CADDreamer: CAD Object Generation from Single-view Images

Supplementary Material

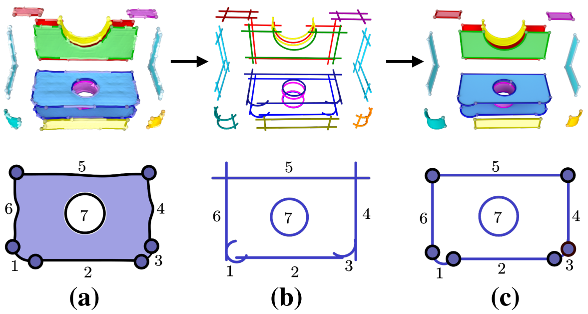

Figure A1. An illustration of the topology-preserving intersection strategy to achieve a watertight B-rep model with CAD vertices, edges, and faces. (a) depicts the topology derived from the mesh patches. (b) shows the CAD curves resulting from primitive intersections. (c) represents the reconstructed B-rep model through topology-preserving trimming.

Our project page and code can be found here. In this supplementary document, we provide more details about the evaluation metrics used in this paper, the ablation study, our limitations, and failure cases, as well as more experimental results.

A. Details of Evaluation Metrics

In this section, we provide more details about each evaluation metric used in this paper.

- **Chamfer Distance (CD).** We uniformly sample 10,000 points from both the generated and ground truth CAD models. The Chamfer Distance is the average minimum distance between these point pairs.
- **Normal Consistency (NC).** We sample 10,000 points uniformly from both the generated and ground truth CAD models. Normal consistency is then computed as the average absolute cosine similarity between corresponding point pairs, identified via nearest neighbor search.
- **Vertex-wise Segmentation Accuracy (SEG(V)).** For each case, 10,000 points are uniformly sampled from both the ground truth and the segmented mesh/CAD model, along with their primitive labels. SEG(V) is the average percentage of points whose nearest neighbor in the ground truth has a matching label.
- **Primitive-wise Segmentation Accuracy (SEG(P)).** For each case, we directly compare the number of primitives in the segmentation result. If the number of segmentation labels is correct, we consider it a correct prediction. SEG(P) is the percentage of cases with correct predictions.
- **Ratio of Hanging Faces (HF).** For each case, if a face has an edge with no neighboring face, we consider it a hanging face. We calculate the ratio between the number of hanging faces and the total number of all faces. HF is the average of these ratios across all cases.

Methods	CD (↓)	NC (↑)	SEG(V) (↑)	SEG(P) (↑)
Model 1	2.56	80.3	58.8	67.8
Model 2	1.28	92.4	63.8	69.4
Model 3	1.27	92.5	95.5	97.6
CADDreamer	**1.27**	**92.6**	**95.7**	**97.9**

Table A1. Statistical results of reconstructed meshes and extracted primitives across ablation models. Evaluation metrics include: Chamfer distance (CD $\times$100), normal consistency (NC, %), vertex-based segmentation accuracy (SEG(V), %), and primitive count-based segmentation accuracy (SEG(P), %). Best values are highlighted in bold.

Methods	HF (↓)	CD (↓)
Model 1	32.4	8.84
Model 2	29.8	8.13
Model 3	2.40	1.37
Model 4	23.5	6.47
Model 5	11.8	3.29
CADDreamer	**2.4**	**1.36**

Table A2. Statistical results of reconstructed B-reps across ablation models, including the percentage of B-reps with hanging faces (HF, %), and Chamfer distance (CD $\times$100) between the reconstructed B-reps and ground truth. Best values are highlighted in bold.

B. Ablation Study

Our method benefits primarily from two distinct technical contributions: a new multi-view generation module that yields higher-quality meshes and improved segmentation results, and a geometric optimization strategy that reconstructs essential topological and geometric relationships to ensure CAD reconstruction through primitive intersection. We conduct ablation studies on these two technical contributions to investigate the performance gains achieved independently by each module.

