Density-Guided Semi-Supervised 3D Semantic Segmentation with Dual-Space Hardness Sampling

Supplementary Material

As introduced in the submitted paper, we propose a method named DDSemi for handling the semi-supervised 3D semantic segmentation task, which includes a density-guided contrastive learning technique and a dual-space hardness sampling strategy. In the supplementary material, we provide more details on the training configuration and more visualization results.

1. Training Configuration

The backbone network can be an arbitrary 3D semantic segmentation network [4, 5, 8, 11]. The classifier and projector are both multi-layer perceptions.

For the S3DIS [1] dataset, we adopt AdamW as the optimizer and use the Cosine learning rate scheduler. The learning rate and weight decay are set to 0.0006 and 0.05, respectively. For the SemanticKITTI [2] and nuScenes [3] datasets, we adopt AdamW as the optimizer and use the poly learning rate scheduler where the power is set to 0.9. The learning rate and weight decay are set to 0.006 and 0.01, respectively.

In addition, it is worth noting that when Kong *et al.* [7] compared their proposed LaserMix with GPC [6], they directly used the scores reported in [6]. However, Kong *et al.* and Jiang *et al.* [6] used different settings (*e.g.*, the data splits), thus, we reproduced the results of GPC with the same settings that Kong *et al.* used on the SemanticKITTI dataset and reported the reproduced results in our submitted paper for a fair comparison.

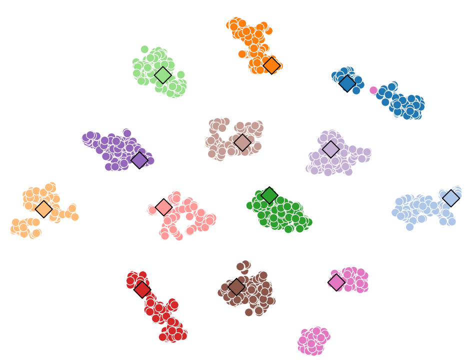

Figure 1. Visualization of the features sampled from the S3DIS [1] dataset and their corresponding anchors. The circles represent the features and the diamonds represent the anchors. Different colors represent different categories.

2. Limitation

The proposed DDSemi requires many k-nearest neighbors searching operations at the training stage, resulting in a relatively higher training cost. For example, compared with GPC [6] which spends 3 seconds on each training iteration, the proposed DDSemi has to spend 6 seconds on each training iteration. However, it is worth noting that the inference speed of DDSemi is the same as the comparative methods, because only the backbone network and classifier are used at the inference stage.

3. More Visualization Results

We visualize the features and their corresponding anchors in Figure 1. As seen from this figure, the features and their corresponding anchors aggregate in the same local regions of the feature space, and the anchors are located in the dense local regions of their corresponding clusters rather than the naive center of the clusters, which is consistent with the revealed finding about clustering in [10].

In addition, we give more visualization of the segmentation results on three public datasets in Figure 2, Figure 3, and Figure 4. As seen from these figures, the proposed DDSemi outperforms the comparative methods.

References

[1] Iro Armeni, Ozan Sener, Amir R Zamir, Helen Jiang, Ioannis Brilakis, Martin Fischer, and Silvio Savarese. 3d semantic parsing of large-scale indoor spaces. In *CVPR*, pages 1534–1543, 2016. 1, 2

[2] Jens Behley, Martin Garbade, Andres Milioto, Jan Quenzel, Sven Behnke, Cyrill Stachniss, and Jurgen Gall. Semantickitti: A dataset for semantic scene understanding of lidar sequences. In *ICCV*, pages 9297–9307, 2019. 1, 2

[3] Holger Caesar, Varun Bankiti, Alex H Lang, Sourabh Vora, Venice Erin Liong, Qiang Xu, Anush Krishnan, Yu Pan, Giancarlo Baldan, and Oscar Beijbom. nuscenes: A multimodal dataset for autonomous driving. In *CVPR*, pages 11621–11631, 2020. 1, 2

