JiSAM: Alleviate Labeling Burden and Corner Case Problems in Autonomous Driving via Minimal Real-World Data

Supplementary Material

1. Visualization of detection results on corner cases.

We manually eliminate the motorcycle labels in training set of NuScenes [2] to evaluate how well JiSAM does in deal with objects not labeled in real dataset. We train the original Transfusion [1] and another model with JiSAM . Then we visualize the detection results in Figure 1. As shown in the two zoom-in parts, where ground truth boxes are highlighted with green color and the red ones are prediction from JiSAM. It can be found that JiSAM is able to detect corner cases in the evaluation sets and meanwhile maintain the performance of detecting other categories.

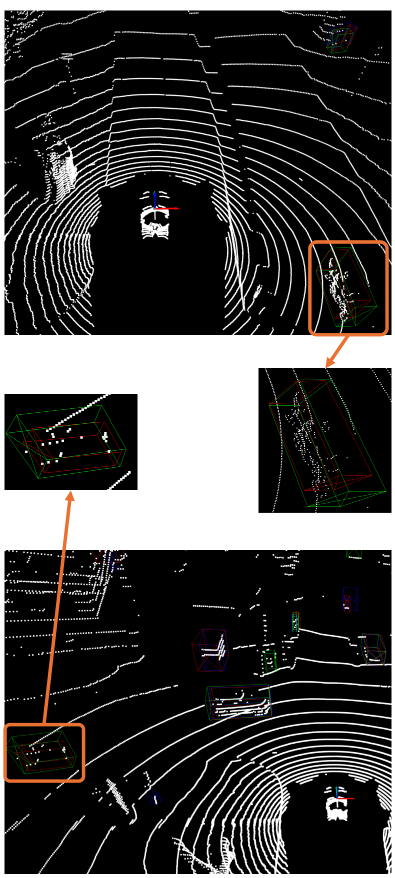

Figure 1. Visualization of two example detection results, where green boxes are ground trutch boxes and red ones are the prediction results of JiSAM , while blue ones are prediction results of Transfusion [1].

2. Visualization of the Inspiration of Sectorized Memory Alignment

We show an example of the inspiration of the sectorized memory alignment in Figure 2. It can be found that when objects in the same category with similar headings appear in the same partition of the surrounding environment, the point ditributions are quite similar. On the contrary when there are some changes in heading, the point distributions differ a lot.

3. Statistic of Real and Simulation Datasets

To help readers better understand the simulation dataset, we use the convention from [2] to calculate statistic of the simulation dataset collected in CARLA [4].

4. Other Results

Results on Once Dataset. We conduct experiments on ONCE [5] with CenterPoint (C.P.) [6]. Due to limited time, we collect a smaller simulation dataset with sensor of ONCE ($\sim$10,000 frames, $\frac{1}{8}$ as that for NuScenes). Together with 5% of the real data, we train JiSAM. Results are shown in table below where C.P. (5%) is the few shot baseline. It can be found that JiSAM achieves comparable performance to C.P. and large improvement over C.P. (5%). Looking into detailed categories, it can be found that JiSAM **surpasses** C.P. on Vehicle and Cyclist classes. For Pedestrian, JiSAM falls behind C.P. with all real data due to the scarce Pedestrian labels in 5% real data and relatively small synthetic dataset. As more synthetic data brings more gains (results in Table 2 in the main paper), we believe JiSAM will further improve if a larger simulation dataset is available.

Method	mAP	Veh.	Ped.	Cyc.
C.P.	64.00	75.98	49.35	66.22
C.P. (5%)	54.65	65.47	41.22	56.40
JiSAM	62.64	78.03	42.32	67.56

Table 1. Results on Once dataset [5].

Baseline Results for Corner Case Problem. In a sense, other sim2real works also help with corner case. JiSAM narrows sim2real gap and further improves corner case ability. We will rephrase to make it clearer. We conduct corner case experiments with sim2real work [3]. Results for Motorcycle (eliminated in real data) are 7.12% for [3] and

