PCDreamer: Point Cloud Completion Through Multi-view Diffusion Priors

Supplementary Material

In this supplementary, we present experiments on additional datasets (*i.e.*, ScanObjectNN [6], KITTI [1], ShapeNet-55 [8]) to demonstrate the generalization ability of our method on real-world scans, as well as randomly cropped general partial points clouds. Besides, we show complementary visual results on PCN [9] and ShapeNet-55 [8] datasets and present additional visual examples when comparing our approach against SDS-Complete [2] and other single-view 3D generation methods.

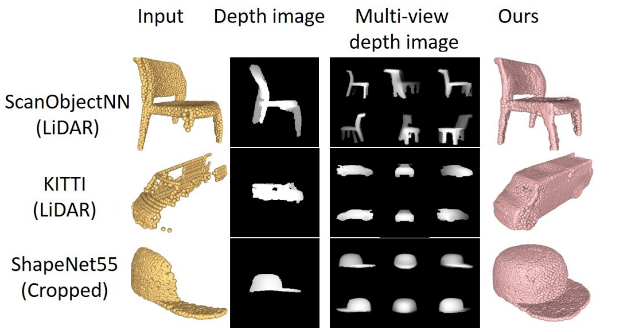

Figure A1. Visual results on real scanned and randomly cropped datasets. The first and second rows are real scanned data from ScanObjectNN and KITTI, while the third row is randomly cropped models from ShapeNet55. The converted single-view depth image and generated multi-view depth images are shown in columns two and three.

Real-world scans and general partial point clouds. As stated in Sec. 1 in the paper, completing a single-view partial point cloud is actively studied and much more challenging than completing a general one, since it usually misses more than half of the points (see the chair and lamp in Fig. 1 in the main paper). Due to the unique setting, we thus design our *PCDreamer*, dedicated to completing a single-view partial point cloud faithfully. However, thanks to the robustness of the large diffusion models and our fuser and consolidator, *PCDreamer* can work on general partial clouds well. Following the convention, we did not emphasize and report this advantage in the paper, but preliminary visual results are presented in Fig. A1, where *PCDreamer* successfully completes both LiDAR point clouds and randomly cropped ShapeNet55 point clouds. Indeed, a perfect single-view depth image is not feasible in this case, we thus select the most informative view and obtain the 'incomplete' depth image serving as the input of our method. To mitigate the incompleteness and potential quality degradation in the initial depth map, we employ a larger guidance scale (*e.g.*, 9.0) along with more complex text prompts in ControlNet.

Figure A2. The visual comparison with SDS-Complete.

This strategy effectively enhances the quality of the generated RGB images. Extending our method to handle arbitrary multi-view depth images is a promising direction, we leave it for future work.

Comparison with SDS-Complete [2]. We have conducted a visual comparison with SDS-Complete, which exploits diffusion priors with SDS optimization. The results in Fig. A2 demonstrate that while SDS-Complete is capable of recovering the overall shape, it fails to capture fine details, such as the leg and the armrest of the chair.

Additional Comparison with 3D generation methods. In Fig. A3, we present additional comparisons with other 3D model generation methods [5, 7]. Moreover, we provide further examples from the PCN dataset for comprehensive evaluation. As illustrated in the figure, the lack of extra views, the randomness in the generation process, and the inherent inconsistencies in multi-view images often prevent single-view 3D model generation methods from effectively capturing geometric details. Consequently, these models may produce shapes that, while plausible, exhibit significant deviations from the ground truth, such as variations in the width of the sofa. Therefore, multi-view images are more appropriate as auxiliary cues for point cloud completion, although directly utilizing them for point cloud completion typically yields suboptimal results.

Additional results on PCN and ShapeNet-55 dataset. Figs. A4 and A5, each presents an extra six results from the PCN and the ShapeNet-55 datasets, respectively. For both datasets, the first three cases in the first three rows possess high-quality and consistent multi-view depth images, whereas depth images with lower quality and less consistency are generated for the last three examples. Specifically,

