81 and 0.80 to 0.76 and 0.73, while validity drops from 88.71% to 79.51%, underscoring its role in topology assembly. Moreover, every two-component removal underperforms its corresponding single-component removals. Disabling all three yields the worst result on every metric and only 68.41% validity. These results demonstrate their distinct and complementary contributions. Additional ablation experiments and analyses are provided in Appendix A.5 and Appendix A.6.

5 Conclusion

We introduced Sketch2BRep, a boundary-first hierarchical latent diffusion framework for sketch-conditioned B-Rep generation. The framework first recovers complete 3D boundary curves from partial sketch cues, then uses the generated boundaries to guide surface synthesis, and finally predicts explicit face–edge relations for topology assembly. Experiments on DeepCAD and ABC datasets demonstrate improved geometric fidelity, topological coherence, and sketch consistency in conditional generation, together with strong coverage and structural richness in unconditional generation. These results show that preserving explicit boundary information throughout geometry and topology generation provides an effective foundation for producing precise and coherent B-Rep CAD models from sparse sketches.

AI USE STATEMENT

AI tools were used to improve the clarity and language of this manuscript. All technical content, experimental design, analysis, and conclusions were developed and verified by the authors.

REPRODUCIBILITY STATEMENT

The model architecture and training objectives are described in the Method section, while dataset preprocessing, implementation details, training settings, and evaluation metrics are provided in the experimental setup. Additional implementation and training details are provided in the appendix.

REFERENCES

Cheng Chen, Jiacheng Wei, Tianrun Chen, Chi Zhang, Xiaofeng Yang, Shangzhan Zhang, Bingchen Yang, Chuan-Sheng Foo, Guosheng Lin, Qixing Huang, and Fayao Liu. CADCrafter: Generating computer-aided design models from unconstrained images. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 11073–11082, 2025.

Yizi Chen, Sidi Wu, Tianyi Xiao, Nina Wiedemann, and Loic Landrieu. Order matters: 3d shape generation from sequential vr sketches. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 34259–34269, 2026.

Rubin Fan, Fazhi He, Yuxin Liu, and Jing Lin. A history-based parametric cad sketch dataset with advanced engineering commands. *Computer-Aided Design*, 182:103848, 2025.

Yaroslav Ganin, Sergey Bartunov, Yujia Li, Ethan Keller, and Stefano Saliceti. Computer-aided design as language. *arXiv preprint arXiv:2105.02769*, 2021.

Hao Guo, Xiaoshui Huang, Jiacheng Hao, Yunpeng Bai, Hongping Gan, and Yilei Shi. BrepGiff: Lightweight generation of complex B-rep with 3d GAT diffusion. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 26587–26596, 2025.

Hao Guo, Liyuan Deng, Yongkang Dai, Ruohan Wang, Jiahao Li, Yunpeng Bai, and Yilei Shi. BrepVGAE: Variational graph autoencoder with unified latent representation for B-rep. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 39230–39238, 2026.

Jiacheng Hao, Chunying Liu, Hao Guo, Ruohan Wang, Hongping Gan, and Yilei Shi. Pp-brep: Few-shot b-rep classification with hybrid graph representation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 41616–41625, 2026.

Jonathan Ho, Ajay Jain, and Pieter Abbeel. Denoising diffusion probabilistic models. In *Advances in Neural Information Processing Systems*, volume 33, pp. 6840–6851, 2020.

Junhao Hou, Chenqi Luo, Pufan Wang, Jiaying Lu, Yusheng Liu, Feiwei Qin, Meie Fang, and Kun Zhou. HiFi-BRep: High-fidelity latent representation for robust B-rep generation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 27199–27208, 2026.

Pradeep Kumar Jayaraman, Aditya Sanghi, Joseph G. Lambourne, Karl D. D. Willis, Thomas Davies, Hooman Shayani, and Nigel Morris. Uv-net: Learning from boundary representations. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 11703–11712, 2021.

Pradeep Kumar Jayaraman, Joseph G. Lambourne, Nishkrit Desai, Karl D. D. Willis, Aditya Sanghi, and Nigel J. W. Morris. Solidgen: An autoregressive model for direct b-rep synthesis, 2023. URL https://arxiv.org/abs/2203.13944.

