OccuSeg: Occupancy-aware 3D Instance Segmentation (Supplementary)

Lei Han[1,2], Tian Zheng[1], Lan Xu[1,2], and Lu Fang[1]

[1]Tsinghua University [2]Hong Kong University of Science and Technology

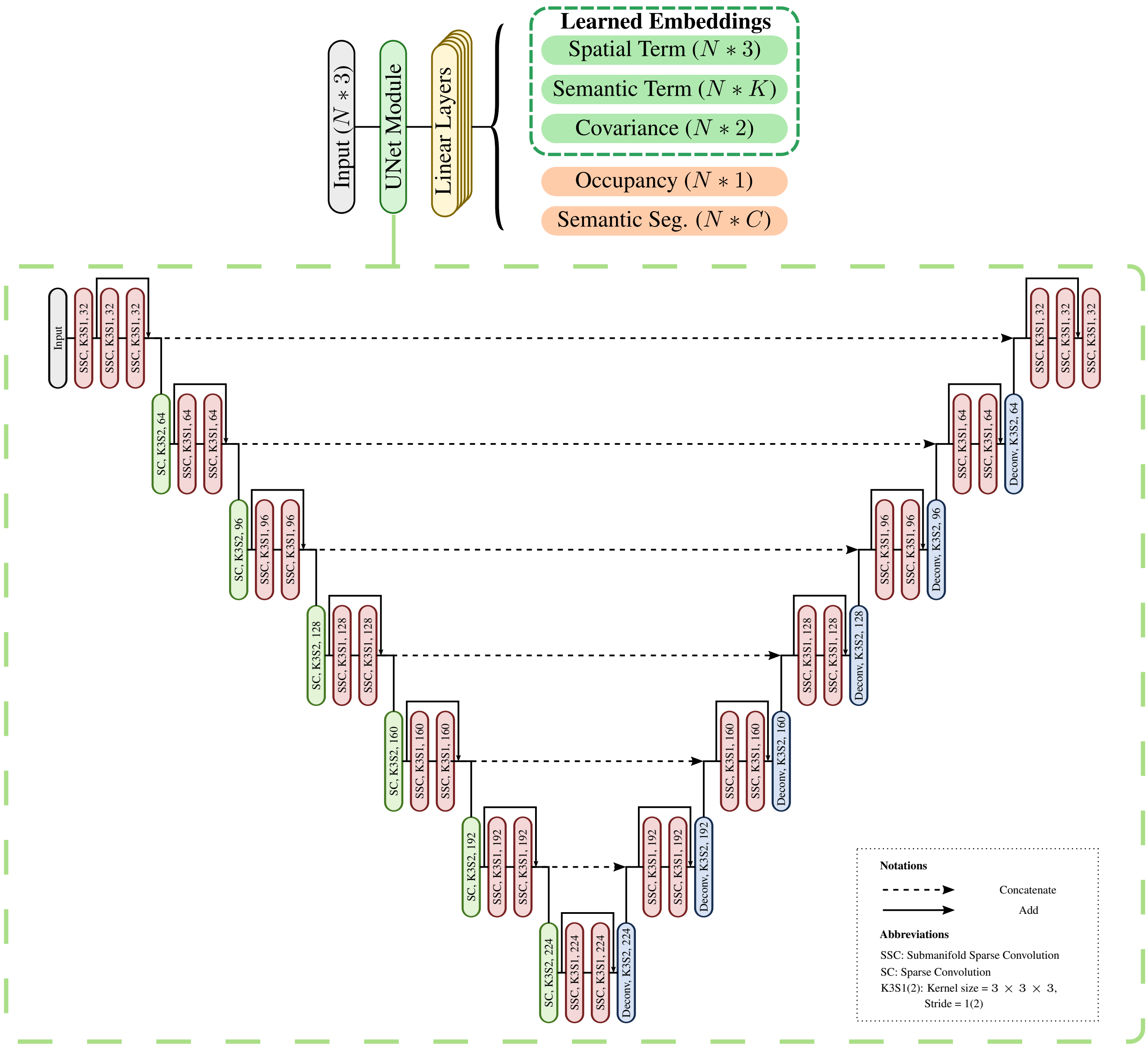

Figure 1. The detailed network architecture of OccuSeg. We use a 3D UNet-like network [5] as the feature extractor. Afterwards, the feature goes through different linear layers, yielding multiple learning objectives. (Best viewed on the display)

