BNV-Fusion: Dense 3D Reconstruction using Bi-level Neural Volume Fusion

Kejie Li[1], Yansong Tang[1,2], Victor Adrian Prisacariu[1], Philip H.S. Torr[1]
[1]University of Oxford, [2]Tsinghua-Berkeley Shenzhen Institute, Tsinghua University

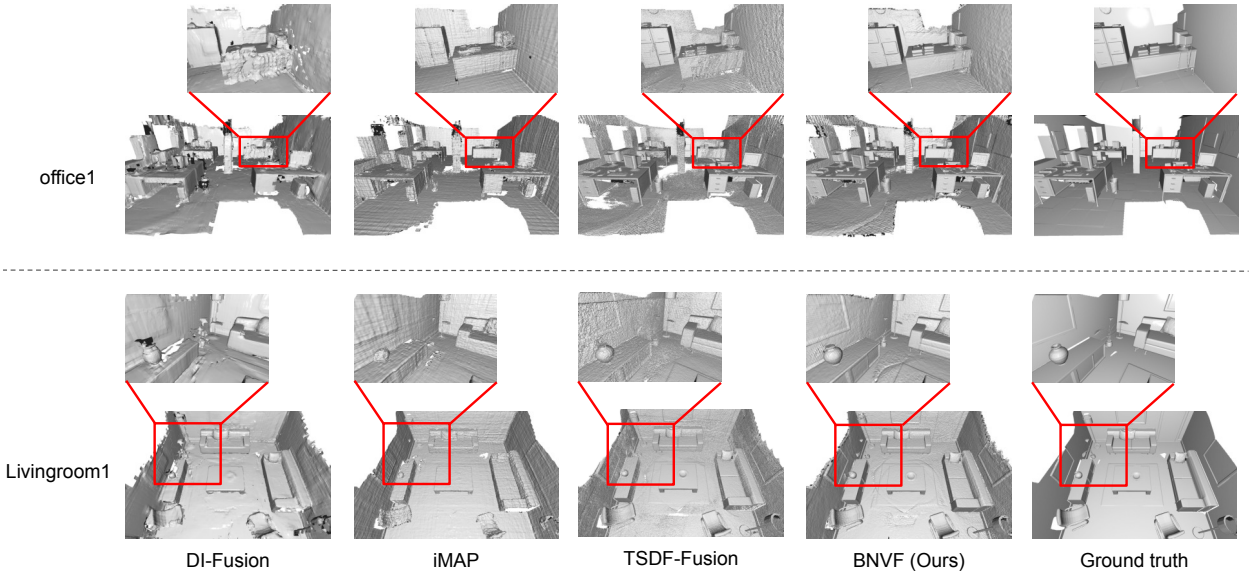

Figure 1. Reconstructions from different methods of two augmented ICL-NUIM sequences. Our method is less affected by noisy depth measurements (*e.g.* flat walls) and is able to reconstruct fine details (*e.g.* cables in the office1 and branches in livingroom1) better than other methods.

1. Qualitative Comparisons

Learning-based methods, iMAP [3] and DI-Fusion [2], fail to reconstruct faithfully as shown in Fig. 1. Although TSDF-Fusion [4] is able to capture more details in the scene than these learning-based methods, it is susceptible to noise in depth measurements. As a result, the walls in both scenes are reconstructed roughly. Our method consistently produces reconstructions that are closer to the ground-truth models than other methods.

Fig. 2 visualizes reconstructions of 5 ScanNet sequences. Note that reconstructions in ScanNet are more challenging than the other two datasets due to camera drift in the poses provided. Nevertheless, BNV-Fusion is less affected by the noise in poses and depth measurements than TSDF-Fusion.

2. Video Demo

Readers are encouraged to watch the video that demonstrates a reconstruction of the "lounge" scene in the 3D scene dataset [1]. We refresh the reconstruction at every 10 frames.

