

Technical Appendices and Supplementary Material

Overview

This supplementary material is organized into several sections that provide additional details and analysis related to our work on ViewCraft3D (VC3D). Specifically, it includes the following aspects:

- In Section A, we provide detailed implementation information for VC3D.
- In Section B, we present additional qualitative results generated by VC3D.
- In Section C, we describe a user study that demonstrates the superiority of our method compared to existing approaches.
- In Section D, we discuss the potential societal impact of VC3D.

A Implementation Details of VC3D

A.1 Algorithm Flow of Salient Point Cloud Extraction

Algorithm S1 shows the details of Salient Point Cloud Extraction (in main paper Sec. 3.2.1).

Algorithm S1: Salient Point Cloud Extraction

Input: Mesh file \mathcal{M} , Boolean variable *enable_find_silhouette_edges*

Output: Salient point cloud \mathcal{P}_s

```

 $\mathcal{P}_s \leftarrow \emptyset$ ; // Initialize the salient point cloud
 $\mathcal{P}_s \leftarrow SES\_Process(\mathcal{M})$ ; // Extract salient points using SES
if enable_find_silhouette_edges then
     $E \leftarrow Find\_Silhouette\_Edges(\mathcal{M})$ ; // Extract silhouette edges
     $\mathcal{P}_e \leftarrow Sample\_Points(E)$ ; // Sample points on silhouette edges
     $\mathcal{P}_s \leftarrow \mathcal{P}_s \cup \mathcal{P}_e$ ; // Merge point cloud
return  $\mathcal{P}_s$ ; // Return the salient point cloud

```

SES stands for Sharp Edge Sampling, a technique proposed in Dora [1] for extracting salient point clouds. The parameter *enable_find_silhouette_edges* is a hyper-parameter that controls whether silhouette edges are detected. We set this parameter to True when the mesh contains many smooth surfaces (e.g., spheres, cylinders). Figure S1 illustrates the results on the same example with *enable_find_silhouette_edges* set to both True and False for comparison.

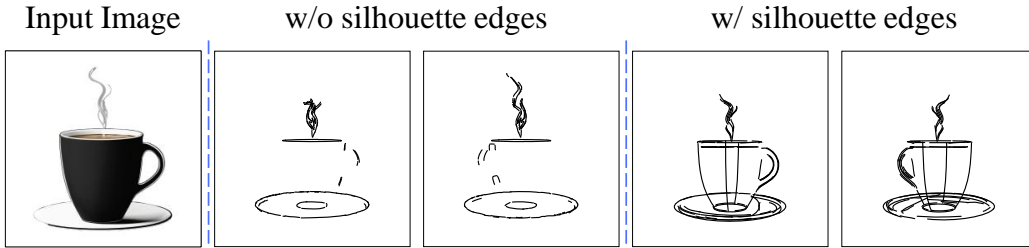


Figure S1: Effect of Silhouette Edge Extraction. Comparison of outputs with (right) and without (middle) the silhouette edge extraction enabled.

When the parameter *enable_find_silhouette_edges* is set to False, the salient edge extraction fails to capture the outline of the cup body, resulting in an incomplete representation. In contrast, enabling this parameter (i.e., setting it to True) allows the extraction process to effectively highlight the cup body, which is clearly reflected in the final result.

23 A.2 Details of Detail Refinement

24 In the Detail Refinement stage, we randomly add N Bézier curves to intricate-to-approximate regions
 25 (as defined in Sec. 3.3 of main paper). By default, N is set to 128, but it can be adjusted based on
 26 the geometric complexity of missing areas. The intricate-to-approximate regions are determined by
 27 the mesh vertices and the point cloud \mathcal{P}_s . Specifically, for a point p , if its distance to all points in
 28 the point cloud \mathcal{P}_s exceeds a threshold (set to 0.03 in our experiments), it is considered to belong
 29 to the intricate-to-approximate regions. All such points comprise a set \mathcal{P}_i , from which vertices are
 30 randomly selected to initialize the Bézier curves.

