WS3D Supplementary Material

Kangcheng Liu*[1], Yuzhi Zhao[2], Qiang Nie[3], Zhi Gao[4], and Ben M. Chen[1]

[1] The Chinese University of Hong Kong, China
[2] City University of Hong Kong, China
[3] Tencent Youtu Lab, China
[4] Wuhan University, China
kcliu@mae.cuhk.edu.hk, yzzhao2-c@my.cityu.edu.hk, qnie.cuhk@gmail.com,
bmchen@mae.cuhk.edu.hk, gaozhinus@gmail.com

Overview

In the supplementary material, additional details that are not included in the main paper due to space limits are provided as follows:

- Details of our experimental datasets (see Section 1).
- Details of our data augmentation approach (see Section 2).
- Training details of the boundary prediction network (see Section 3).
- Some other experimental setting details (see Section 4).

1 Experimental Datasets Details

To demonstrate the effectiveness of our proposed **WS3D**, we tested it on both indoor and outdoor 3D scene segmentation benchmarks, including S3DIS [1], SemanticKITTI [2], and ScanNet [3] for semantic segmentation, and ScanNet [3] for instance segmentation. Many prior works have used these datasets for benchmark comparisons [6–13]. Therefore, we continue to use these datasets in our work. The detailed information of each dataset is listed as follows:

ScanNet-V2 is a standard large-scale and widely-acknowledged 3D indoor scene understanding benchmark consisting of more than 1,600 3D scene scans. It contains densely reconstructed point clouds from RGB-D images captured by the depth camera. For the dataset partition in fully-supervised learning, we follow the official partition of ScanNet-V2 [3] using 1,201 scans as the training set, 312 scans as the validation set, and 100 scans as the test set.

SemanticKITTI [2] is a large-scale point cloud understanding benchmark for self-driving applications. The point cloud of individual scene is obtained densely by Velodyne VLP-64 LiDAR. The dataset covers long-range road scenes, thanks to the vehicle driving in the complex road scenarios with mounted LiDAR to capture the point clouds. More than 43,550 labeled LiDAR scans in total are split into 21 sequences. Each LiDAR scan contains approximately 10^5 points. We follow the official setting [2] that uses the sequences 00 to 07, and 09 to 10

for training, the sequence 08 for validation, and the sequence 11 to 21 for testing.

S3DIS is a commonly used indoor 3D scene understanding benchmark that contains various indoor scenes. It includes six indoor areas which are made up of 271 rooms. Each room contains the magnitude of 10^6 points. The typical room size is $20 \times 15 \times 5$ meters. In our experiments, we follow **GPC** [5] using area 5 as the validation set and the other areas as the training set.

2 Data Augmentation Details

2.1 Data Augmentation Approaches Details

We perform three kinds of data augmentations to enhance the unlabeled data and to perform region-level contrastive learning, including: 1. **Geometric rotations and mirroring** of the point cloud scene. 2. **Adding minor disturbance** to the point cloud scene. 3. **Random scaling** of the point cloud scene.

Geometric rotations and mirroring. The original point cloud scene $P \in \mathbb{R}^{N_p \times (3+f_a)}$ can be partitioned into the coordinate channels of point clouds 3D positions $P_x \in \mathbb{R}^{N_p \times 3}$ and the feature channels $P_f \in \mathbb{R}^{N_p \times f_a}$ with diverse attributes such as the color and normal. We perform rotation of the point clouds around the X axis. Denote the rotated point clouds as P_x^r, and the rotation matrix as R_x, we can obtain the transformed point clouds as: $P_x^r = P_x R_x$, where $R_x = \begin{bmatrix} 1 & 0 & 0 \\ 0 & cos(\phi) & -sin(\phi) \\ 0 & sin(\phi) & cos(\phi) \end{bmatrix}$. R_x is the rotation matrix of ϕ around x axis. The rotated degrees for all input point clouds scenes samples obey the uniform distribution of $U[0, 2\pi]$. We perform point clouds mirroring with respect to the Z axis. Denote the mirroring transformation as $R_m = diag(1, 1, -1)$, the final augmented point clouds scene P_x^m is denoted as $P_x^m = P_x^r R_m$.

Adding minor disturbance. We add the noise for the normalized point clouds coordinates, which can be represented as $P_x^d = P_x^m + T_d$. T_d is the Gaussian noise with the mean of 0 and variance of 0.01. We set T_d in the range of $[-0.1, 0.1]$.

Random scaling. We do random scaling to the point cloud scene. Random scaling of $(0.9 - 1.1)$ is done to the point clouds coordinates. Denote the scaling parameter as η, the final augmented point clouds scene $P_x^{'}$ is given as $P_x^{'} = \eta P_x^d$.