Generation module. We first evaluate the contributions of our generation module through three comparative models. Model 1 employs the non-finetuned Wonder3D (RGB inputs and RGB&normal outputs) combined with Point2CAD to establish a baseline for comparison with other fine-tuned solutions. Model 2 utilizes the fine-tuned Wonder3D (RGB inputs and RGB&normal outputs) with Point2CAD to assess the effectiveness of our new cross-domain generation strategy (normal&semantic map). Model 3 implements CADDreamer's pipeline with RGB inputs to evaluate the impact of our normal inputs. As demon-

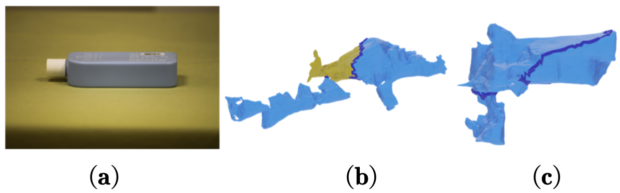

Figure A2. Using RGB images as input for fine-tuning will lead to poor generalization performance on real images. (a) real image input, (b-c) are the segmentation results of Model 2-3, which are fine-tuned with RGB inputs.

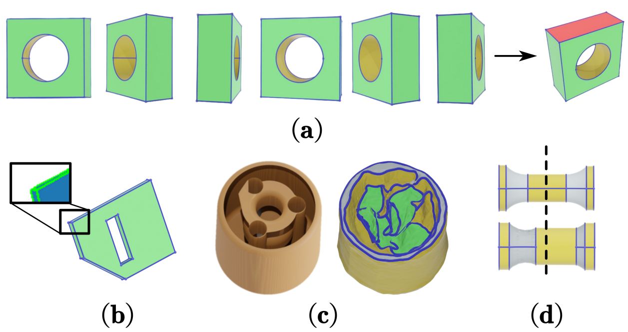

Figure A3. The limitations of our methods. (a) The multi-view images of six horizontal angles cannot cover the top face (red) of the CAD object. (b) An extremely thin and elongated plane (less than 1 pixel wide). (c) A failure case of complex geometry (original image and segmentation result). (d) Our generated B-reps cannot preserve symmetry of the CAD structure.

strated in Table A1, all models fine-tuned on primitive-based shapes outperform the original Wonder3D (Model 1), which lacks prior knowledge of primitives. Models that incorporate our cross-domain generation strategy for mesh segmentation (Models 3-5) achieve superior segmentation results compared to the post-processing segmentation approach applied to noisy meshes (Models 1-2), leading to more accurate B-reps reconstruction outcomes. Although using RGB images as inputs (Model 3) does not diminish CADDreamer's performance on synthetic data, it impairs generalization to real-world image inputs, as illustrated in Figure A2. This limitation stems from the training data's restricted texture variety and fixed lighting conditions, which fail to capture the complexity and variability of textures and illumination encountered in real-world scenarios.

Geometry optimization module. Next, we investigate the contribution of our geometric optimization strategy and the importance of enforcing parallelism, perpendicularity, and collinearity constraints. We construct two variant models: (1) *Model 4*, which removes all geometric optimization steps from CADDreamer, and (2) *Model 5*, which maintains primitive stitching but removes the three geometric constraints. Since Models 4 and 5 share the same mesh reconstruction and segmentation results with our method, we focus our analysis on comparing the reconstructed B-reps with ground truth, as shown in Table A2. Given that our method provides mesh and segmentation results closer to ground truth, the hanging faces and CD metrics of Models 4 and 5 demonstrate a significant improvement over Point2CAD (Model 1-2). Without geometric optimization, reconstruction fitting errors introduce gaps between primitives, compromising geometric relationships and causing missing or incorrect intersections (see Figure 4 in the main manuscript). Model 5, incorporating primitive stitching, effectively reduces intersection failures, thus achieving lower hanging face ratios and CD values compared to Model 4. Additionally, our model with three geometric constraints reduces reconstruction failures caused by erroneous geometric relationships, resulting in fewer hanging faces and smaller Chamfer distances than Model 5.

