Supplementary Material for
RoutedFusion: Learning Real-time Depth Map Fusion

Abstract

In this supplementary document, we provide further visualizations and qualitative results of our learned depth map fusion approach in comparison to multiple baselines. Furthermore, we discuss the choice of different hyperparameters.

1. Hyperparameters

Number of Samples S. First, we discuss our choice for the number of samples S. With figure 1 and 2, we show that sampling 9 values inside the local window centered around the surface leads to the best reconstruction performance. The number of samples S is closely related to the truncation distance in standard TSDF fusion [2]. Since the spacing between samples in the window is fixed to the scene's resolution, the size of the local window is dependent on S. Therefore, an increase in S leads to an increase of the local window size. By increasing S and the window size, we feed more information along the ray to the depth fusion network and we can account for larger noise levels. However, if we increase the number of samples beyond 9, the performance decreases again, which is experimentally shown in figure 1 and 2. Having empirically evaluated the influence of S on our depth map fusion pipeline, we decided to keep the number of samples $S = 9$ constant across all experiments.

Outlier Post-filtering. In order to reduce the amount of outliers in the scene, we have chosen to introduce outlier post-filtering according to the accumulated update weights during TSDF integration. Therefore, after every 100 frames integrated, we re-initialize all voxels, where the accumulated weights are smaller than 3.

2. Qualitative Results

In figures 3 and 4, we show more examples of reconstructions of ShapeNet [1] objects using DeepSDF [4], OccupancyNetworks [3], standard TSDF fusion [2] and our proposed method. We can clearly show that our method is superior to all other approaches shown in reconstructing these objects from noisy depth measurements.

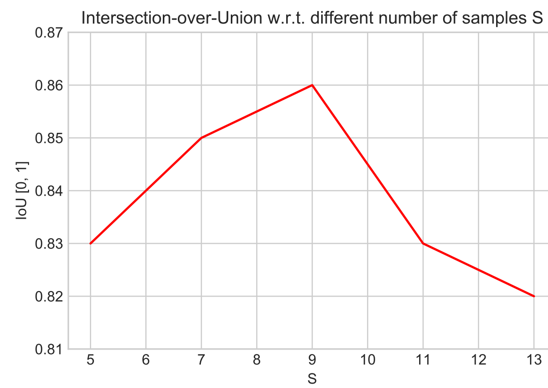


Figure 1. **Intersection over Union on Modelnet [5] test data for different numbers of samples** S. When sampling 9 SDF values inside the local window, our pipeline shows the best performance in reconstructing models from noisy depth measurements.

In figures 5 and 6, we qualitatively compare our method to standard TSDF fusion in reconstructing real-world scenes from the 3D scene dataset [6]. Our method significantly reduces the noise artifacts in the result, mitigates the surface thickening effect and generates very clean edges and corners.

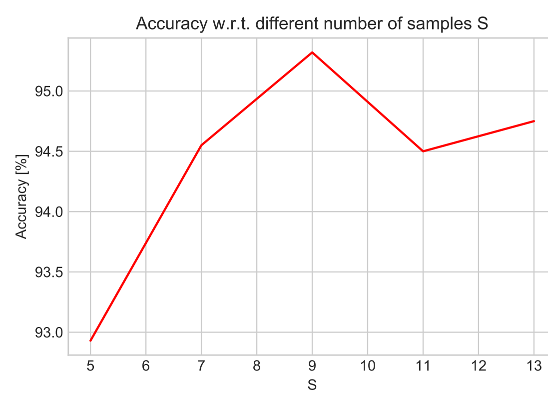


Figure 2. **Accuracy on Modelnet [5] test data for different numbers of samples** S. As it is the case for intersection-over-union, the accuracy peaks at 9 samples inside the window.

