Curricular Object Manipulation in LiDAR-based Object Detection
——— Supplemental File ———

1. Supplementary

In the supplemental file, we provide more details of our work to supply our main paper, including

- **Details of our object clustering** about the occupancy ratio and the clustering rules in our experiments,
- **More visualization results** of our clustering rules and different clustering groups,
- **Deeper analysis of COMAug** about the influence of the hyper parameters and the effectiveness of anti-curriculum learning.

1.1. Details of our object clustering

Explanation of occupancy ratio. In our main paper, we have clarified that the occupancy ratio f_o is the ratio of non-empty 3D voxels inside the bounding box of an object. Considering the shape differences of the point cloud objects, we employ different voxel division strategies for vehicles and pedestrians in our experiments.

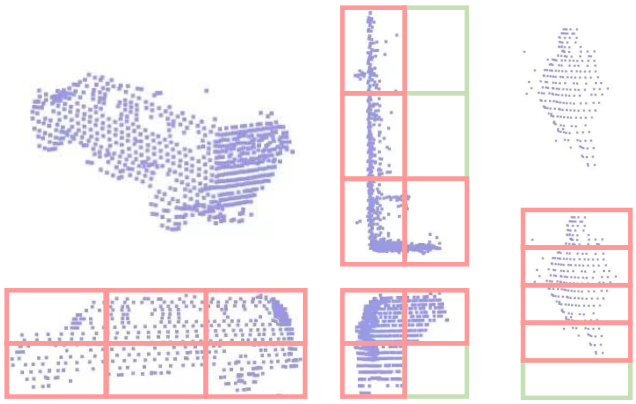

Figure 1. Voxel divisions for the Vehicle and Pedestrian. Red boxes are non-empty boxes and the green ones are empty. Occupancy ratio denotes the ratio of non-empty boxes.

We illustrate our voxel division strategies in Fig. 1. For the vehicle, we divide the bounding box into $3 \times 2 \times 2$ voxels along the length, width, and height. For Pedestrian we divide the box into 5 voxels by height.

The clustering rules in our experiments. We cluster objects in the ground-truth database into groups by the distance f_d, size f_s, relative angle f_a, and occupancy ratio f_o. In our experiments,
(1) the distances of objects are divided into $[0m, 30m)$, $[30m, 50m)$, and $[50m, +\infty)$.
(2) Following [1], the sizes of the vehicles are divided into $[0m, 4m)$, $[4m, 8m)$, and $[8m, +\infty)$.
(3) For the relative angle f_a, we divide $f_a(mod\frac{\pi}{2})$ of the vehicles into $[0, \frac{\pi}{6})$, $[\frac{\pi}{6}, \frac{\pi}{3})$, and $[\frac{\pi}{3}, \frac{\pi}{2}]$.
(4) The occupancy ratios of the objects range from 0 to 1. We divide them into $[0, 0.2)$, $[0.2, 0.4)$, $[0.4, 0.6)$, $[0.6, 0.8)$, and $[0.8, 1]$.

Importantly, in the Tab. 1 of our main paper, we use the distance and occupancy ratio to cluster pedestrians. In this way, the pedestrians are clustered into 15 groups. In the Tab. 2 of our main paper, all of the factors are utilized and the vehicles are clustered into 135 groups. In the Tab. 3 of our main paper, different combinations of (1), (2), (3), and (4) are implemented for clustering the vehicle class.

1.2. A deeper analysis of COMAug

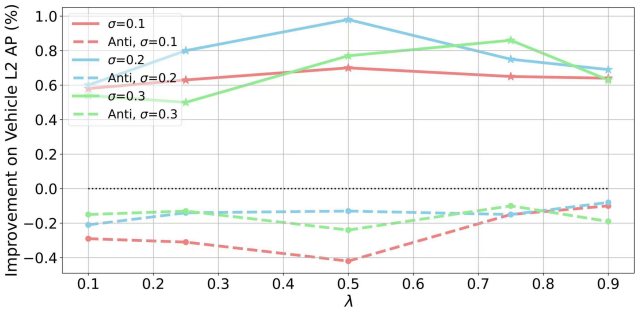

Figure 2. The performance improvement with different hyper parameters settings in COMAug. Solid lines represent curriculum learning, while dashed lines represent anti-curriculum learning.

