Supplementary Material for Towards Balanced RGB-TSDF Fusion for Consistent Semantic Scene Completion by 3D RGB Feature Completion and a Classwise Entropy Loss Function

Laiyan Ding[1][0009−0006−3093−5335], Panwen Hu[1][0000−0001−6183−6598],
Jie Li[2][0000−0001−6254−1724], and Rui Huang[1✉][0000−0002−7950−1662]

[1] School of Science and Engineering, The Chinese University of Hong Kong (Shenzhen), Shenzhen, Guangdong, China
{laiyanding,panwenhu}@link.cuhk.edu.cn, ruihuang@cuhk.edu.cn
[2] School of Artificial Intelligence, Shenzhen Polytechnic University, Shenzhen, Guangdong, China
jieli1@szpt.edu.cn

1 Performance on NYU dataset

We compare our methods with other literatures that do not use extra data or iterative learning strategy on NYU dataset [4] here.

Table 1. Semantic scene completion results on NYU dataset. **Bold** numbers and $\underline{\text{underlined}}$ numbers represent the best and the second best scores among similar methods, respectively.

	Semantic Scene Completion						
Methods	wall	win.	chair	tvs	furn	objs	avg.
SSCNet [5]	24.4	0.0	12.6	7.8	27.1	10.1	24.7
DDRNet [2]	33.5	6.8	14.8	13.9	35.3	13.2	30.4
SketchNet [1]	40.5	$\underline{24.3}$	30.0	$\underline{14.3}$	$\underline{42.5}$	$\underline{28.6}$	41.1
PVANet [6]	**49.9**	15.9	**41.9**	12.9	**48.5**	**29.1**	**46.0**
Ours	$\underline{41.9}$	**26.8**	$\underline{32.9}$	**21.8**	41.8	27.2	$\underline{42.3}$

The results are in Table 1. We achieve the second highest mIoU among similar methods. Nevertheless, our method can boost the performance less compared with NYUCAD dataset. We presume that the effectiveness of both our FCM and classwise entropy loss depends on the preliminary results. Considering the last row in Figure 6 in the main paper, when most of the oven is classified wrongly, our method will still try to produce consistent results, i.e., more will be classified as *furniture* instead of *objects.* These cases occur more on NYU dataset. The recent work, PVANet [6], obtains better performance on NYU dataset, where depth missing and misalignment are severe. The point stream, that constitues of the main part of their network, can benefit from its point cloud input representation and network design. PointNet-like network is robust, i.e., it can produce stable results under tolerable shape corruptions [3]. Yet, our focus is RGB-TSDF fusion, and we outperform SketchNet [1] which also takes RGB-TSDF pairs as inputs, by 1.2% on SSC mIoU.

References

1. Chen, X., Lin, K.Y., Qian, C., Zeng, G., Li, H.: 3d sketch-aware semantic scene completion via semi-supervised structure prior. In: CVPR. pp. 4193–4202 (2020)
2. Li, J., Liu, Y., Gong, D., Shi, Q., Yuan, X., Zhao, C., Reid, I.: Rgbd based dimensional decomposition residual network for 3d semantic scene completion. In: CVPR. pp. 7693–7702 (2019)
3. Qi, C.R., Su, H., Mo, K., Guibas, L.J.: Pointnet: Deep learning on point sets for 3d classification and segmentation. In: CVPR. pp. 652–660 (2017)
4. Silberman, N., Hoiem, D., Kohli, P., Fergus, R.: Indoor segmentation and support inference from rgbd images. In: ECCV. pp. 746–760. Springer (2012)
5. Song, S., Yu, F., Zeng, A., Chang, A.X., Savva, M., Funkhouser, T.: Semantic scene completion from a single depth image. In: CVPR. pp. 1746–1754 (2017)
6. Tang, J., Chen, X., Wang, J., Zeng, G.: Not all voxels are equal: Semantic scene completion from the point-voxel perspective. In: AAAI. vol. 36, pp. 2352–2360 (2022)