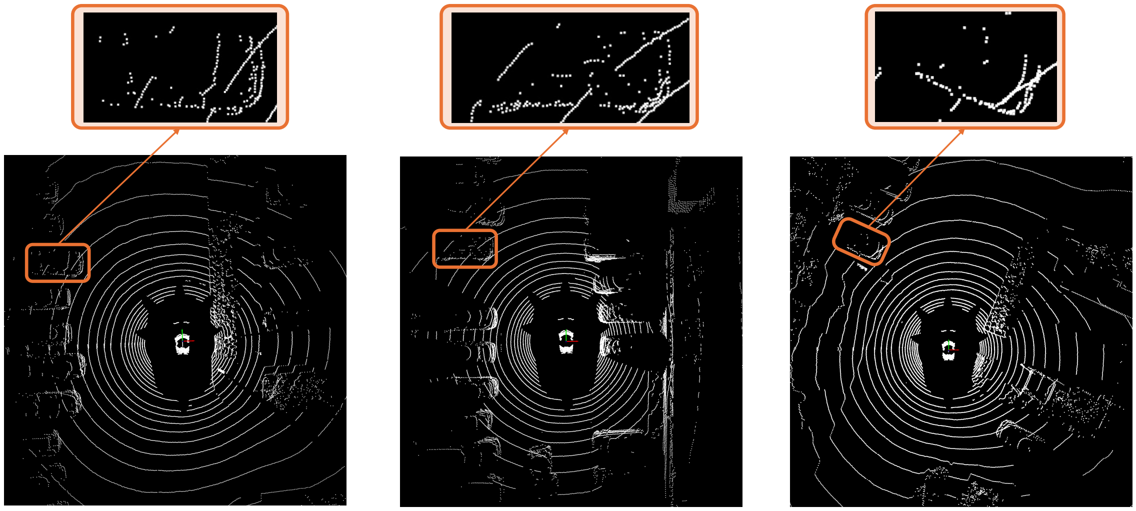

Figure 2. Illustration of the inspiration of sectorized memory alignment. In the first two examples, it can be found the point ditributions are quite similar of the objects, which is of the same category (car) with similar headings and in the same partition of surrounding environment. However, when we compare the first/second one with the third one, it can be found that some changes in heading result in much more different point distributions.

Dataset	Car	Truck	Bus	Cyclist	Pedestrian	Motorcycle
NuScenes [2]	294k	60k	11k	7k	148k	7k
CARLA	903k	87k	26k	46k	180k	67k

Table 2. Statistic of the NuScenes [2] dataset and the corresponding dataset we collected in CARLA [4].

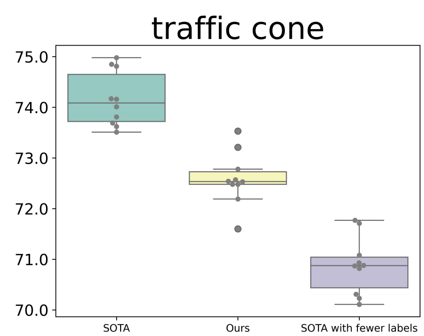

Figure 3. Additional results for traffic cone on NuScenes Dataset.

16.38% for JiSAM, showing the large improvement brought by JiSAM.

Baseline Results for Sim-Real Joint Training. We use our synthetic data and the same real subset to train [3] and ResimAD. Results of mAP are 63.95% for JiSAM, 62.33% for [3] and 59.09% for [7]. It can be found that JiSAM achieves the best performance, showing its strength.

Results for adding more real data. We use 5% of real data in NuScenes [2] to train JiSAM. The result is 64.28% in mAP.

Additional Results. We provide some additional detailed results on traffic cone category in Figure 3. The metrics are the same as those in main paper.

References

[1] Xuyang Bai, Zeyu Hu, Xinge Zhu, Qingqiu Huang, Yilun Chen, Hongbo Fu, and Chiew-Lan Tai. Transfusion: Robust lidar-camera fusion for 3d object detection with transformers. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 1090–1099, 2022. 1

[2] 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/CVF Conference on Computer Vision and Pattern Recognition*, pages 11621–11631, 2020. 1, 2

[3] Robert DeBortoli and others. Adversarial training on point clouds for sim-to-real 3d object detection. *IEEE Robotics and Automation Letters*, 2021. 1, 2

[4] Alexey Dosovitskiy, German Ros, Felipe Codevilla, Antonio Lopez, and Vladlen Koltun. CARLA: An open urban driving simulator. In *Proceedings of the 1st Annual Conference on Robot Learning*, pages 1–16, 2017. 1, 2

[5] Jiageng Mao et al. One million scenes for autonomous driving: Once dataset. Neurips Datasets and Benchmarks Track, 2024. 1

[6] Tianwei Yin et al. Center-based 3d object detection & tracking. CVPR, 2021. 1

[7] Bo Zhang, Xinyu Cai, Jiakang Yuan, Donglin Yang, Jianfei Guo, Renqiu Xia, Botian Shi, Min Dou, Tao Chen, Si Liu, et al. Resimad: Zero-shot 3d domain transfer for autonomous driving with source reconstruction and target simulation. *arXiv preprint arXiv:2309.05527*, 2023. 2