Sebastian Koch, Albert Matveev, Zhongshi Jiang, Francis Williams, Alexey Artemov, Evgeny Burnaev, Marc Alexa, Denis Zorin, and Daniele Panozzo. Abc: A big cad model dataset for geometric deep learning, 2019. URL https://arxiv.org/abs/1812.06216.

Weilin Lai, Tie Xu, Ben Niu, and Hu Wang. GraphBRep: Explicit graph diffusion of B-rep topology for efficient CAD generation. *Journal of Computational Design and Engineering*, 13(1):259–274, 2026. doi: 10.1093/jcde/qwaf136.

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

Mingi Lee, Dongsu Zhang, Clément Jambon, and Young Min Kim. BrepDiff: Single-stage B-rep diffusion model. In *Proceedings of the Special Interest Group on Computer Graphics and Interactive Techniques Conference Conference Papers*, pp. 1–11, 2025.

Jing Li, Yihang Fu, and Falai Chen. DTGBrepGen: A novel B-rep generative model through decoupling topology and geometry. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 21438–21447, 2025a.

Jing Li, Yihang Fu, and Falai Chen. BrepForge: Factorized B-rep synthesis via wireframe composition and boundary-conditioned surface instantiation. *arXiv preprint arXiv:2605.19411*, 2026.

Pu Li, Jie Guo, Xiaopeng Zhang, and Dong-Ming Yan. Secad-net: Self-supervised cad reconstruction by learning sketch-extrude operations. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 16816–16826, 2023.

Pu Li, Wenhao Zhang, Jinglu Chen, and Dongming Yan. Stitch-a-shape: Bottom-up learning for B-rep generation. In *Proceedings of the Special Interest Group on Computer Graphics and Interactive Techniques Conference Conference Papers*, pp. 1–12, 2025b. doi: 10.1145/3721238.3730661.

Yuan Li, Cheng Lin, Yuan Liu, Xiaoxiao Long, Chenxu Zhang, Ningna Wang, Xin Li, Wenping Wang, and Xiaohu Guo. CADDreamer: CAD object generation from single-view images. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 21448–21457, 2025c.

Shurui Liu, Weide Chen, and Ancong Wu. Hidigen: Hierarchical diffusion for b-rep generation with explicit topological constraints, 2026a. URL https://arxiv.org/abs/2604.02847.

Yilin Liu, Duoteng Xu, Xingyao Yu, Xiang Xu, Daniel Cohen-Or, Hao Zhang, and Hui Huang. HoLa: B-Rep generation using a holistic latent representation. *ACM Transactions on Graphics*, 44(4), 2025. doi: 10.1145/3730842.

Yilin Liu, Pradeep Jayaraman, Chinthala Reddy, Xiang Xu, and Hooman Shayani. Dualbrep: A dual-field continuous representation for b-rep modelling. In *Proceedings of the Special Interest Group on Computer Graphics and Interactive Techniques Conference Conference Papers*, pp. 1–11, 2026b.

Junran Lu, Yuanqi Li, Hengji Li, Jie Guo, and Yanwen Guo. Flatten the complex: Joint b-rep generation via compositional k-cell particles, 2026. URL https://arxiv.org/abs/2601.17733.

Paulius Micikevicius, Sharan Narang, Jonah Alben, Gregory Diamos, Erich Elsen, David Garcia, Boris Ginsburg, Michael Houston, Oleksii Kuchaiev, Ganesh Venkatesh, and Hao Wu. Mixed precision training. In *International Conference on Learning Representations*, pp. 1–12, 2018.

Maxime Oquab, Timothée Darcet, Théo Moutakanni, Huy Vo, Marc Szafraniec, Vasil Khalidov, Pierre Fernandez, Daniel Haziza, Francisco Massa, Alaaeldin El-Nouby, Mahmoud Assran, Nicolas Ballas, Wojciech Galuba, Russell Howes, Po-Yao Huang, Shang-Wen Li, Ishan Misra, Michael Rabbat, Vasu Sharma, Gabriel Synnaeve, Hu Xu, Hervé Jegou, Julien Mairal, Patrick Labatut, Armand Joulin, and Piotr Bojanowski. Dinov2: Learning robust visual features without supervision, 2024. URL https://arxiv.org/abs/2304.07193.

