Supplementary Materials for DOtA

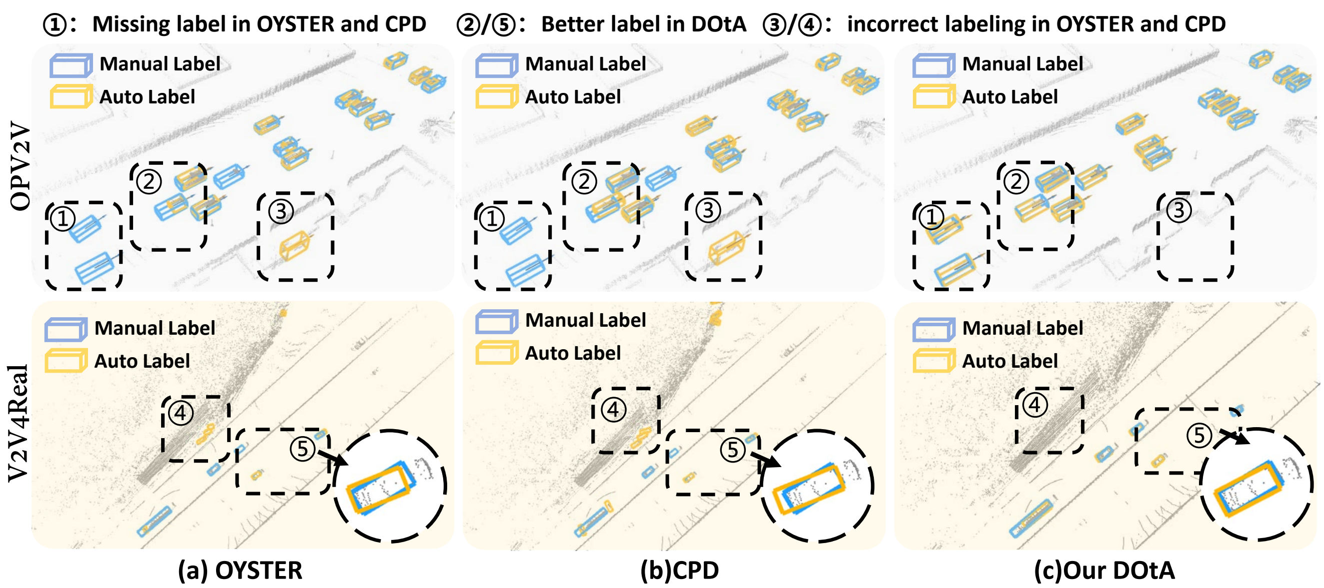

Figure 1. Visualization comparison of different labels on OPV2V $train$ split.

1. Label visualization comparison of different methods on two datasets.

As shown in the Fig. 1, we present a visual comparison of the automatic labeling results for different unsupervised methods on both the simulated dataset OPV2V [3] (top) and the real-world dataset V2V4Real [4] (bottom). For ease of observation, we have removed the ground point cloud from the original point cloud scene. In the simulated dataset OPV2V, within the areas under multi-agent synchronous observation, both existing unsupervised approaches and our approach can achieve commendable results in automatic labeling. However, objects located at the edge of the collaborative observation area are still affected by occlusion and the sparsity of point clouds, leading to incomplete structures in the point cloud descriptions. This phenomenon can degrade the performance of traditional unsupervised clustering-based algorithms. For instance, in area ①, due to the limited range of the point clouds output by clustering, traditional methods may discard the fitted bounding boxes for this part; in areas ② and ⑤, the incomplete structures resulting from clustering lead to poor bounding box fitting effects. Additionally, traditional fitting-based methods only consider the dimensions of the fitting box for selecting bounding boxes, which can lead to incorrect labels,$e.g.$, in many areas labeled as ③ and ④.

2. Additional Study on Robustness to localization noise.

Due to localization errors and communication delays, there is noise in real-world multi-agent collaborative observation. To further investigate the impact of this issue on our method, we follow the localization noise setting in Where2comm [1] (Gaussian noise with a mean of $0m$ and a standard deviation of $0m$-$0.6m$) and conduct experiments on OPV2V dataset to validate the robustness against realistic localization noise. As the results shown in Fig. 2, compared to the traditional unsupervised method CPD [2], our DOtA demonstrates stronger robustness. To further investigate the sources of robustness in DOtA, we statistically analyze the distribution of Intersection over Union (IOU) between DOtA labels and the ground truth labels, as depicted in the Fig. 3.