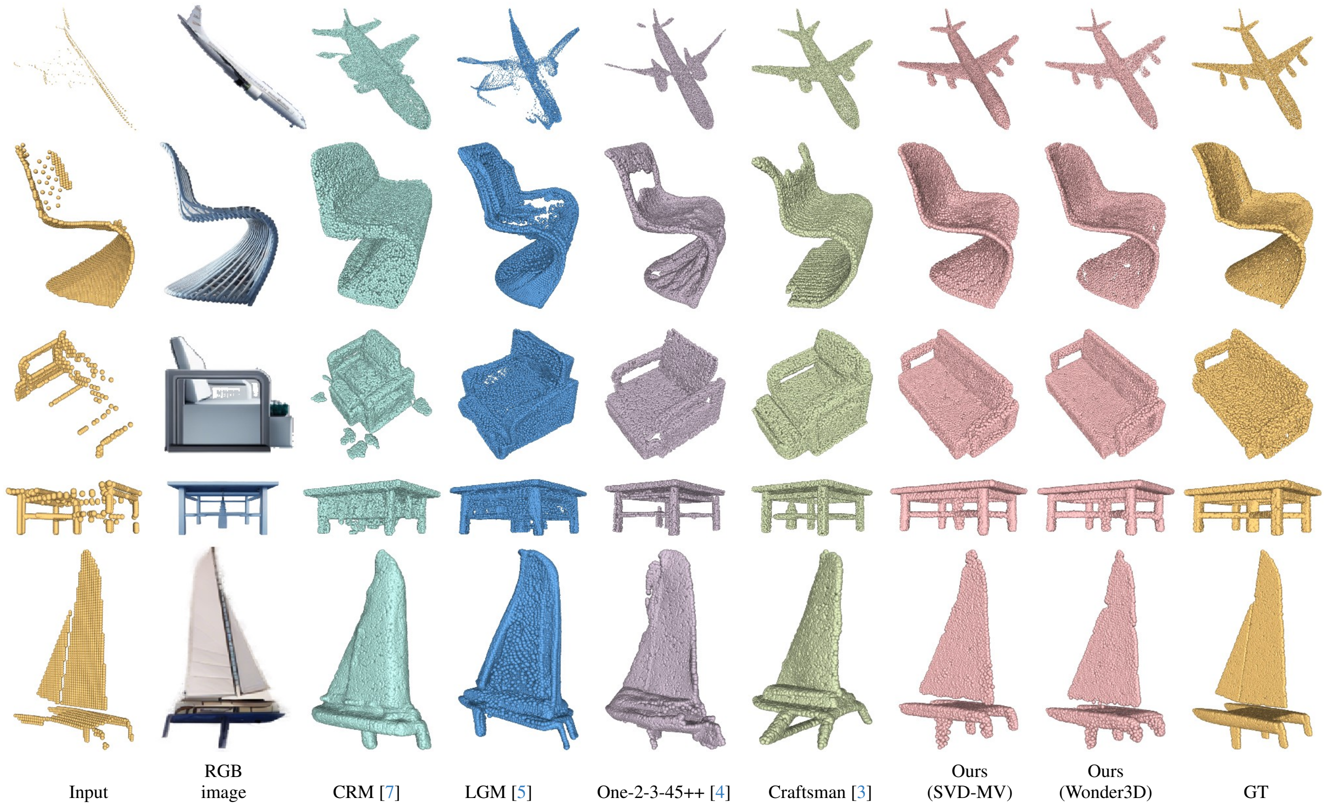

Figure A3. Additional visual comparisons with 3D model generation baesd on single-view RGB image.

the high randomness of the generation process, combined with limitations of the initial view, can result in multi-view depth images exhibiting inconsistencies across views (*e.g.*, the lid of the *Dustbin*), missing geometric details(*e.g.*, the *Helmet*), and the presence of noisy regions (*e.g.*, the leg of the *Table*). Furthermore, shapes imagined by the generative model may exhibit scale discrepancies relative to their real-world counterparts (*e.g.*, the *Sofa* and the *Table*). In addition, the depth estimation process may also occasionally produce suboptimal outcomes (*e.g.*, the *Dishwasher*). However, the proposed confidence-based shape consolidator effectively addresses these issues by eliminating unreliable points caused by inconsistencies in diffusion priors. As a result, our models produce accurate and reasonable outputs, as demonstrated in the second and fourth columns of the last three examples.