Wamiq Reyaz Para, Shariq Farooq Bhat, Paul Guerrero, Tom Kelly, Niloy Mitra, Leonidas J. Guibas, and Peter Wonka. Sketchgen: Generating constrained CAD sketches. In *Advances in Neural Information Processing Systems*, volume 34, pp. 5077–5088, 2021.

Daxuan Ren, Jianmin Zheng, Jianfei Cai, Jiatong Li, and Junzhe Zhang. Extrudenet: Unsupervised inverse sketch-and-extrude for shape parsing. *arXiv preprint arXiv:2209.15632*, 2022.

Ari Seff, Yaniv Ovadia, Wenda Zhou, and Ryan P. Adams. Sketchgraphs: A large-scale dataset for modeling relational geometry in computer-aided design. *arXiv preprint arXiv:2007.08506*, 2020.

Ari Seff, Wenda Zhou, Nick Richardson, and Ryan P. Adams. Vitruvion: A generative model of parametric CAD sketches. *arXiv preprint arXiv:2109.14124*, 2021.

Mikaela Angelina Uy, Yen-Yu Chang, Minhyuk Sung, Purvi Goel, Joseph Lambourne, Tolga Birdal, and Leonidas Guibas. Point2cyl: Reverse engineering 3d objects from point clouds to extrusion cylinders. *arXiv preprint arXiv:2112.09329*, 2022.

Karl D. D. Willis, Yewen Pu, Jieliang Luo, Hang Chu, Tao Du, Joseph G. Lambourne, Armando Solar-Lezama, and Wojciech Matusik. Fusion 360 gallery: A dataset and environment for programmatic cad construction from human design sequences, 2021. URL `https://arxiv.org/abs/2010.02392`.

Jianyu Wu, Yizhou Wang, Xiangyu Yue, Xinzhu Ma, Jinyang Guo, Dongzhan Zhou, Wanli Ouyang, and Shixiang Tang. CMT: A cascade MAR with topology predictor for multimodal conditional CAD generation. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 7014–7024, 2025.

Rundi Wu, Chang Xiao, and Changxi Zheng. Deepcad: A deep generative network for computer-aided design models. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 6772–6782, 2021.

Xiang Xu, Karl D. D. Willis, Joseph G. Lambourne, Chin-Yi Cheng, Pradeep Kumar Jayaraman, and Yasutaka Furukawa. Skexgen: Autoregressive generation of cad construction sequences with disentangled codebooks. In *International Conference on Machine Learning*, pp. 24698–24724, 2022.

Xiang Xu, Pradeep Kumar Jayaraman, Joseph G. Lambourne, Karl D. D. Willis, and Yasutaka Furukawa. Hierarchical neural coding for controllable cad model generation. In *International Conference on Machine Learning*, pp. 38443 – 38461, 2023.

Xiang Xu, Joseph George Lambourne, Pradeep Kumar Jayaraman, Zhengqing Wang, Karl D. D. Willis, and Yasutaka Furukawa. Brepgen: A b-rep generative diffusion model with structured latent geometry. *ACM Transactions on Graphics*, 43(4):1–14, 2024.

Xiang Xu, Pradeep Kumar Jayaraman, Joseph G. Lambourne, Yilin Liu, Durvesh Malpure, and Pete Meltzer. AutoBrep: Autoregressive B-rep generation with unified topology and geometry. In *Proceedings of the SIGGRAPH Asia Conference Papers*, pp. 1–12, 2025. doi: 10.1145/3757377.3763814.

Yi Yang, Hao Pan, Yijing Cui, Alla Sheffer, and Changjian Li. Sketch2arti: Sketch-based articulation modeling of cad objects. *ACM Transactions on Graphics (TOG)*, 45(4):1–17, 2026.

Jiaxing Yu, Dongyang Ren, Hangyu Xu, Zhouyuxiao Yang, Yuanqi Li, Jie Guo, Zhengkang Zhou, and Yanwen Guo. BrepGaussian: CAD reconstruction from multi-view images with gaussian splatting. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 26104–26113, 2026.