3 Training Details of the Boundary Prediction Network

As mentioned in our main paper, we have utilized the JSE-Net [4] as our boundary prediction network to obtain the boundary labels. In this Section, we provide the training details of the boundary prediction network, which follows the settings in the JSE-Net [4]. The input point clouds are down-sampled with the grid size of $4cm$. The gradient descent with a momentum of 0.98 is used for optimization and the initial learning rate is 0.01. We reduce the learning rate

exponentially and the learning rate is multiplied by 0.1 every 100 epochs. The network is trained on a single GTX 1080Ti GPU. we follow JSENet in other settings except substituting the binary cross-entropy loss L_{bce} with the focal loss L_{foc}, as mentioned in our main paper.

4 Experimental Setting Details

The supervised branch of our **WS3D** is the same as the **GPC** [5]. Therefore, the 'Sup-only' and 'Sup-only-GPC' both represent the supervised-only model in GPC that is merely trained with labeled data. To the best of our knowledge, the weakly-supervised/semi-supervised 3D scene semantic segmentation problem under the limited reconstruction setting has only been explored in **GPC** [5]. Therefore, we use **GPC** as a counterpart for fair comparisons in our experiments.

References

1. Armeni, I., Sener, O., Zamir, A.R., Jiang, H., Brilakis, I., Fischer, M., Savarese, S.: 3d semantic parsing of large-scale indoor spaces. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). pp. 1534–1543 (2016) 1
2. Behley, J., Garbade, M., Milioto, A., Quenzel, J., Behnke, S., Stachniss, C., Gall, J.: Semantickitti: A dataset for semantic scene understanding of lidar sequences. In: Proceedings of the IEEE International Conference on Computer Vision (ICCV). pp. 9297–9307 (2019) 1
3. Dai, A., Chang, A.X., Savva, M., Halber, M., Funkhouser, T., Nießner, M.: Scannet: Richly-annotated 3d reconstructions of indoor scenes. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). pp. 5828–5839 (2017) 1
4. Hu, Z., Zhen, M., Bai, X., Fu, H., Tai, C.l.: Jsenet: Joint semantic segmentation and edge detection network for 3d point clouds. In: European Conference on Computer Vision (ECCV). pp. 222–239. Springer Nature (2020) 2
5. Jiang, L., Shi, S., Tian, Z., Lai, X., Liu, S., Fu, C.W., Jia, J.: Guided point contrastive learning for semi-supervised point cloud semantic segmentation. In: Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV). pp. 6423–6432 (2021) 2, 3
6. Liu, K.: Robust industrial uav/ugv-based unsupervised domain adaptive crack recognitions with depth and edge awareness: From system and database constructions to real-site inspections. In: 2022 30th ACM International Conference on Multimedia (ACM MM) (2022) 1
7. Liu, K., Gao, Z., Lin, F., Chen, B.M.: Fg-net: Fast large-scale lidar point cloud-sunderstanding network leveraging correlatedfeature mining and geometric-aware modelling. arXiv preprint arXiv:2012.09439 (2020) 1
8. Liu, K., Gao, Z., Lin, F., Chen, B.M.: Fg-conv: Large-scale lidar point clouds understanding leveraging feature correlation mining and geometric-aware modeling. In: 2021 IEEE International Conference on Robotics and Automation (ICRA). pp. 12896–12902. IEEE (2021) 1

9. Liu, K., Gao, Z., Lin, F., Chen, B.M.: Fg-net: A fast and accurate framework for large-scale lidar point cloud understanding. IEEE Transactions on Cybernetics (2022) 1
10. Liu, K., Han, X., Chen, B.M.: Deep learning based automatic crack detection and segmentation for unmanned aerial vehicle inspections. In: 2019 IEEE International Conference on Robotics and Biomimetics (ROBIO). pp. 381–387. IEEE (2019) 1
11. Liu, K., Qu, Y., Kim, H.M., Song, H.: Avoiding frequency second dip in power unreserved control during wind power rotational speed recovery. IEEE transactions on power systems **33**(3), 3097–3106 (2017) 1
12. Liu, K., Zhou, X., Chen, B.M.: An enhanced lidar inertial localization and mapping system for unmanned ground vehicles. In: 2022 17th IEEE International Conference on Control and Automation (ICCA). IEEE (2022) 1
13. Liu, K., Zhou, X., Zhao, B., Ou, H., Chen, B.M.: An integrated visual system for unmanned aerial vehicles following ground vehicles: Simulations and experiments. In: 2022 17th IEEE International Conference on Control and Automation (ICCA). IEEE (2022) 1