[4] Benjamin Graham, Martin Engelcke, and Laurens van der Maaten. 3d semantic segmentation with submanifold sparse convolutional networks. *CVPR*, pages 9224–9232, 2018. 1

[5] Qingyong Hu, Bo Yang, Linhai Xie, Stefano Rosa, Yulan Guo, Zhihua Wang, Niki Trigoni, and Andrew Markham. Randla-net: Efficient semantic segmentation of large-scale point clouds. In *CVPR*, pages 11108–11117, 2020. 1

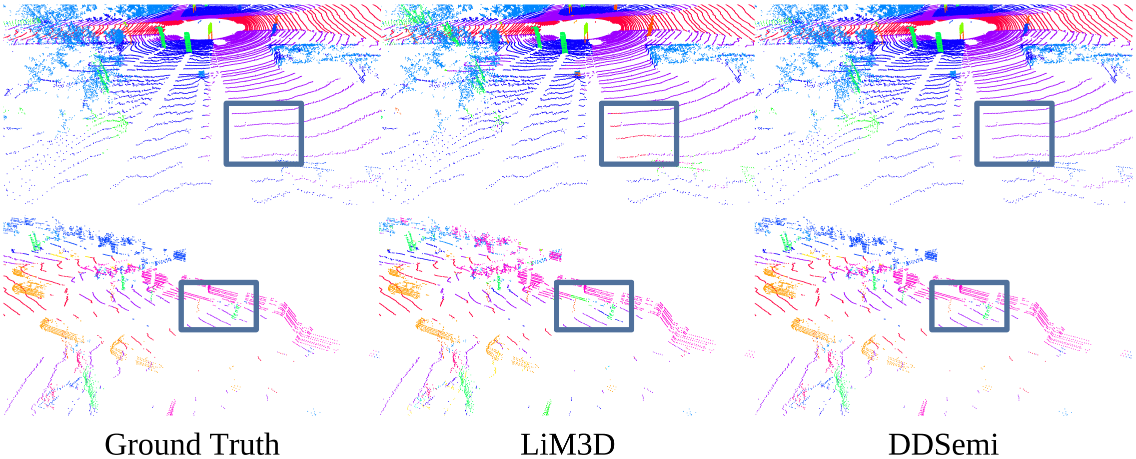

Figure 2. Visualization of the semantic segmentation results on the SemanticKITTI [2] dataset by LiM3D [9] and the proposed DDSemi.

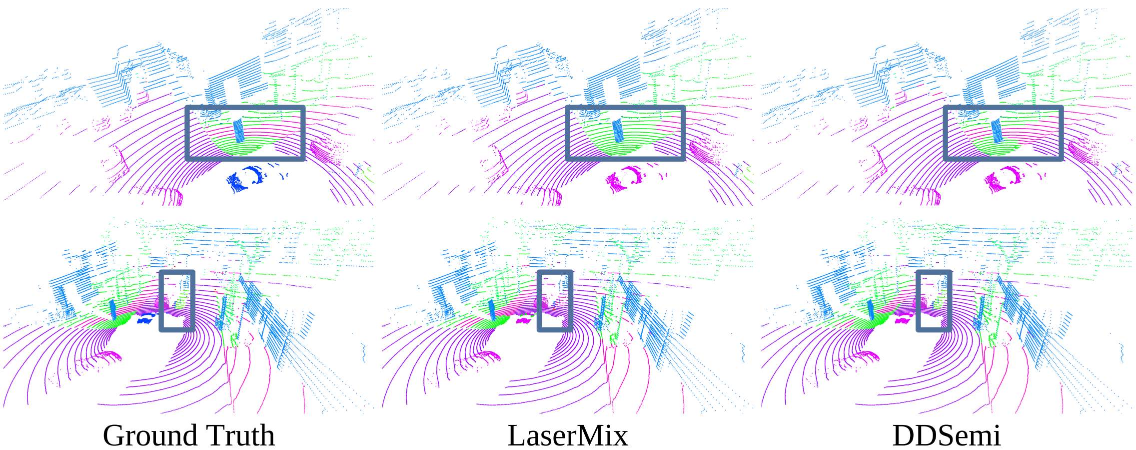

Figure 3. Visualization of the semantic segmentation results on the nuScenes [3] dataset by LaserMix [7] and the proposed DDSemi.