Effects of the parameter λ and σ. The λ in our COMAug controls the pacing speed, which determines how early the probabilities of sampling semi-hard/hard objects are raised. The pacing speed is faster with a larger λ. As is shown in Fig. 2, when $\sigma = 0.1, 0.2$, setting λ to 0.5 can lead to better performances. When $\sigma = 0.3$, assigning λ with 0.75

is optimal. Overall, the λ should be adjusted to a moderate value, where the pacing speed is neither too fast nor too slow. This conclusion is similar to that in other curriculum learning works [2, 4].

The parameter σ control the diversity of the sampling. When $\sigma \to 0$, only one group whose score s_i equals μ_t will be sampled at each training epoch t. When $\sigma \to \infty$, every group will have an equal probability of being sampled. In Fig. 2, we set σ to 0.1, 0.2, and 0.3. The experiments show that $\sigma = 0.2$ can typically achieve better performances.

Effects of anti-curriculum learning. In the introduction of our main paper, we claim that selecting too many hard examples at early stages may overwhelm the training, while selecting too many easy samples at the later stages may slow the model convergence. Here we employ anti-curriculum COMAug to validate this point.

In contrast to curriculum learning, anti-curriculum learning emphasizes training samples from difficult to easy. Even if counter-intuitive, [3, 5, 6] have shown the effectiveness of anti-curriculum in certain scenarios. However, most works in curriculum learning demonstrate the superiority of curriculum than anti-curriculum or random ordering.

We conduct experiments to validate whether anti-curricular can be effective for our task. Specifically, we reverse the sorted group scores $\{s_i\}_1^G$, so that our COMAug selects increasingly easy objects for augmentation as training proceeds. The settings of λ and σ are kept the same as the curricular experiments in Fig. 2. Through the experiment results, we find that anti-curriculum is inferior to standard training. Notice that when $\lambda = 0.1$ and $\sigma = 0.1$, many difficult objects are sampled during the early training stage. In this case, anti-curriculum brings a 0.27% performance drop. This phenomenon demonstrates that the hard-sample-first strategy can hinder convergence. Moreover, However, the performances are still worse than the standard training. It shows that selecting too many easy objects during the later stage does not benefit the training.

The minor improvement lead by COM on Pedestrian class. The improvements of COMAug+COMLoss over COMAug are not as significant for class pedestrian as those for class vehicle. The phenomenon could be caused by data distribution as shown in Tab. 1. For vehicle, each frame contains 30.2 source objects and 1.8 augmented objects on average. In contrast, numbers for pedestrian are 14.1 and 5.9. Recall that COMLoss essentially improves model performances by treating objects differently based on their difficulties. COMAug, on the other hand, selects objects of similar difficulty levels (suitable for current training) for augmentation. That decreases the variety of object difficulties and weakens the effects of COMLoss. As a result, the improvements for class pedestrian are less evident due to its high ratio of augmented objects (29.5%), while class vehicle is less affected because of its low ratio (5.6%). Besides, we found that the phenomenon does not occur for GT-Aug. The reason could be that GT-Aug selects objects at random, which results in diverse object difficulties.

Table 1. Average number of objects per frame in Waymo Dataset.

	Source	Augmented	Ratio of augmented object
Vehicle	30.2	1.8	5.6%
Pedestrian	14.1	5.9	29.5%

1.3. More visualization results

In this work, we cluster the vehicles according to four factors: distance, size, relative angle, and occupancy ratio. To demonstrate how the four factors affect the distribution of the vehicle point clouds, we present a few examples in Fig. 3. We can observe the following four phenomena: (1) The objects' point clouds become sparser as the distance increases. (2) The sizes and shapes of the vehicles can vary greatly. (3) Our relative angle can reveal the faces of the vehicle being observed. (4) Objects with low occupancy ratios can be difficult to recognize.

Further, we cluster the pedestrians into 15 groups according to the distance and occupancy ratio in the Tab. 1 of our main paper. In Fig. 4 we provide the visualization examples of the pedestrian groups. The sizes of the pedestrians are typically similar, but their point clouds vary widely in sparsity and shape integrity.