Ying Zang, Chunan Yu, Jiahao Zhang, Jing Li, Shengyuan Zhang, Lanyun Zhu, Chaotao Ding, Renjun Xu, and Tianrun Chen. From sketch to reality: Enabling high-quality, cross-category 3d model generation from free-hand sketches with minimal data. *IEEE Transactions on Visualization and Computer Graphics*, 32(7):5088–5097, 2026.

Fuyang Zhang, Pradeep Kumar Jayaraman, Xiang Xu, and Yasutaka Furukawa. B-rep distance functions (br-df): How to represent a b-rep model by volumetric distance functions?, 2025. URL `https://arxiv.org/abs/2511.14870`.

A Appendix

A.1 Implementation Details.

We implement Sketch2BRep in PyTorch and train its components sequentially using mixed precision. Input sketches are first resized to 656×528 with aspect-ratio-preserving padding. A frozen DINOv2-Base encoder (Oquab et al., 2024) extracts sketch features, which are precomputed and cached. First, the edge and surface VAEs (Xu et al., 2024) are pretrained and frozen to provide geometric latents. Next, Shape-Layout Coupled Sketch Diffusion and Boundary-Conditioned Joint Surface Diffusion are trained with the noise-prediction objective using 1,000 forward diffusion steps and a linear variance schedule from 10^{-4} to 0.02. The Boundary DiT in Shape-Layout Coupled Sketch Diffusion is initialized from a pretrained unconditional edge model. Its Boundary Encoder projections, Sketch Adapter, and bidirectional cross-attention layers are optimized for 1,000 epochs with a batch size of 8. To reduce the train–test mismatch in the cascaded pipeline, each downstream stage is initially trained with ground-truth upstream geometry and subsequently fine-tuned with generated outputs from the preceding stage. Specifically, the Surface DiT is first trained with ground-truth boundaries and then fine-tuned with generated boundaries. Geometry-Aware RelationNet is trained with binary cross-entropy, first using ground-truth boundaries and surfaces and then fine-tuned with their generated counterparts. All trainable modules use AdamW with an initial learning rate of 10^{-4}, weight decay of 10^{-6}, five warm-up epochs, cosine decay to 10^{-6}, and gradient clipping at 1.0. During inference, face–edge adjacency is obtained by thresholding the predicted probabilities at 0.5. All experiments are conducted on four NVIDIA RTX 4090 GPUs.

A.2 Metrics.

For sketch-conditioned generation, we follow the HoLa evaluation protocol (Liu et al., 2025) and compare each generated B-Rep with its ground truth at the geometry, topology, and conditioning levels using the following metrics:

- **Chamfer Distance (CD)** $\downarrow$ measures the bidirectional geometric discrepancy between matched primitives, reported separately for vertices, edges, and faces. Table entries are displayed with a factor of 10^{-1}, while values discussed in the text are unscaled.
- **Geometry** $\uparrow$ measures the F1 scores of matched vertices, edges, and faces under a distance threshold of 0.1.
- **Topology** $\uparrow$ measures the F1 scores of Face–Edge (FE) and Edge–Vertex (EV) incidence relations after geometric primitive matching.
- **CLIP Similarity (CLIP)** $\uparrow$ measures the cosine similarity between CLIP image features of the conditioning sketch and the generated B-Rep rendered using the same orthographic hidden-line view.
- **Structural Similarity (SSIM)** $\uparrow$ measures the grayscale structural similarity between the conditioning sketch and the corresponding rendering of the generated B-Rep.

For unconditional generation (Xu et al., 2024), we evaluate the fidelity, coverage, distribution, and structural richness of the generated B-reps using the following metrics:

- **Coverage (COV)** $\uparrow$ measures the fraction of reference shapes represented by the generated set.
- **Minimum Matching Distance (MMD)** $\downarrow$ measures the average nearest-neighbor distance between the reference shapes and the generated set.
- **Jensen–Shannon Divergence (JSD)** $\downarrow$ measures the distributional discrepancy between the generated and reference shape sets.
- **Cyclomatic Complexity (CC)** $\uparrow$ measures topological complexity by counting the independent closed loops in the wireframe topology graph of each B-Rep. A higher value indicates more holes, nested contours, and concave–convex structures.
- **Mean Curvature (MC)** $\uparrow$ measures geometric complexity by averaging the mean curvature over points sampled from all surfaces of each B-Rep. A higher value indicates richer curved geometry, including arcs and bends.