C. Limitations and Failure Cases

CADDreamer consists of two main modules: the generation module and the geometry optimization module. We discuss the limitations of these two modules separately.

Generation module. Constrained by limited multi-view angles and image resolutions, the generation module fails to maintain robustness under extreme viewing angles and struggles with CAD shapes containing fine geometric structures. First, CADDreamer generates mesh segmentation schemes using multi-view images, requiring the selected viewpoints to include all CAD faces. Since our method uses Wonder3D as a pretrained model with the same multi-view setup (six 256×256 images), our method requires that the CAD faces must be included in these six viewpoints. As illustrated in Figure A3(a), when the six viewpoints used for generation are all horizontal, the top face of the CAD object does not appear in any of the six views. Consequently, the top face is omitted in the final generation, leading to inaccurate topology and reconstruction failure. Second, some B-reps contain extremely thin and elongated CAD faces that, when projected into multi-view images, manifest as sub-pixel features with widths less than one pixel. As demonstrated in Figure A3(b), the elongated top face is segmented into isolated pixels by feature lines, which is not a continuous patch and cannot generate meaningful segmentation results. The combination of these limitations means that CADDreamer might fail to reconstruct and segment complex CAD shapes, especially shapes with fine structures and excessive occlusion, as shown in Figure A3(c). A potential solution to these limitations is to increase the number of views and improve image resolution: adding a top view ensures coverage of the top face, while using higher image resolution makes fine structures more visible, enabling better face coverage. However, this approach demands prohibitive computational resources and introduces greater convergence challenges.

Geometry optimization module only reconstructs basic geometric relationships between primitives, such as parallelism, perpendicularity, and collinearity. It does not consider more complex relationships between multiple primitives, such as symmetry. As shown in Figure A3(d), while the ground truth (upper) exhibits perfect symmetry between primitives, such property is difficult to be preserved in our results (lower) due to inevitable reconstruction and fitting errors. Furthermore, since detecting and optimizing such complex relationships remains an open problem, implementing their reconstruction within the geometric optimization module presents significant challenges.

D. More Comparison Results

We add additional comparative experiments to analyze the advantages of CADDreamer. First, we add the comparison with DeepCAD [3] to analyze the advantages of CADDreamer compared to sketch-extrude methods. In addition, we notice some closed-source (Tripo [2]) and open-source (Trellis [4], MeshAnything [1]) models released concurrently with CADDreamer. We analyze the performance of these methods on some typical examples and explain why these methods are not suitable for CAD generation.

Since DeepCAD cannot handle the image inputs, we use our generated meshes as inputs and use DeepCAD to generate B-reps. As shown in Figure A4, since sketch-extrude approaches can only support planes and cylinders, complex primitives like cones are simplified, and the reconstructed B-reps have higher errors (CD=4.21, NC=67.8) than CADDreamer (CD=1.36, NC=91.9) on our testing dataset.

We also evaluate Tripo, Trellis, and MeshAnything on 10 diverse CAD models for mesh reconstruction, followed by Point2CAD for CAD generation to obtain CAD models. Tripo and Trellis tend to add unnecessary details and components due to their focus on generating diverse and complex shapes, which results in a large number of segmented primitives and faces, as shown in Figure A4. MeshAnything, based on a finite-length autoregressive generation progress, overlooks strong constraints on mesh connectivity. Consequently, it tends to generate simple and topology-breaking meshes, resulting in numerous isolated primitives and inferior outcomes than CADDreamer.

E. More Reconstruction Results

Figure A5, A6, and A7 showcase more reconstruction results of our method, including input images, reconstructed meshes, reconstructed B-reps, as well as their CAD vertices and edges.