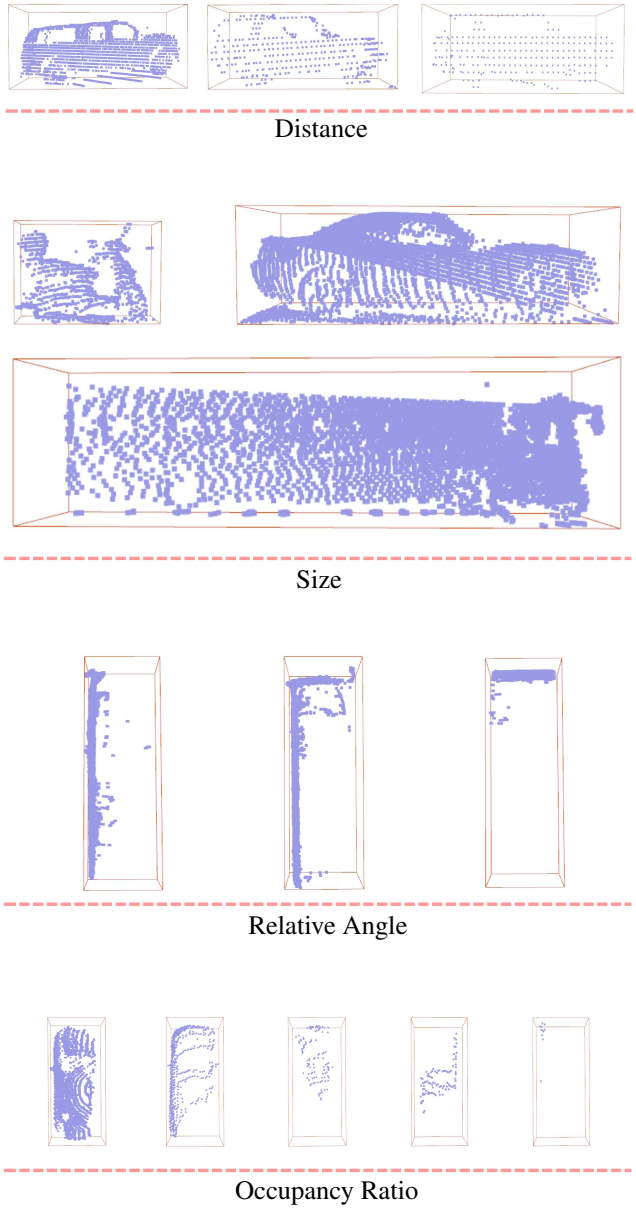

Figure 3. Visualization of the vehicles with different distances, sizes, relative angles, and occupancy ratios.

Figure 4. Visualization of the 15 pedestrian groups.

References

[1] Lue Fan, Ziqi Pang, Tianyuan Zhang, Yu-Xiong Wang, Hang Zhao, Feng Wang, Naiyan Wang, and Zhaoxiang Zhang. Embracing single stride 3d object detector with sparse transformer. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 8458–8468, 2022. 1

[2] Yuge Huang, Yuhan Wang, Ying Tai, Xiaoming Liu, Pengcheng Shen, Shaoxin Li, Jilin Li, and Feiyue Huang. Curricularface: adaptive curriculum learning loss for deep face recognition. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 5901–5910, 2020. 2

[3] Tom Kocmi and Ondrej Bojar. Curriculum learning and minibatch bucketing in neural machine translation. *arXiv preprint arXiv:1707.09533*, 2017. 2

[4] Yajing Kong, Liu Liu, Jun Wang, and Dacheng Tao. Adaptive curriculum learning. In *International Conference on Computer Vision*, pages 5067–5076, 2021. 2

[5] Xuan Zhang, Gaurav Kumar, Huda Khayrallah, Kenton Murray, Jeremy Gwinnup, Marianna J Martindale, Paul McNamee, Kevin Duh, and Marine Carpuat. An empirical exploration of curriculum learning for neural machine translation. *arXiv preprint arXiv:1811.00739*, 2018. 2

[6] Xuan Zhang, Pamela Shapiro, Gaurav Kumar, Paul McNamee, Marine Carpuat, and Kevin Duh. Curriculum learning for domain adaptation in neural machine translation. *arXiv preprint arXiv:1905.05816*, 2019. 2