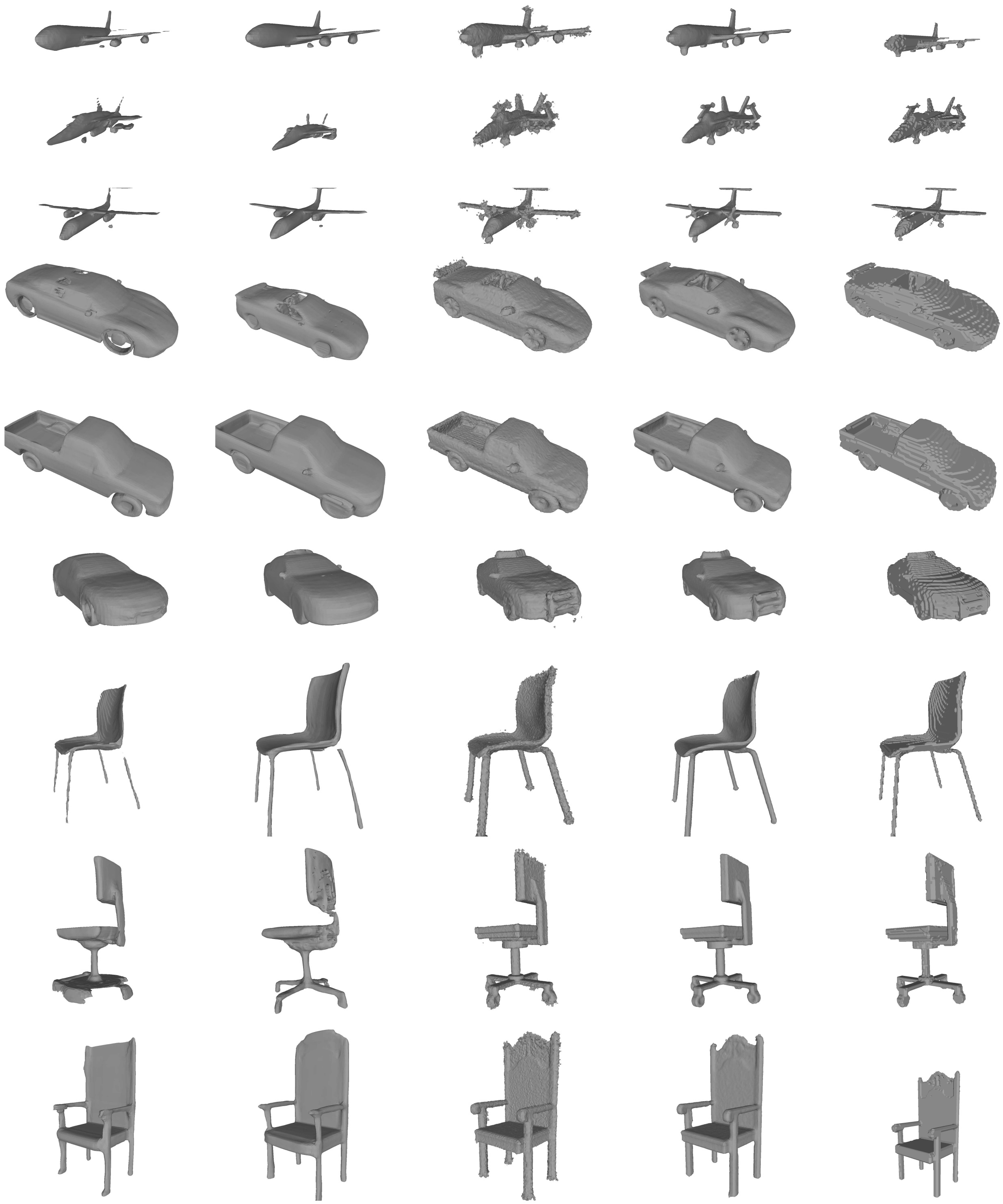
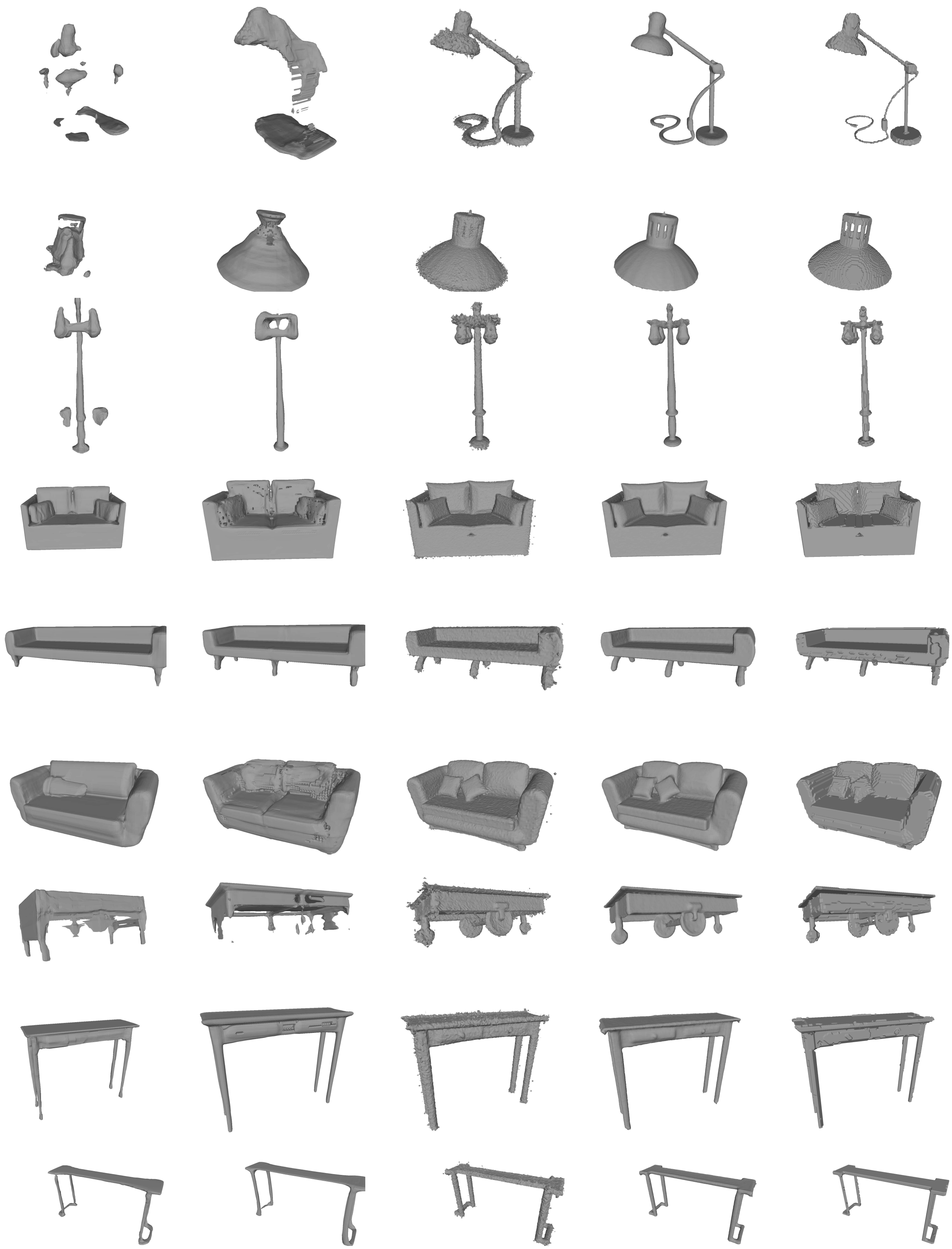