31 For both the Bézier curves from the first stage and those newly added in the second stage, we sample
 32 64 points per curve to construct the combined point cloud $\mathcal{P}_{combined}$. Based on the mesh \mathcal{M} generated
 33 in the first stage, each point in $\mathcal{P}_{combined}$ is assigned an approximate normal direction as a geometric
 34 prior. Both $\mathcal{P}_{combined}$ and corresponding normal directions are then encoded using the VAE encoder
 35 of TripoSG [2] to generate structured latent representations. To ensure stable optimization, the SDS
 36 loss [3] employs medium-strength noise with the range $[0.15, 0.4]$ and multiple denoising steps.
 37 This strategy effectively balances global shape consistency with local detail preservation, thereby
 38 enhancing the visual quality of the generated vector graphics.

39 B More Qualitative Results

40 B.1 More Qualitative Results of VC3D

41 As shown in Figure S2, VC3D is capable of generating high-quality results across a diverse range of
 42 3D object categories, such as animals, industrial products, and furniture.

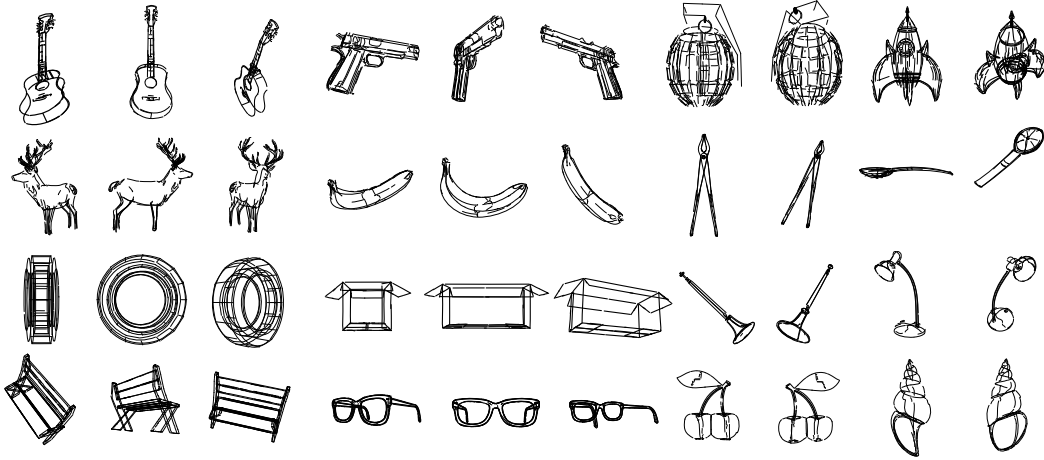


Figure S2: More qualitative results generated by our method VC3D.

43 B.2 More Comparisons with Other Methods

44 Figure S3 presents additional comparisons between our method and baseline methods. It can be
 45 observed that VC3D outperforms these methods in terms of fidelity and view-consistency.

46 C User Study

47 We conducted a user study to evaluate the quality of the synthesized 3D vector graphics. The test
 48 set used in the study is identical to that in the main paper and consists of 40 cases, each with results
 49 generated by three different approaches. For the user study, each questionnaire contained 20 cases
 50 randomly sampled from these 40.