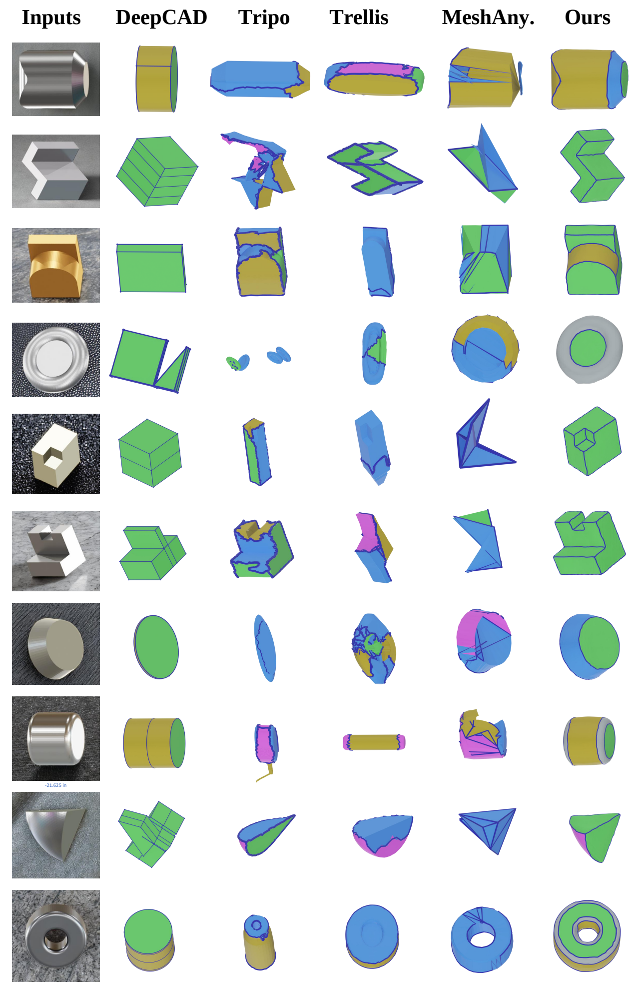

Figure A4. Comparison with other methods. From left to right, (a) Input image, (b) DeepCAD results [3], (c) Tripo results [2], (d) Trellis results [4], (e) MeshAnything results [1]), and (f) our segmentation results. Color definitions match those used in the main manuscript figures.

References

[1] Yiwen Chen, Tong He, Di Huang, Weicai Ye, Sijin Chen, Jiaxiang Tang, Xin Chen, Zhongang Cai, Lei Yang, Gang Yu, Guosheng Lin, and Chi Zhang. Meshanything: Artist-created mesh generation with autoregressive transformers. *arXiv preprint arXiv:2406.10163*, 2024. 3

[2] Dmitry Tochilkin, David Pankratz, Zexiang Liu, Zixuan Huang, Adam Letts, Yangguang Li, Ding Liang, Christian Laforte, Varun Jampani, and Yan-Pei Cao. Triposr: Fast 3d object reconstruction from a single image. *arXiv preprint arXiv:2403.02151*, 2024. 3

[3] 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*, pages 6772–6782, 2021. 3

[4] Jianfeng Xiang, Zelong Lv, Sicheng Xu, Yu Deng, Ruicheng Wang, Bowen Zhang, Dong Chen, Xin Tong, and Jiaolong Yang. Structured 3d latents for scalable and versatile 3d generation. *arXiv preprint arXiv:2412.01506*, 2024. 3

Figure A5. Reconstruction results from the given images are shown from left to right: input image, reconstructed mesh, B-reps, and their CAD vertices and edges. Different colors indicate different primitives: plane (◯), cylinder (◯), cone (◯), sphere (◯), and torus (◯).

Figure A6. Reconstruction results from the given images are shown from left to right: input image, reconstructed mesh, B-reps, and their CAD vertices and edges. Different colors indicate different primitives: plane (◯), cylinder (◯), cone (◯), sphere (◯), and torus (◯).

