Commonsense Prototype for Outdoor Unsupervised 3D Object Detection

Supplementary Material

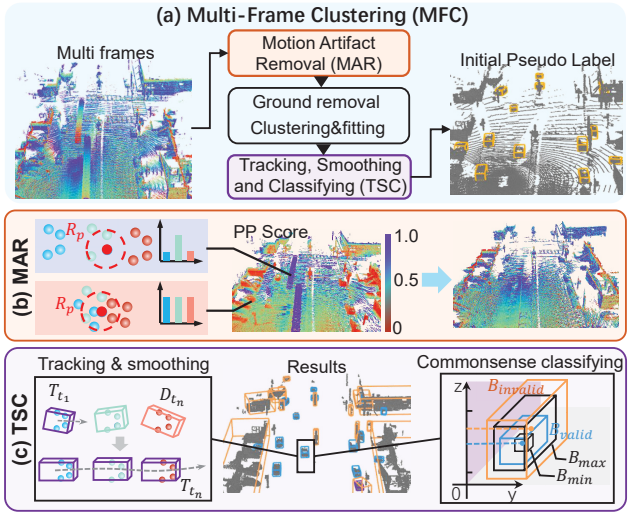

Figure 1. The MFC consists of motion artifact removal, clustering (ground removal, points clustering, and box fitting), and post-processing (tracking, smoothing, and classifying).

1. More Details of Method

More details of MFC. In our main paper section 3.1, we introduced the Multi-Frame Clustering (MFC) for initial label generation. For a more intuitive understanding, we provide a framework illustration in Fig. 1. Here we present more details of post-processing. As mentioned in our main paper, we pre-defined a set of class-specific size thresholds based on human commonsense to classify pseudo labels into different categories. Taking the WOD as an example, we predefine five categories: 'Discard Small', 'Pedestrian', 'Cyclist', 'Vehicle', and 'Discard Large'. Formally, for a cluster box b_j, we determine the class identity β by sequentially matching from the thresholds:

$$\beta = \begin{cases} DisSmall & h \leq 0.8, \\ Vehicle & 1 < h \leq 3, 0.5 < w \leq 3, 0.5 < l \leq 8, \\ Pedestrian & 0.8 < h \leq 2.3, 0.2 < w \leq 1., 0.2 < l \leq 1., \\ Cyclist & 1.4 < h \leq 2., 0.5 < w \leq 1., 1. < l \leq 2.5, \\ DisLarge & others. \end{cases} \tag{1}$$

Where l, w, h refers to the length, width, and height of b_j, respectively. The 'Discard Large' boxes mostly with trees and buildings are directly removed. The 'Discard Small' boxes contain both potential foreground objects and background objects. We then apply class-agnostic tracking to associate the small background objects with foreground trajectories, and enhance the consistency of objects' sizes by using temporal coherency.

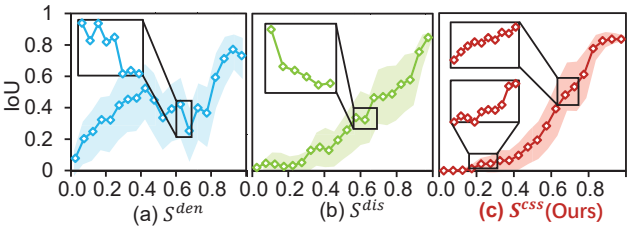

Figure 2. The comparison of different scoring methods.

More details of CSS scoring. In our main paper section 3.2, we presented the CSS scoring. To better understand how the CSS scoring approximates the IoU score, we present the IoU-score carve in Fig. 2, where we show three methods: density scoring (s^{den}), distance scoring (s^{dis}) and our CSS scoring (s^{css}). Intuitively, good scoring should keep consistent with IoU scoring. In other words, with the increase of score, the selected pseudo labels should have larger IoUs with the ground truth. We found that our CSS scoring keeps the most consistent increase along with the IoU increase. Here we also provide the length, width and height of the template box for calculating the Size Similarity in the main paper Eq. 3:

```
{
'Vehicle': [5.065, 1.86, 1.49],
'Pedestrian': [1.0, 1.0, 2.0],
'Cyclist': [1.9, 0.85, 1.8]
}
```

2. More Experimental Results

More visualization results. To better understand how our method improves detection results, here we present more visualization results. From Fig. 3, we observe that both the recognition and localization performance of our method (3.1-3.4) are much better than previous methods(1.1-1.4, 2.1-2.4), thanks to our CProto-based design.

BEV AP and 3D APH results on WOD validation set. Some fully supervised methods also reported the BEV AP L2 and 3D APH performance. Here we presented the results in Table 1 and Table 2, respectively. Our CPD outperforms the previous MODEST and OYSTER in both BEV AP L2 and APH L2 by a large margin, further demonstrating the effectiveness of our method.