References

[1] Sungjoon Choi, Qian-Yi Zhou, and Vladlen Koltun. Robust reconstruction of indoor scenes. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2015. 1

[2] Jiahui Huang, Shi-Sheng Huang, Haoxuan Song, and Shi-Min Hu. Di-fusion: Online implicit 3d reconstruction with deep priors. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 8932–8941, 2021. 1

[3] Edgar Sucar, Shikun Liu, Joseph Ortiz, and Andrew J Davison. imap: Implicit mapping and positioning in real-time. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 6229–6238, 2021. 1

[4] Qian-Yi Zhou, Jaesik Park, and Vladlen Koltun. Open3d: A modern library for 3d data processing. *arXiv preprint arXiv:1801.09847*, 2018. 1

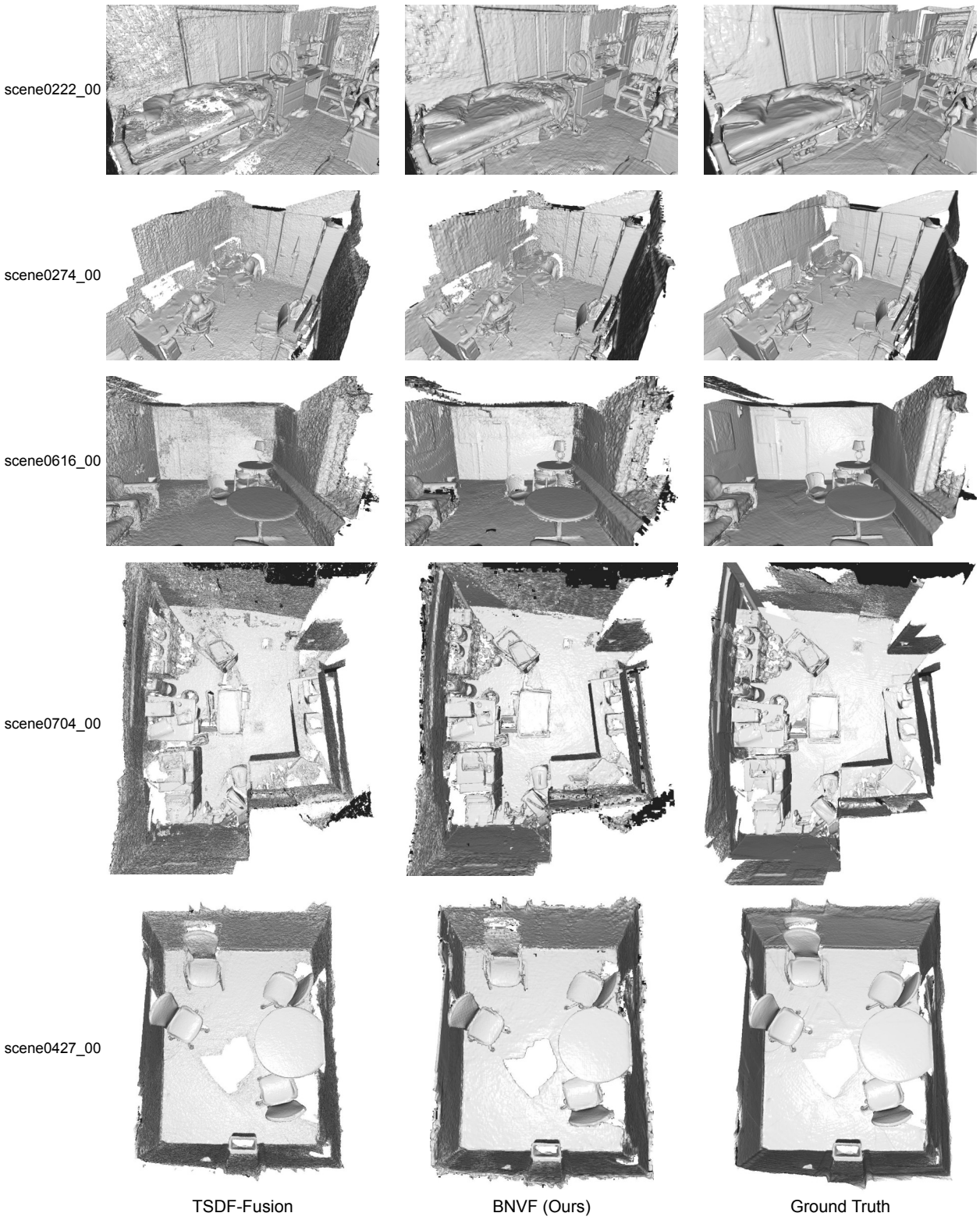

Figure 2. Reconstructions of 5 sequences in ScanNet. Note that our reconstructions are generally smoother and more complete.