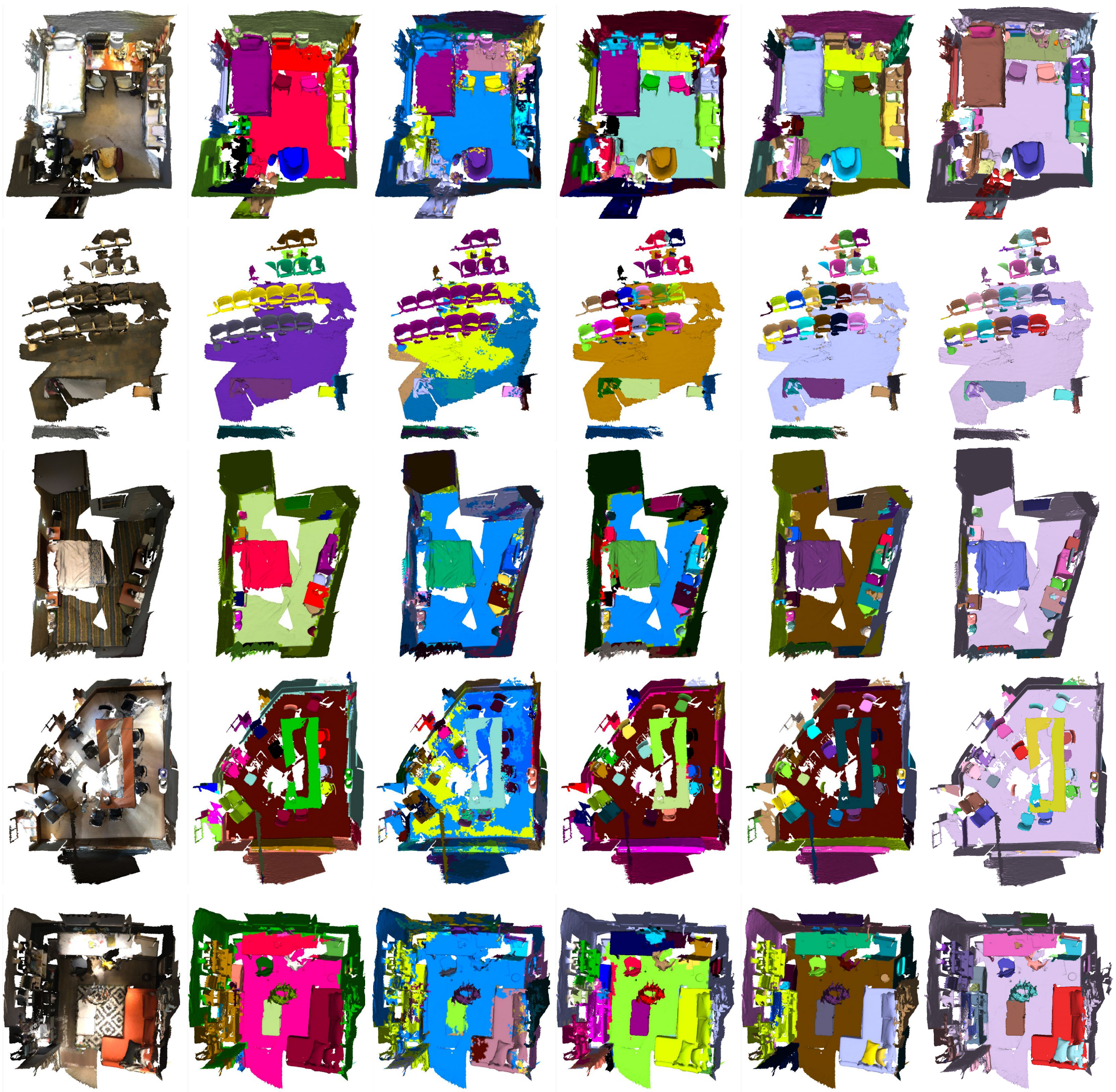

Figure 2. More qualitative comparisons on ScanNetV2 [1]. All the results of the previous methods are taken from Lahoud *et al*. [4]. Note that the instance segmentation result of sparse convolution networks(SCN) [3] is obtained by showing connected components of its semantic segmentation results.

1. Network Architecture Details

The network architecture adopted in OccuSeg is presented in this section. We employ the widely used UNet-style network [5] for feature learning as shown in Fig. 1. The network is mainly built upon submanifold sparse convolution(SSC) layers and sparse convolution layers, both of which are originally introduced by Graham [2]. In Fig. 1, K represents the dimension of the semantic term of the learned embedding, which is set to 32 in our experiment. Additionally, C stands for the semantic class numbers, which conforms to the training dataset.

2. Additional Qualitative Comparisons

In this section, we show some more qualitative comparisons on the ScanNetV2[1]. As shown in Fig. 2, our results are compared with sparse convolutional networks[3], SGPN [6] and multi-task metric learning method from Lahoud *et al*. [4]. Our results generally show a better capacity of dealing with small objects, as well as produce less noise.

References

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

[2] Benjamin Graham. Spatially-sparse convolutional neural networks. *arXiv preprint arXiv:1409.6070*, 2014. 2

[3] Benjamin Graham, Martin Engelcke, and Laurens van der Maaten. 3d semantic segmentation with submanifold sparse convolutional networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 9224–9232, 2018. 2, 3

[4] Jean Lahoud, Bernard Ghanem, Marc Pollefeys, and Martin R Oswald. 3d instance segmentation via multi-task metric learning. *arXiv preprint arXiv:1906.08650*, 2019. 2, 3

[5] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical image computing and computer-assisted intervention*, pages 234–241. Springer, 2015. 1, 2

[6] Weiyue Wang, Ronald Yu, Qiangui Huang, and Ulrich Neumann. Sgpn: Similarity group proposal network for 3d point cloud instance segmentation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 2569–2578, 2018. 2, 3