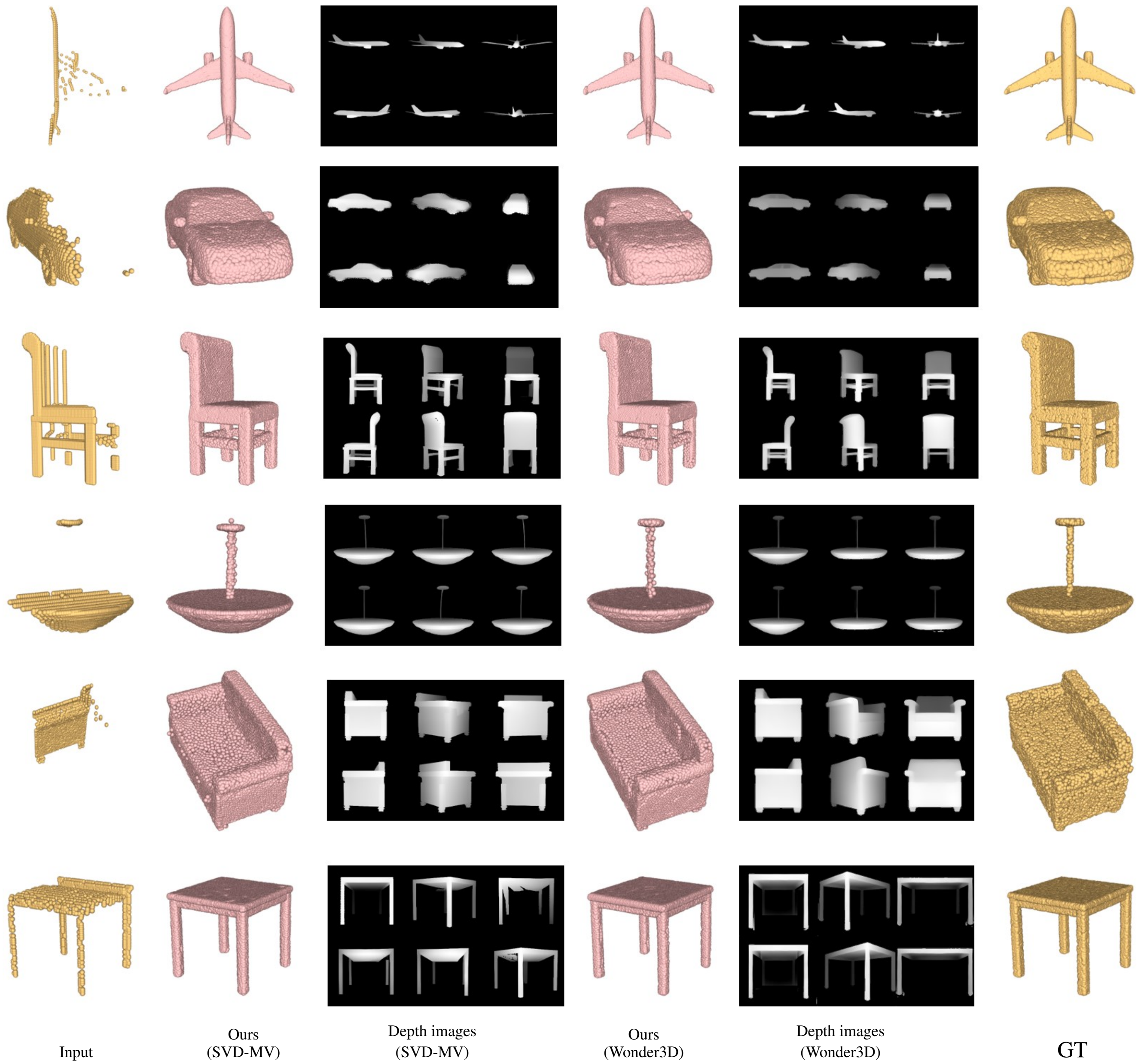

Figure A4. **Visual results as well as multi-view depth images on the PCN dataset.** The generated multi-view depth images corresponding to the last three rows exhibit lower quality and less consistency than the first three rows.

