Supplementary Material for GenPC: Zero-shot Point Cloud Completion via 3D Generative Priors

An Li† Zhe Zhu† Mingqiang Wei‡
Nanjing University of Aeronautics and Astronautics
{lian, zhuzhe0619, mqwei}@nuaa.edu.cn *

A. Implementation Details

Depth Prompting: To select viewpoints, we uniformly distribute 2048 virtual cameras on a spherical surface around the partial point cloud. Each camera has a resolution of 256, is positioned at a distance of 1.6 units, and has a field of view (FOV) of 49.1°. When projecting the point cloud into depth maps, we assign a pixel size of 1 for dense point clouds and 2 for sparse ones, with a consistent mask pixel size rate of 3. For depth inpainting, we make use of a pre-trained diffusion model [3] with a resolution of 256×256. Meanwhile, depth-conditioned generative modeling leverages ControlNet [7] with 30 inference steps and a conditioning scale of 0.99, generating outputs at 1024×1024 resolution.

Image-to-3D generative model: The Image-to-3D generative model, InstantMesh [5], uses a base-scale configuration with 75 inference steps, a scale of 1, a distance parameter of 4.5, and six views for rendering.

Geometric Preserving Fusion: In the Refining step, the parameter setup for G_{miss} is shown in Table 1. During the refining process, the output resolution is set to 256×256 with a batch size of 12, and the process is run for 100 iterations.

B. Failure Case and Limitation

Our method starts by obtaining depth from the scanned viewpoint of the input partial point cloud. While this approach effectively addresses the issue of missing points caused by self-occlusion in the partial point cloud, it faces challenges in scenarios where large areas of the object are heavily occluded by other objects. In such cases, the quality of the geometric details may be affected, as illustrated in Figure 1. However, by integrating both 2D and 3D generative models, our method alleviates these challenges to some extent, producing results with globally correct shapes. In contrast, SDS-Complete struggles to handle such severe occlusion challenges, often resulting in incomplete reconstructions.

*This work was supported by the National Natural Science Foundation of China (No. T2322012, No. 62172218).
† Equal contribution ‡ Corresponding author

Table 1. 3D Gaussian Parameter Setup for G_{miss}.

Parameter	Value
Initial Opacity	1
Color Learning Rate	0.005
Opacity Learning Rate	0.005
Scaling Learning Rate	0.005
Spherical Harmonics Degree	0
Initial Position Learning Rate	0.001
Final Position Learning Rate	0.0003

Table 2. Comparison of Completion Time on the Redwood Dataset. Ours* represents GenPC without Refining (ℓ^1 CD and EMD $\times 10^2$).

Methods	CD↓	EMD↓	Time↓
SDS-Complete	2.72	4.06	≈40 Hours
Ours*	1.98	3.16	≈**1 Minute**
Ours	**1.74**	**2.88**	≈2 Minutes

Another limitation of our method is its reliance on several pre-trained models, whose performance significantly affects the final completion results. However, this also indicates that: With advanced techniques and data, we can utilize more powerful pre-trained models to improve GenPC.

Table 3. Quantitative comparison on unseen categories of the ShapeNet dataset (ℓ^1 CD $\times 10^3$).

Methods	PoinTr	SnowflakeNet	AdaPoinTr	Ours
Guitar	8.96	6.37	5.11	**4.62**
pistol	10.81	8.23	7.90	**6.84**

C. Complexity Analysis

We show the complexity analysis in Table 2, where the quantitative results and completion time on the Redwood dataset are shown. Compared with the previous zero-shot method SDS-complete [4], our method significantly reduces completion time while improving completion quality. This is

Figure 1. Failure cases. When P_{in} suffers from both self-occlusion (black boxes) and external occlusion (red boxes), the external one may undermine the quality of the generated image I_{gen} and affect the overall completion performance.

Table 4. Results on the Redwood dataset with different incompleteness types (CD and EMD).

Methods	PoinTr	Snowflake	AdaPoinTr	ShapeFormer	Ours
N1	2.93/6.61	3.08/5.87	4.47/7.23	3.83/6.04	**1.97/3.37**
N2	3.16/6.93	3.18/5.93	4.64/7.62	3.94/5.89	**2.33/3.84**
DR	3.05/6.98	2.96/5.56	4.41/7.21	4.58/6.50	**1.92/3.45**
DV	3.41/7.45	3.31/6.10	4.50/7.32	4.54/5.72	**2.18/3.71**
CUT	4.83/7.76	4.24/6.01	5.07/7.89	5.23/6.81	**2.76/4.89**
SCAN3	2.81/5.25	2.67/5.32	4.33/6.73	3.43/4.69	**1.51/2.43**
SCAN5	2.60/5.98	2.60/5.26	4.22/6.79	2.96/4.10	**1.38/2.26**

Table 5. Results using different prompts (CD and EMD).

Prompt	Null	Category	Detail
CUT	4.39/6.33	2.76/4.89	**2.45/4.47**
Standard	1.84/3.10	1.74/2.88	**1.66/2.73**

Table 6. Comparison with generative method [6] (TMD and UHD).