Figure 8: Sketch-conditioned B-Rep generation results of Sketch2BRep (ours).

For both settings, **Validity (Val.)** ↑ reports the percentage of generated outputs that are valid B-Reps. All other metrics are macro-averaged over the evaluated shapes unless otherwise stated.

A.3 Sketch-Conditioned Qualitative Comparison

Figure 8 shows input sketches and three views of the generated B-Reps. Sketch2BRep generates contours and openings as specified by the sketches. The consistent geometry across viewpoints verifies the coherent inference of hidden surfaces and the complete generation of 3D structures from sparse sketches.

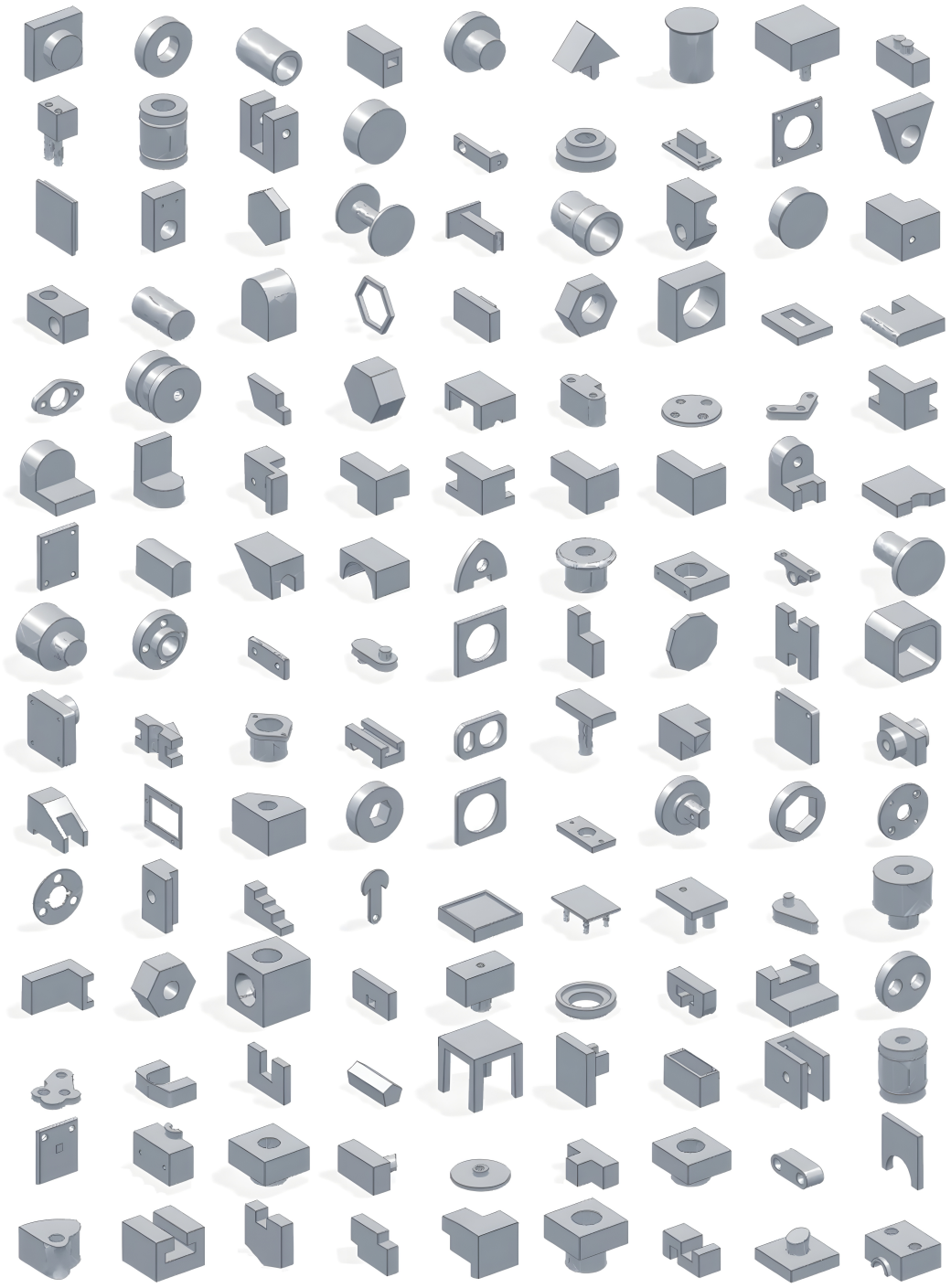

Figure 9: Unconditional B-Rep generation results of Sketch2BRep (ours).

A.4 Unconditional Qualitative Comparison

Figure 9 shows diverse unconditional samples, including prismatic and rotational parts, plates, frames, openings, cavities, and curved components. These shapes maintain clean boundaries, coherent face connectivity, and complete solid structures. Therefore, the results further demonstrate both structural richness and B-Rep consistency.

Table 3: Sketch-conditioned boundary-first versus surface-first generation on DeepCAD.

Setting		Chamfer ↓ $\times 10^{-1}$			Geometry ↑			Topology ↑		Sketch ↑		Val. ↑
Dataset	Order	Vertex	Edge	Face	Vertex	Edge	Face	FE	EV	CLIP	SSIM	
DeepCAD	Surface-First	0.85	0.32	0.27	0.66	0.72	0.68	0.65	0.68	0.71	0.70	80.33%
	Boundary-First	**0.76**	**0.18**	**0.11**	**0.79**	**0.82**	**0.83**	**0.81**	**0.80**	**0.86**	**0.81**	**88.71%**

A.5 Detailed Ablation Analysis

Table 2 shows all eight combinations of A1, A2, and A3. Among the single-component removals, disabling A1 increases the edge and face Chamfer distances from 0.018 and 0.011 to 0.045 and 0.048, decreases their Geometry F1 scores from 0.82 and 0.83 to 0.65 and 0.70, and lowers CLIP and SSIM from 0.86 and 0.81 to 0.73 and 0.70. These changes demonstrate the importance of bidirectional feature interaction for sketch-aligned geometry generation. Removing A2 increases the edge and face Chamfer distances to 0.027 and 0.025, lowers