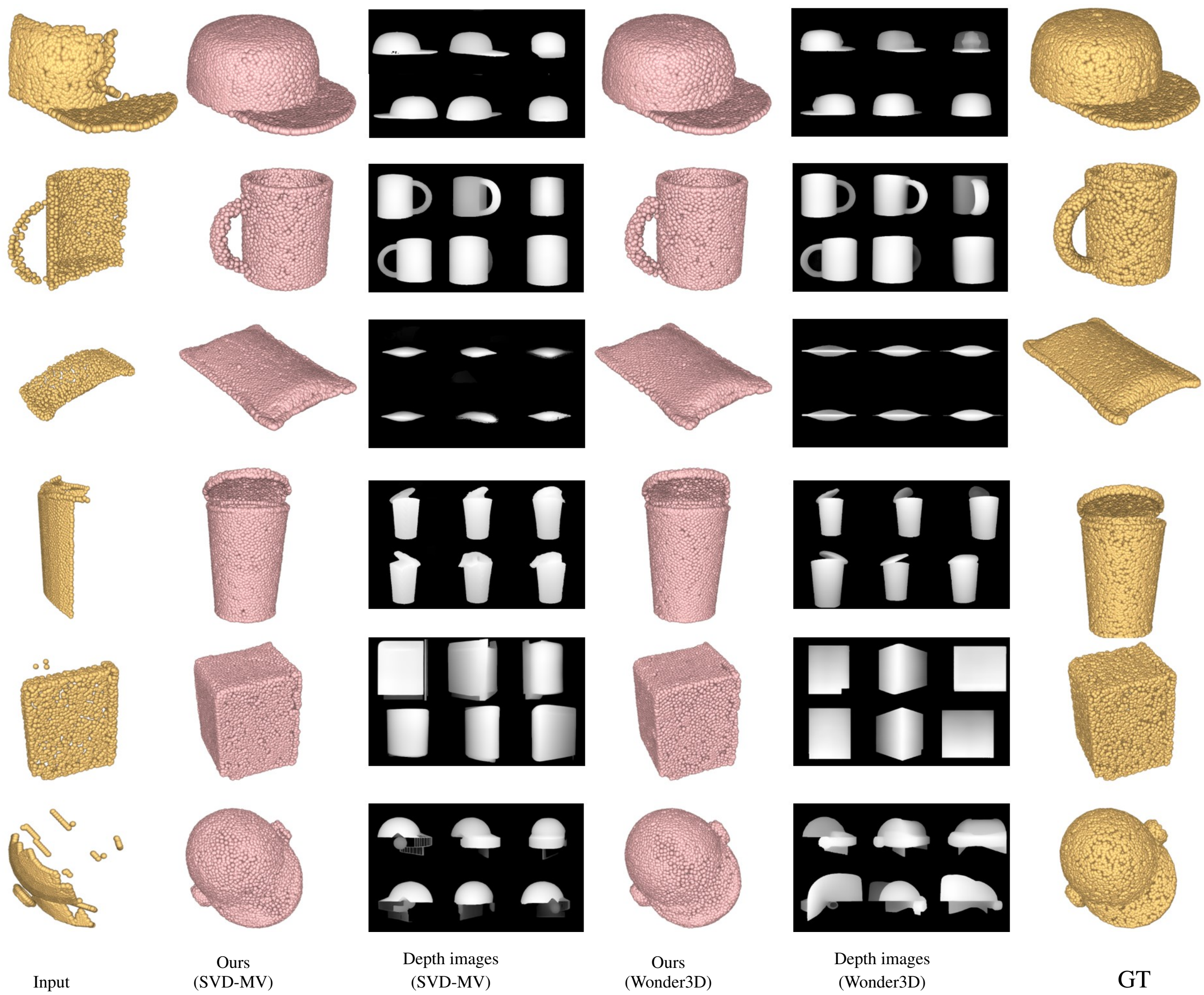

Figure A5. **Visual results as well as multi-view depth images on the ShapeNet-55 dataset.** The generated multi-view depth images corresponding to the last three rows exhibit lower quality and less consistency than the first three rows.

References

[1] Andreas Geiger, Philip Lenz, Christoph Stiller, and Raquel Urtasun. Vision meets robotics: The kitti dataset. *The International Journal of Robotics Research*, 32(11):1231–1237, 2013. 1

[2] Yoni Kasten, Ohad Rahamim, and Gal Chechik. Point-cloud completion with pretrained text-to-image diffusion models. *arXiv preprint arXiv:2306.10533*, 2023. 1

[3] Weiyu Li, Jiarui Liu, Rui Chen, Yixun Liang, Xuelin Chen, Ping Tan, and Xiaoxiao Long. Craftsman: High-fidelity mesh generation with 3d native generation and interactive geometry refiner. *arXiv preprint arXiv:2405.14979*, 2024. 2

[4] Minghua Liu, Ruoxi Shi, Linghao Chen, Zhuoyang Zhang, Chao Xu, Xinyue Wei, Hansheng Chen, Chong Zeng, Jiayuan Gu, and Hao Su. One-2-3-45++: Fast single image to 3d objects with consistent multi-view generation and 3d diffusion. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 10072–10083, 2024. 2

[5] Jiaxiang Tang, Zhaoxi Chen, Xiaokang Chen, Tengfei Wang, Gang Zeng, and Ziwei Liu. Lgm: Large multi-view gaussian model for high-resolution 3d content creation. In *European Conference on Computer Vision*, pages 1–18. Springer, 2025. 1, 2

[6] Mikaela Angelina Uy, Quang-Hieu Pham, Binh-Son Hua, Duc Thanh Nguyen, and Sai-Kit Yeung. Revisiting point cloud classification: A new benchmark dataset and classification model on real-world data. In *International Conference on Computer Vision (ICCV)*, 2019. 1

[7] Zhengyi Wang, Yikai Wang, Yifei Chen, Chendong Xiang, Shuo Chen, Dajiang Yu, Chongxuan Li, Hang Su, and Jun Zhu. Crm: Single image to 3d textured mesh with convolutional reconstruction model. *arXiv preprint arXiv:2403.05034*, 2024. 1, 2

[8] Xumin Yu, Yongming Rao, Ziyi Wang, Zuyan Liu, Jiwen Lu, and Jie Zhou. Pointr: Diverse point cloud completion with geometry-aware transformers. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 12498–12507, 2021. 1

[9] Wentao Yuan, Tejas Khot, David Held, Christoph Mertz, and Martial Hebert. Pcn: Point completion network. In *2018 international conference on 3D vision (3DV)*, pages 728–737. IEEE, 2018. 1