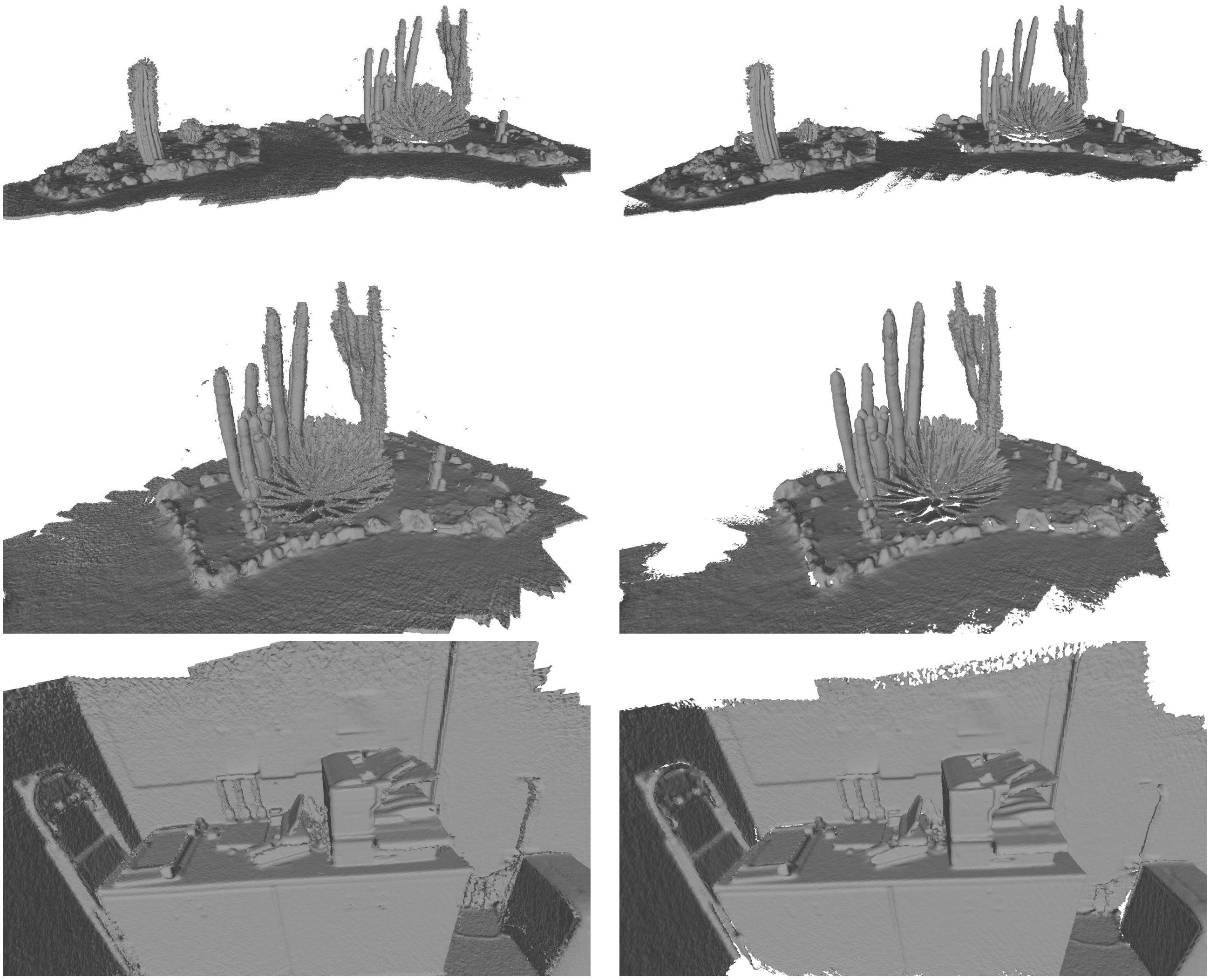