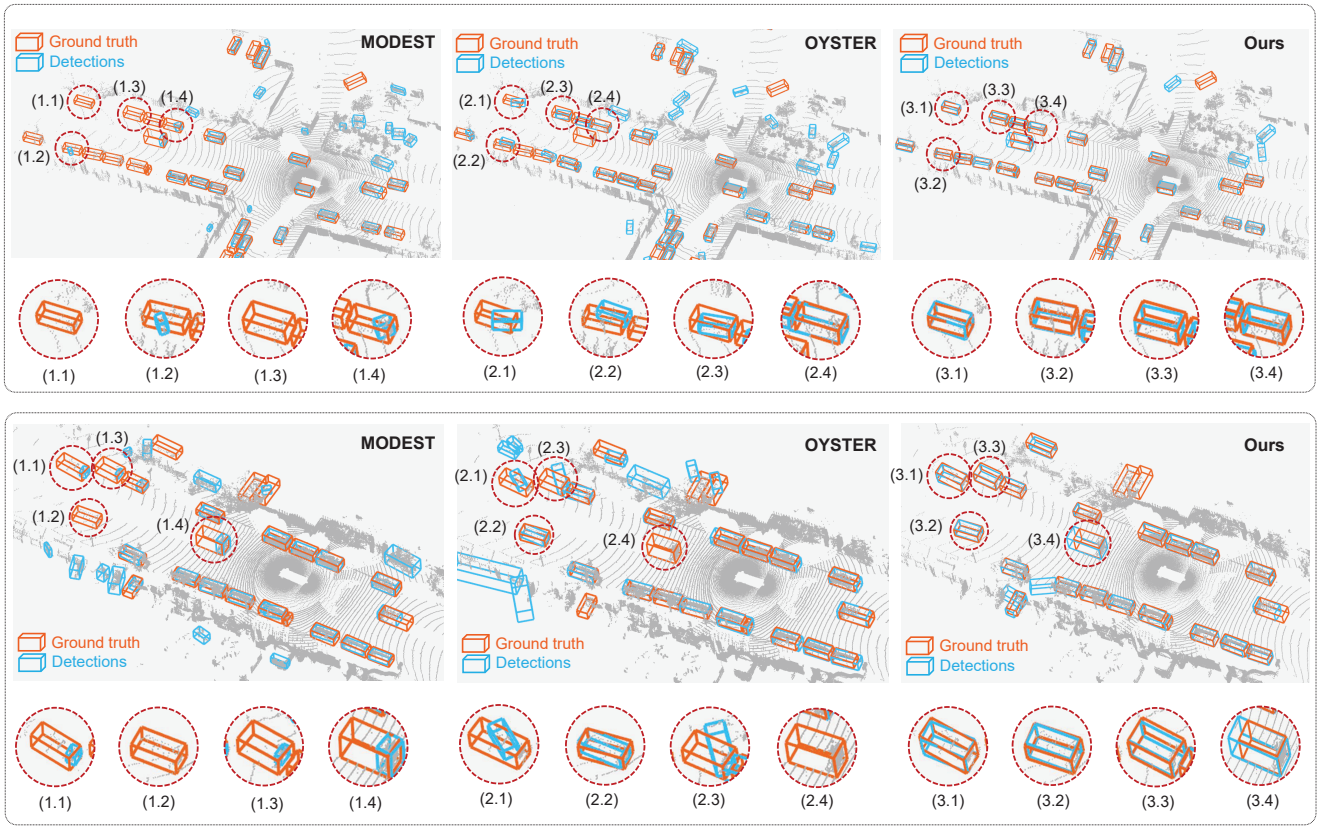

Figure 3. The visualization results predicted by different unsupervised detectors.

Method	Vehicle				Pedestrian				Cyclist			
	3D AP L2		BEV AP L2		3D AP L2		BEV AP L2		3D AP L2		BEV AP L2	
	$IoU_{0.5}$	$IoU_{0.7}$	$IoU_{0.5}$	$IoU_{0.7}$	$IoU_{0.3}$	$IoU_{0.5}$	$IoU_{0.3}$	$IoU_{0.5}$	$IoU_{0.3}$	$IoU_{0.5}$	$IoU_{0.3}$	$IoU_{0.5}$
DBSCAN	1.94	0.25	3.97	1.44	0.19	0	2.07	0	0.2	0	0.25	0.06
DBSCAN+init-train	14.87	2.29	20.6	11.95	1.35	0	6.49	0.1	0.43	0.2	0.73	0.24
MODEST	15.83	5.48	19.63	13.31	8.96	0.1	14.06	0.13	1.17	1.01	2.38	1.07
OYSTER	26.21	14.6	32.31	25.04	3.52	0.14	11.76	0.3	1.24	0.32	1.65	0.33
Proto-vanilla	31.58	18.36	34.91	28.88	14.62	8.59	17.94	15.9	3.8	3.31	4.05	3.48
CPD(Ours)	**50.67**	**32.13**	**52.66**	**47.48**	**20.01**	**15.22**	**20.21**	**17.26**	**5.61**	**4.87**	**5.68**	**5.22**

Table 1. 3D AP L2 and BEV AP L2 results on WOD validation set.

Method	Vehicle 3D APH				Pedestrian 3D APH				Cyclist 3D APH			
	L1		L2		L1		L2		L1		L2	
	$IoU_{0.5}$	$IoU_{0.7}$	$IoU_{0.5}$	$IoU_{0.7}$	$IoU_{0.3}$	$IoU_{0.5}$	$IoU_{0.3}$	$IoU_{0.5}$	$IoU_{0.3}$	$IoU_{0.5}$	$IoU_{0.3}$	$IoU_{0.5}$
MODEST	16.43	4.25	14.04	3.63	5.59	0.11	4.18	0.05	1.07	0.82	0.45	0.07
OYSTER	28.56	12.87	25.01	12.54	3.12	0.12	2.03	0.06	0.87	0.24	0.82	0.21
Proto-vanilla	32.34	19.2	29.71	16.23	9.12	6.3	8.12	5.26	2.84	2.51	2.73	2.42
CPD(Ours)	**54.19**	**34.97**	**46.99**	**30.09**	**12.01**	**9.24**	**10.06**	**7.68**	**3.68**	**3.26**	**3.55**	**3.14**

Table 2. 3D APH results on WOD validation set.