Method	ShapeFormer	Ours
CUT	12.53/9.66	**3.84/0.95**
Standard	12.96/10.84	**2.51/0.79**

achieved by leveraging a feed-forward 3D generative model to generate an explicit 3D prior in just a few seconds. By doing so, our method avoids the need to optimize an SDF from scratch, as in SDS-Complete, thus accelerating the completion process and enhancing the quality of the results.

D. Robustness to Incompleteness Types

To evaluate the robustness of GenPC on different types of incompleteness, in Table 4, we constructed four types of degradation benchmarks for the partial point clouds in Redwood: **a)** Adding Gaussian noise with $\sigma = 0.005/0.01$ (N1/N2); **b)** Geometrically cutting 50% points (CUT); **c)** Randomly deleting 98% points and downsampling with a voxel size of 0.07 (DR/DV); **d)** Fusing depth maps from 3 and 5 consecutive viewpoints (SCAN3/5); As shown in Table 4, GenPC demonstrates robustness to point cloud sparsity and noise. While large missing regions have some impact on performance, our method still achieves SOTA results. Notably, more fused depth maps boost completeness as the input contains more structural information.

E. Effect of Text Prompt

In our main experiments, the text prompt is set as the category name. To explore the impact of the text prompt, we conduct experiments on the standard and CUT versions of Redwood under three prompt configurations: Null prompts, category prompts, and prompts with detailed structure descriptions. As shown in Table 5, the absence of text prompts has a negligible impact on completion performance in typical scenarios. However, in extreme occlusion scenarios, it leads to decreased performance. In such cases as shown in Figure 2 (left), category text prompts significantly improve the structural reasoning of occluded regions. Additionally, using more detailed prompts can produce varying structures in the missing parts as shown in Figure 2 (right).

F. Comparison with Generative Method

Comparison with the generative method ShapeFormer [6] using multi-modal metrics is shown in Table 6, where GenPC significantly outperforms ShapeFormer in both metrics. We find that structural uncertainties in the CUT dataset cause increased TMD compared to the standard setting. Meanwhile, the negligibly small UHD values indicate that GenPC effectively preserves the input structure.

G. Experimental Results on ShapeNet

We conducted quantitative comparisons on unseen categories from the ShapeNet [1] dataset, with the results presented in Table 3. Qualitative comparisons are provided in Figures 3 and 4. Our method surpasses previous approaches in both quantitative performance and visual quality on unseen categories, maintaining consistent shapes and capturing fine details without introducing noise.

H. More Visual Results on the Scannet Dataset

We provide additional visual results on the ScanNet [2] dataset in Figure 5. Our method is capable of preserving

Figure 2. I_{gen} generated with different prompts. (Left) Under severe occlusion, prompts describing category names can enhance image generation quality compared to null prompts. (Right) Varying types of results can be generated by using diverse prompts.

Figure 3. Qualitative comparison of guitar category in ShapeNet dataset.

Figure 4. Qualitative comparison of pistol category in ShapeNet dataset.

Figure 5. Visualization results on the ScanNet dataset.

geometric fidelity while accurately reconstructing missing structures and maintaining fine details.

References

[1] Angel X. Chang, Thomas A. Funkhouser, Leonidas J. Guibas, Pat Hanrahan, Qi-Xing Huang, Zimo Li, Silvio Savarese, Manolis Savva, Shuran Song, Hao Su, Jianxiong Xiao, Li Yi, and Fisher Yu. Shapenet: An information-rich 3d model repository. *CoRR*, abs/1512.03012, 2015. 2

[2] Angela Dai, Angel X Chang, Manolis Savva, Maciej Halber, Thomas Funkhouser, and Matthias Nießner. Scannet: Richly-annotated 3d reconstructions of indoor scenes. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 5828–5839, 2017. 2

[3] Prafulla Dhariwal and Alexander Quinn Nichol. Diffusion models beat gans on image synthesis. In *Advances in Neural Information Processing Systems 34: Annual Conference on Neural Information Processing Systems 2021, NeurIPS 2021, December 6-14, 2021, virtual*, pages 8780–8794, 2021. 1

[4] Yoni Kasten, Ohad Rahamim, and Gal Chechik. Point cloud completion with pretrained text-to-image diffusion models. In *Advances in Neural Information Processing Systems 36: Annual Conference on Neural Information Processing Systems 2023, NeurIPS 2023, New Orleans, LA, USA, December 10 - 16, 2023*, 2023. 1

[5] Jiale Xu, Weihao Cheng, Yiming Gao, Xintao Wang, Shenghua Gao, and Ying Shan. Instantmesh: Efficient 3d mesh generation from a single image with sparse-view large reconstruction models. *CoRR*, abs/2404.07191, 2024. 1

[6] Xingguang Yan, Liqiang Lin, Niloy J. Mitra, Dani Lischinski, Daniel Cohen-Or, and Hui Huang. Shapeformer: Transformer-based shape completion via sparse representation. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition, CVPR 2022, New Orleans, LA, USA, June 18-24, 2022*, pages 6229–6239. IEEE, 2022. 2

[7] Lvmin Zhang, Anyi Rao, and Maneesh Agrawala. Adding conditional control to text-to-image diffusion models. In *IEEE/CVF International Conference on Computer Vision, ICCV 2023, Paris, France, October 1-6, 2023*, pages 3813–3824. IEEE, 2023. 1