Figure 3. **More qualitative results on ShapeNet test data** They illustrate the significant performance difference in reconstructing fine geometries and clean edges between our proposed method and standard TSDF as well as recent learning-based approaches.



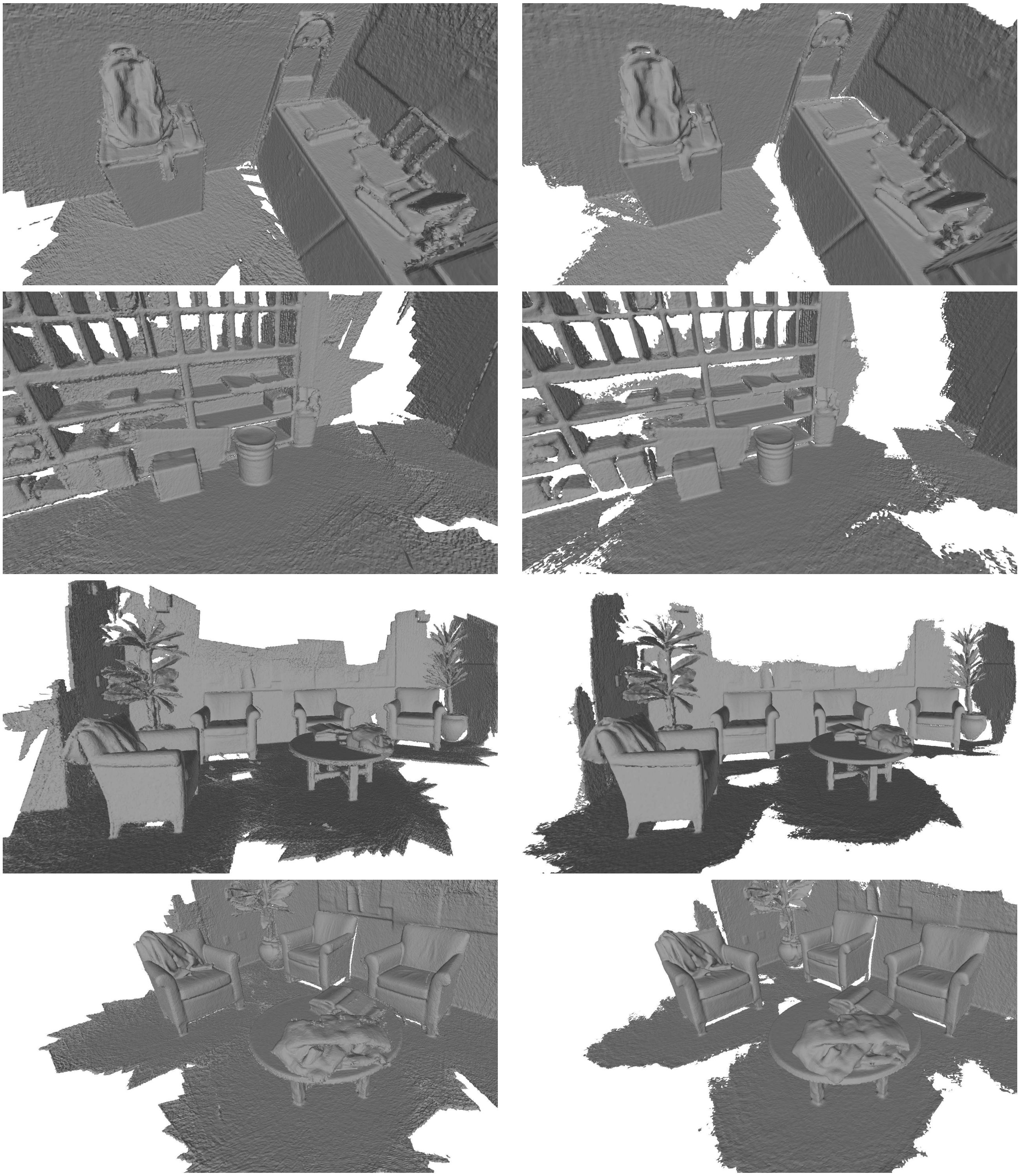
DeepSDF [4] OccupancyNetworks [3] TSDF [2] Ours GT

