DOLPHINS: Dataset for Collaborative Perception enabled Harmonious and Interconnected Self-driving (Supplementary Material)

Ruiqing Mao[1][0000−0001−7169−3922], Jingyu Guo[1][0000−0002−1040−8259], Yukuan Jia[1][0000−0002−2378−9433], Yuxuan Sun[1,2][0000−0001−9720−6993], Sheng Zhou[1][0000−0003−0651−0071], and Zhisheng Niu[1][0000−0003−0420−2024]

[1] Department of Electronic Engineering, Tsinghua University, Beijing 100084, China
{mrq20,guojy18,jyk20}@mails.tsinghua.edu.cn
{sheng.zhou,niuzhs}@tsinghua.edu.cn
[2] School of Electronic and Information Engineering, Beijing Jiaotong University, Beijing 100044, China
yxsun@bjtu.edu.cn

1 The position of sensors on the vehicle

The position of sensors on the vehicle and the corresponding coordinates as mentioned in Sec. 3.2 are illustrated in Fig. 1.

Fig. 1: Sketch map of a fully equipped unit (take a vehicle as an example) and the coordinate system of each sensor.

2 Analysis of Sec. 4.3

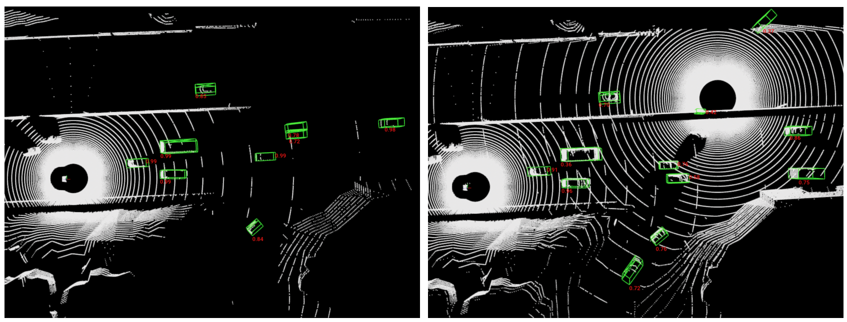
As for the rise of AP in hard and moderate tasks compared with easy tasks in 2D object detection, we further investigate the recall of each task and each method. Scenario 3 is a good case to study since about 40% of the cars and more than half of the pedestrians are moderate and hard objects, which leads to a more significant rise in these tasks. As Table 1 shows (noted that the moderate tasks include the easy ones, and the hard tasks include the former two), the recall of the easy objects is near 100%, which means all the easy objects can be identified in this scenario. However, in order to detect enough moderate and hard objects, all of the algorithms tend to raise more predictions, which in deed benefit the detection of these targets. However, too many predictions cause a lot of False Positives (FPs), which reduces the precision of an easy object. In Scenario 3, due to the high percentage of moderate and hard objects, the pros of more predictions overcome the cons of FPs in easy objects, which finally leads to the increment.

Table 1: Recall analysis of 2D object detection on Scenario 3

Scenario	Method		Car AP@IoU=0.5			Pedestrian AP@IoU=0.25		
			Easy	Moderate	Hard	Easy	Moderate	Hard
3	Ground Truths		3631	4699	5748	789	1678	1828
	Faster R-CNN	Detections	8179	8201	8225	2711	2737	2742
		Recall	99.2%	95.95%	87.67%	99.87%	98.33%	96.01%
	YOLOv3	Detections	12155	8910	8929	3239	3279	3293
		Recall	93.96%	93.23%	84.3%	99.87%	98.51%	97.65%
	YOLOX-S	Detections	21112	21244	21355	6267	6338	6377
		Recall	98.49%	93.23	86.46%	100%	98.51%	92.79%
	YOLOX-L	Detections	12155	12210	12265	5581	5638	5675
		Recall	98.98%	94.25%	85.35%	100%	96.61%	97.43%
	TOOD	Detections	18185	18232	18306	4774	4823	4842
		Recall	99.37%	95.84%	85.9%	99.87%	99.41%	98.36%
	DETR	Detections	102966	103760	104181	36834	37240	37419
		Recall	99.45%	99.05%	98.13%	100%	99.88%	98.96%

3 Analysis of Sec. 4.5

We conduct several experiments on the surprising performance loss in multi-view collaborative perception compared with stand-alone perception. This result suggests that the richer information is not equal to the higher precision. When we directly merge two point clouds from the ego vehicle and the RSU (as we did in Sec. 4.5), the increasing number of points helps to detect more occluded objects, that is, to improve the Recall rate. However, it may also lead to more False Positives due to the noises caused by the large parallax (as illustrated in Fig. 2). The results of multi-view collaborative perception indicate two potential tendencies. First of all, **the benefits of large object detection (such as cars) are larger than small ones (such as pedestrians)**. It might be because the small objects are more easily influenced by the additional points, i.e, more false positives since there are usually few points on a small object in the original point clouds. Secondly, **the benefits of hard tasks are larger than easy tasks**. As the multi-view collaborative perception aims to enhance the detection of occluded objects, those newly recalled objects are usually categorized as hard tasks. Thus, as for the easy tasks, the increase in AP brought by new recalls is smaller than the decrease brought by false positives, which leads to negative gains. These two tendencies suggest that the multi-view collaborative perception algorithms should take the noises into consideration, which is also one of our future works.



(a) Ego vehicle alone (b) Ego vehicle + RSU

Fig. 2: Compared with ego vehicle alone (a), the multi-view collaborative perception with point clouds from the RSU (b) contains three vehicles that are not detected in stand-alone perception but also introduces two false positives.