Supplementary Materials: PVP: Polar Representation Boost for 3D Semantic Occupancy Prediction

Yujing Xue*
NUS
xueyj14@outlook.com

Jiaxiang Liu*
Zhejiang University
jiaxiang.21@intl.zju.edu.cn

Jiawei Du†
A*STAR, NUS
dujiawei@u.nus.edu

Joey Tianyi Zhou
A*STAR
joey.tianyi.zhou@gmail.com

Table 1. The information of 3D occupancy prediction dataset: OpenOccupancy

	OpenOccupancy [2]
Source	nuScenes [1]
Class	17 class
Annotation	The Augmenting And Purifying pipeline [2]
Modal	LiDAR(10 LiDAR sweeps)/Image(1600×900)
Training	LiDAR(700)/Cameras(28,130)
Validation	LiDAR(150)/Cameras(6,019)

1. Dataset

The OpenOccupancy dataset originates from the nuScenes dataset [1], which has been extended to include dense semantic occupancy annotations through the AAP (Augmenting And Purifying) pipeline [2] to efficiently annotate and densify the occupancy labels. This extension features 17 semantic categories and includes data modalities such as LiDAR and cameras. For LiDAR-based methods, the input comprises 10 LiDAR sweeps. For camera-based methods, the input image size is set to 1600 × 900. For the LiDAR modality, the OpenOccupancy dataset includes 700 training point cloud sequences and 150 validation point cloud sequences. For the camera modality, it comprises 28,130 training frames and 6,019 validation frames, as shown in Table 1.

2. Evaluation Protocol

2.1. Geometric Metric

The evaluation range is set from $[-51.2m, 51.2m]$ for the X, Y axes, and $[-3m, 5m]$ for the Z axis. In line with previous studies, the voxel resolution is set at 0.2m, resulting in a volume of $40 \times 512 \times 512$ voxels for occupancy prediction. For evaluation metrics, we use the Intersection of Union (IoU) as the geometric metric, which identifies a voxel as either occupied or empty (treating all occupied voxels as a single category).

$$\mathrm{IoU} = \frac{\mathrm{TP}_o}{\mathrm{TP}_o + \mathrm{FP}_o + \mathrm{FN}_o}, \tag{2}$$

where TP_o, FP_o, FN_o represent the number of true positive, false positive, and false negative predictions for occupied voxels, respectively.

2.2. Semantic Metric

We calculate the mean IoU (mIoU) for each class to serve as the semantic metric:

$$\mathrm{mIoU} = \frac{1}{C_{\text{sem}}} \sum_{c=1}^{C_{\text{sem}}} \frac{\mathrm{TP}_c}{\mathrm{TP}_c + \mathrm{FP}_c + \mathrm{FN}_c}, \tag{3}$$

where TP_c, FP_c, FN_c denote the number of true positive, false positive, and false negative predictions for class c, and C_{sem} represents the total number of semantic classes. In accordance with previous research [3], the noise class [3] is excluded from the evaluation.

3. Performance by Category

From the main text, we can see that PVP has surpassed state-of-the-art (SOTA) methods. Figure 1 visually demonstrates the gains achieved by our method (represented by the red line) across different semantic categories relative to SOTA methods. The red line (ours) consistently remains above the others, outperforming SOTA methods in predictions across nearly all categories. This confirms the high predictive accuracy and robustness of the PVP method for each category.

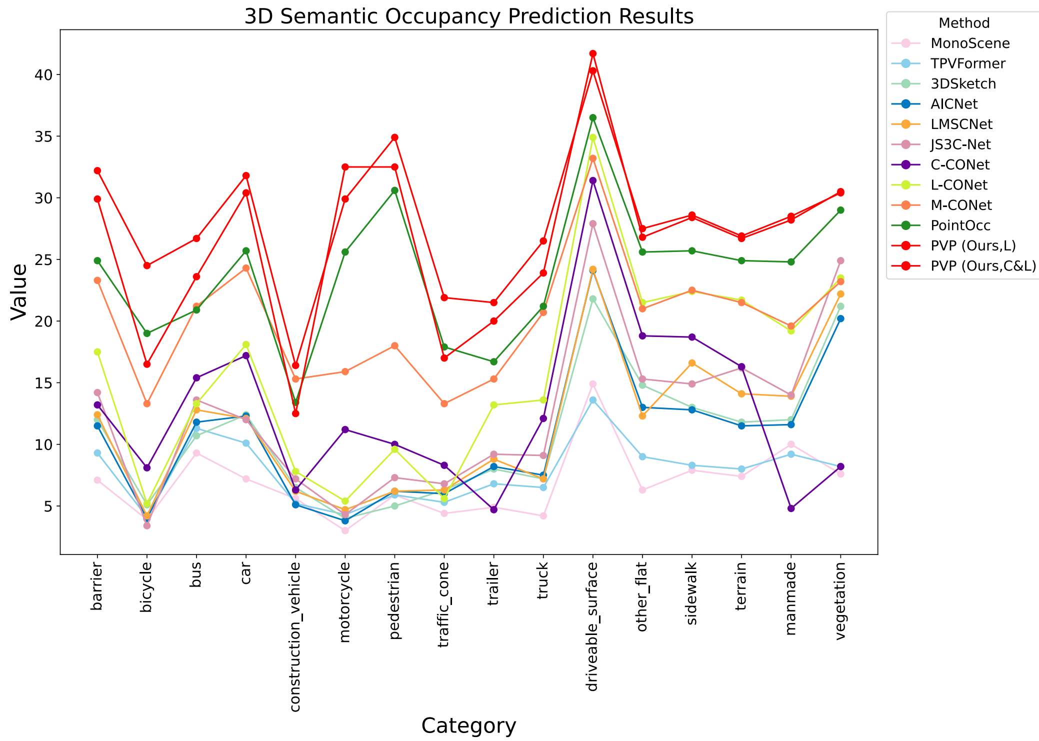

Figure 1. Performance Chart on 3D Occupancy Prediction: Our method *PVP* achieved the best results. C&L signifies that the input includes both camera images and LiDAR point clouds. L indicates that the input is exclusively LiDAR point clouds.

References

[1] 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 *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2020. 1

[2] Xiaofeng Wang, Zheng Zhu, Wenbo Xu, Yunpeng Zhang, Yi Wei, Xu Chi, Yun Ye, Dalong Du, Jiwen Lu, and Xingang Wang. Openoccupancy: A large scale benchmark for surrounding semantic occupancy perception. 2023. 1

[3] Whye Kit Fong, Rohit Mohan, Juana Valeria Hurtado, Lubing Zhou, Holger Caesar, Oscar Beijbom, and Abhinav Valada. Panoptic nuscenes: A large-scale benchmark for lidar panoptic segmentation and tracking. *IEEE Robotics and Automation Letters*, 7(2):3795–3802, 2022. 1