Figure 4. **More qualitative results on ShapeNet test data** They illustrate the significant performance difference in reconstructing fine geometries and clean edges between our proposed method and standard TSDF as well as recent learning-based approaches



TSDF [2] Ours

Figure 5. **More qualitative results of standard TSDF and our method on scene 3D data.** They illustrate the significant performance difference in reconstructing fine geometries and clean edges.



TSDF [2] Ours

Figure 6. **More qualitative results of standard TSDF and our method on scene 3D data.** They illustrate the significant performance difference in reconstructing fine geometries and clean edges.

References

[1] Angel X. Chang, Thomas Funkhouser, Leonidas Guibas, Pat Hanrahan, Qixing 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. Technical Report arXiv:1512.03012 [cs.GR], Stanford University — Princeton University — Toyota Technological Institute at Chicago, 2015. 1

[2] Brian Curless and Marc Levoy. A volumetric method for building complex models from range images. In *Proceedings of the 23rd Annual Conference on Computer Graphics and Interactive Techniques, SIGGRAPH 1996, New Orleans, LA, USA, August 4-9, 1996*, pages 303–312, 1996. 1, 2, 3, 4, 5

[3] Lars Mescheder, Michael Oechsle, Michael Niemeyer, Sebastian Nowozin, and Andreas Geiger. Occupancy networks: Learning 3d reconstruction in function space. In *Proc. International Conference on Computer Vision and Pattern Recognition (CVPR)*, 2019. 1, 2, 3

[4] Jeong Joon Park, Peter Florence, Julian Straub, Richard Newcombe, and Steven Lovegrove. Deepsdf: Learning continuous signed distance functions for shape representation. In *Proc. International Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2019. 1, 2, 3

[5] Zhirong Wu, S. Song, A. Khosla, Fisher Yu, Linguang Zhang, Xiaoou Tang, and J. Xiao. 3d shapenets: A deep representation for volumetric shapes. In *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1912–1920, June 2015. 1

[6] Qian-Yi Zhou and Vladlen Koltun. Dense scene reconstruction with points of interest. *ACM Trans. Graph.*, 32(4):112:1–112:8, 2013. 1