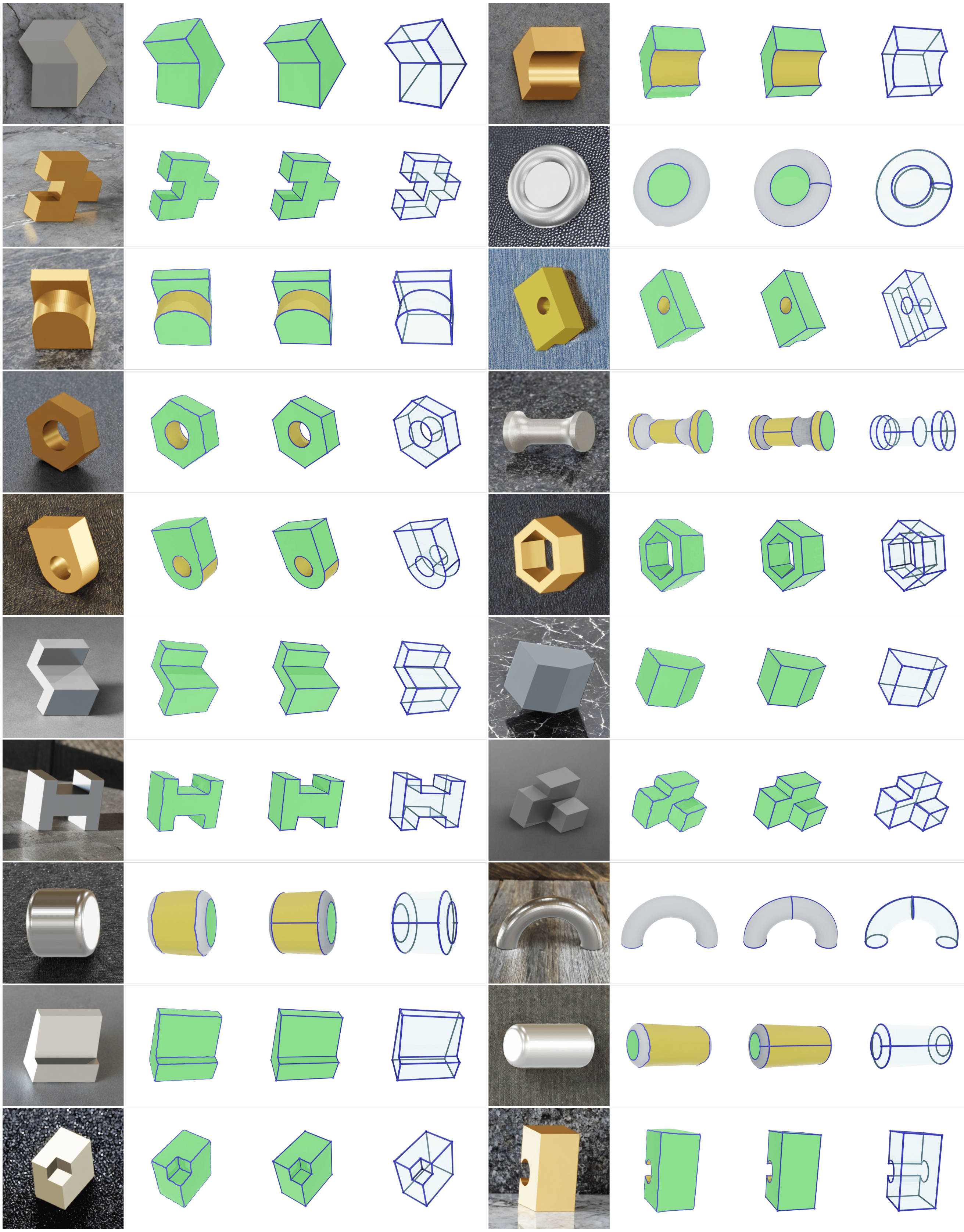

Figure A7. Reconstruction results from the given images are shown from left to right: input image, reconstructed mesh, B-reps, and their CAD vertices and edges. Different colors indicate different primitives: plane (◯), cylinder (◯), cone (◯), sphere (◯), and torus (◯).