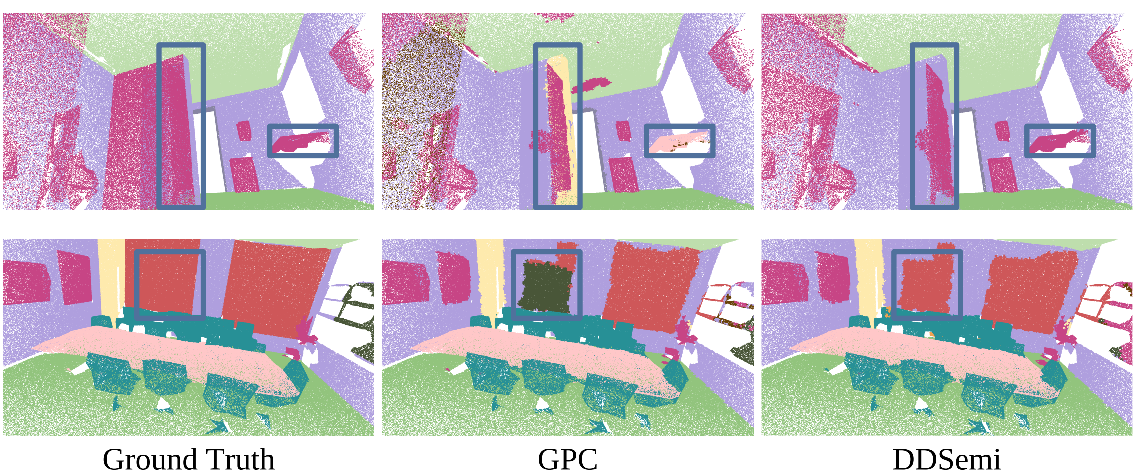

Figure 4. Visualization of the semantic segmentation results on the S3DIS [1] dataset by GPC [6] and the proposed DDSemi.

[6] Li Jiang, Shaoshuai Shi, Zhuotao Tian, Xin Lai, Shu Liu, Chi-Wing Fu, and Jiaya Jia. Guided point contrastive learning for semi-supervised point cloud semantic segmentation. In *ICCV*, pages 6403–6412, 2021. 1, 2

[7] Lingdong Kong, Jiawei Ren, Liang Pan, and Ziwei Liu. Lasermix for semi-supervised lidar semantic segmentation. In *CVPR*, pages 21705–21715, 2023. 1, 2

[8] Xin Lai, Yukang Chen, Fanbin Lu, Jianhui Liu, and Jiaya Jia. Spherical transformer for lidar-based 3d recognition. In *CVPR*, pages 17545–17555, 2023. 1

[9] Li Li, Hubert P. H. Shum, and Toby P. Breckon. Less is more: Reducing task and model complexity for 3d point cloud semantic segmentation. In *CVPR*, pages 9361–9371, 2023. 2

[10] Alex Rodriguez and Alessandro Laio. Clustering by fast search and find of density peaks. *Science*, 344:1492 – 1496, 2014. 1

[11] Xinge Zhu, Hui Zhou, Tai Wang, Fangzhou Hong, Yuexin Ma, Wei Li, Hongsheng Li, and Dahua Lin. Cylindrical and asymmetrical 3d convolution networks for lidar segmentation. In *CVPR*, pages 9939–9948, 2021. 1