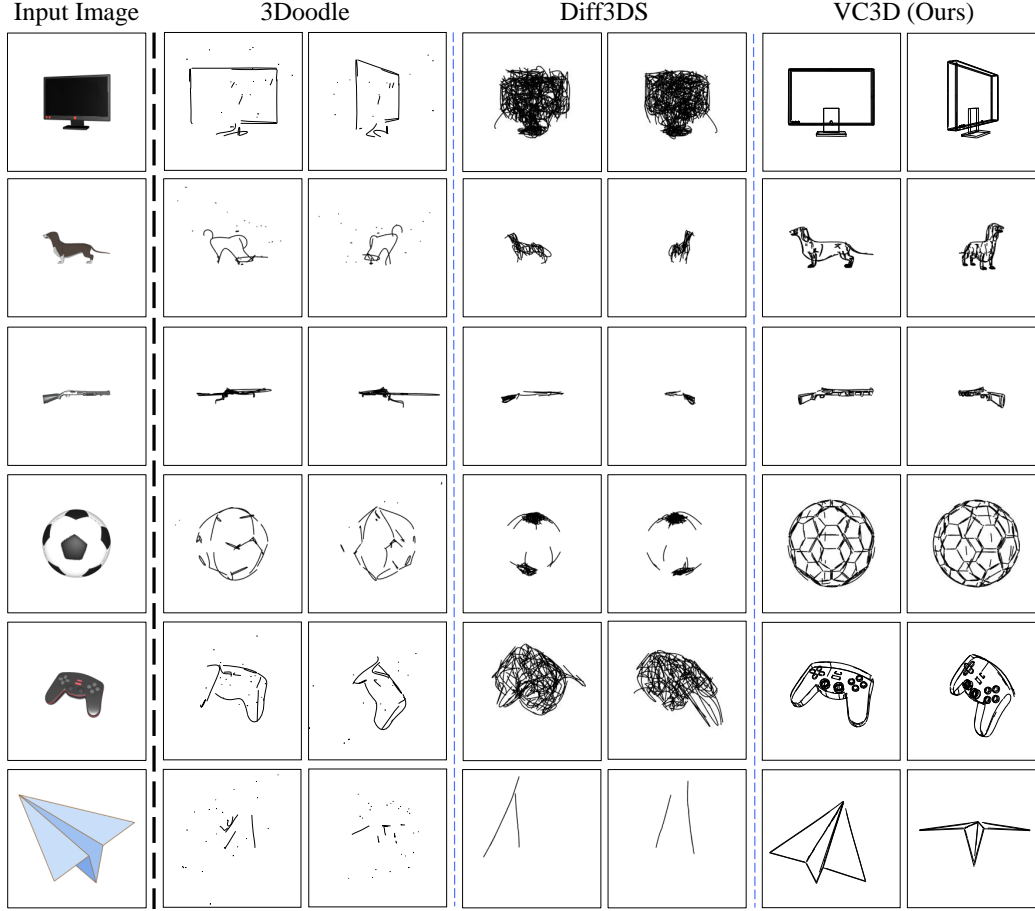


Figure S3: More qualitative comparisons between our method and baseline methods, 3Doodle and Diff3DS.

We invited 30 volunteers to participate in the study, with a gender ratio of approximately 5:1 (male to female). Most participants are university students from various science and engineering disciplines, with ages ranging from 18 to 27. Volunteers were asked to choose the best result for each case based on two criteria: fidelity and view-consistency. As shown in Fig. S4, our approach received 71.17% of the votes, outperforming 3Doodle (13.83%) and Diff3DS (15%).

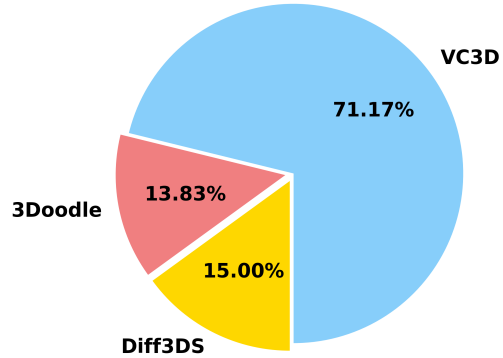


Figure S4: The results of the user study.

D Societal Impact

Our method, VC3D, significantly enhances the fidelity and view-consistency of 3D vector graphics generated from a single image. However, as with other generative models, there exists a critical concern regarding the potential negative societal impact if our model is misused to create inauthentic or misleading 3D vector content. It is imperative to ensure that the model is used responsibly to mitigate the risk of adverse social consequences.

69 **References**

- 70 [1] Rui Chen, Jianfeng Zhang, Yixun Liang, Guan Luo, Weiyu Li, Jiarui Liu, Xiu Li, Xiaoxiao
71 Long, Jiashi Feng, and Ping Tan. Dora: Sampling and benchmarking for 3d shape variational
72 auto-encoders. *arXiv preprint arXiv:2412.17808*, 2024.
- 73 [2] Yangguang Li, Zi-Xin Zou, Zexiang Liu, Dehu Wang, Yuan Liang, Zhipeng Yu, Xingchao Liu,
74 Yuan-Chen Guo, Ding Liang, Wanli Ouyang, et al. Triposg: High-fidelity 3d shape synthesis
75 using large-scale rectified flow models. *arXiv preprint arXiv:2502.06608*, 2025.
- 76 [3] Ben Poole, Ajay Jain, Jonathan T Barron, and Ben Mildenhall. Dreamfusion: Text-to-3d using
77 2d diffusion. In *The Eleventh International Conference on Learning Representations*, 2022.