FE and EV Topology F1 to 0.77 and 0.75, and reduces validity to 81.13%, showing that identity-aware tokens help preserve entity-specific structure. Without A3, FE and EV Topology F1 decrease to 0.76 and 0.73, while validity falls to 79.51%, highlighting the role of explicit geometry-aware relations in reliable topology assembly.

The multi-component ablations further demonstrate complementarity. Every two-component removal performs worse than its corresponding single-component removals. Removing all three produces the worst result on every metric, including edge and face Chamfer distances of 0.088 and 0.071, FE and EV Topology F1 scores of 0.41 and 0.46, and a validity of 68.41%. Therefore, their combination is essential to the complete model.

A.6 Comparison of Boundary-First and Surface-First Generation

To isolate the effect of generation order, both variants use the same sketch encoder, geometric representations, training data, and topology reconstruction procedure. Boundary-first generation first recovers 3D boundary curves from the input sketch and then synthesizes surfaces conditioned on these boundaries. Surface-first generation reverses this order by generating surfaces from the sketch before predicting their boundary curves.

Table 3 compares the two generation orders on the DeepCAD dataset under the same evaluation protocol. Boundary-first generation consistently outperforms surface-first generation across all metrics. It reduces the vertex, edge, and face Chamfer distances from 0.085, 0.032, and 0.027 to 0.076, 0.018, and 0.011, while increasing the corresponding Geometry F1 scores from 0.66, 0.72, and 0.68 to 0.79, 0.82, and 0.83. FE and EV Topology F1 improve from 0.65 and 0.68 to 0.81 and 0.80, and CLIP and SSIM rise from 0.71 and 0.70 to 0.86 and 0.81. Validity also increases by 8.38 percentage points, from 80.33% to 88.71%. These results show that establishing explicit boundaries before surface synthesis provides stronger structural constraints, improving geometric accuracy, topology coherence, sketch alignment, and B-Rep validity for sketch-conditioned generation.