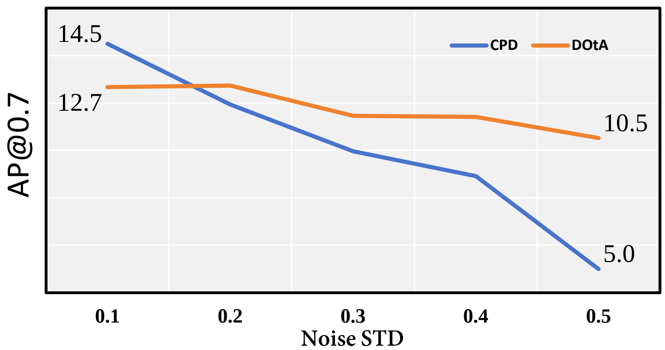

Figure 2. Additional Study on Robustness to localization noise.

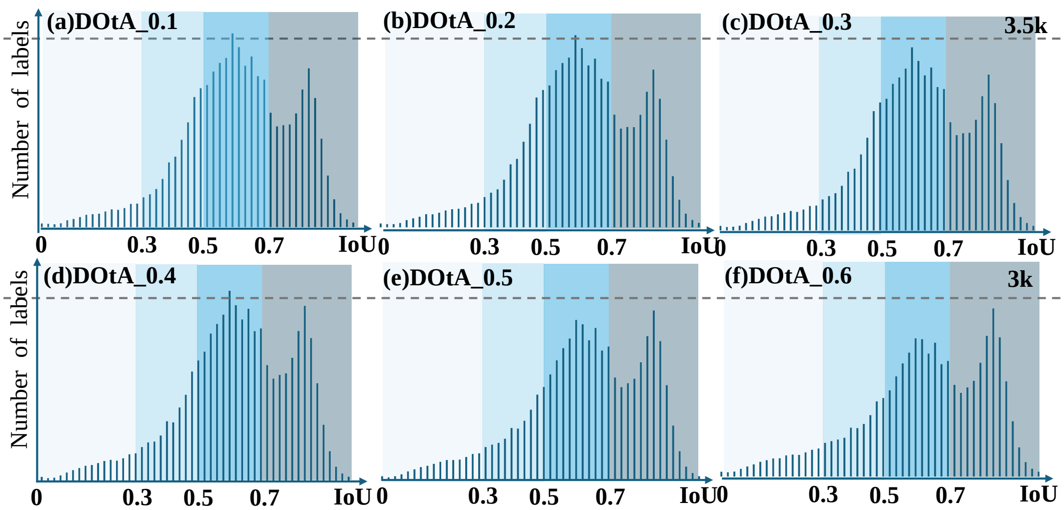

Figure 3. Distribution of Intersection over Union (IOU) between DOtA labels and the ground truth labels.

3. The study of collision tolerance parameter φ_r and alignment parameter φ_o.

In this section, we study the impact of different collision tolerance parameter and alignment parameter on detector performance. As shown in Tab. 1, we final set $\varphi_r = 0.1$ and $\varphi_o = 0.7$.

φ_r	0.01	0.05	0.10	0.15	0.20
AP@0.5	20.66	35.27	**40.01**	38.25	37.63

φ_o	0.1	0.3	0.5	0.7	0.8	0.9
AP@0.5	40.23	40.79	41.45	**43.27**	43.08	42.77

Table 1. The study of collision tolerance parameter φ_r and alignment parameter φ_o.

4. The study of scaling factor η_e.

In this section, we study the impact of different scaling factor on label recall rates. As shown in Tab. 2, the optimal scaling factors are [0.5, 0.2].

η_e	[0.4, 0.2]	[0.4, 0.3]	[0.5, 0.2]	[0.5, 0.3]	[0.6, 0.2]	[0.6, 0.3]	[0.6, 0.4]
Reacll@0.5	51.66	51.45	**51.87**	51.73	51.61	51.70	50.21

Table 2. The study of scaling factor η_e.

References

[1] Yue Hu, Shaoheng Fang, Zixing Lei, Yiqi Zhong, and Siheng Chen. Where2comm: Communication-efficient collaborative perception via spatial confidence maps. *Advances in neural information processing systems*, 35:4874–4886, 2022. 1

[2] Hai Wu, Shijia Zhao, Xun Huang, Chenglu Wen, Xin Li, and Cheng Wang. Commonsense prototype for outdoor unsupervised 3d object detection. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 14968–14977, 2024. 1

[3] Runsheng Xu, Hao Xiang, Xin Xia, Xu Han, Jinlong Li, and Jiaqi Ma. Opv2v: An open benchmark dataset and fusion pipeline for perception with vehicle-to-vehicle communication. In *2022 International Conference on Robotics and Automation (ICRA)*, pages 2583–2589. IEEE, 2022. 1

[4] Runsheng Xu, Xin Xia, Jinlong Li, Hanzhao Li, Shuo Zhang, Zhengzhong Tu, Zonglin Meng, Hao Xiang, Xiaoyu Dong, Rui Song, et al. V2v4real: A real-world large-scale dataset for vehicle-to-vehicle cooperative perception. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 13712–13722, 2023. 1