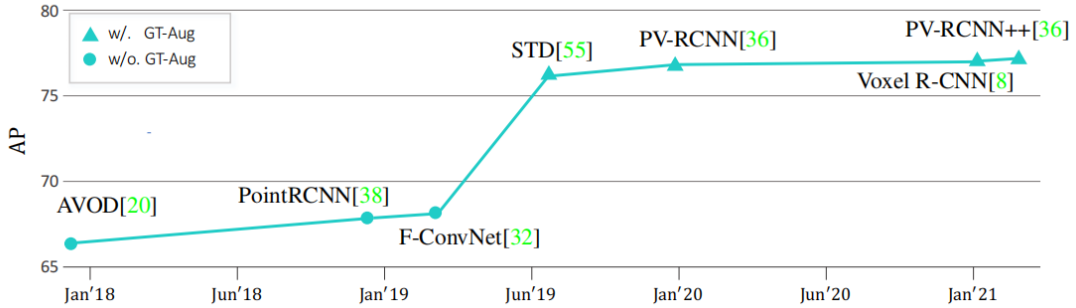
Curricular Object Manipulation in LiDAR-Based Object Detection

(a) Early stage.

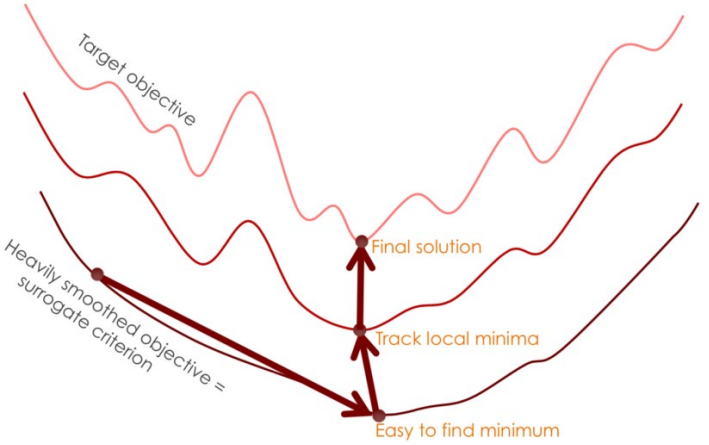
(b) Later stage.

Prerequisites

- GT-Aug: Ground Truth Augmentation (Copy Paste)

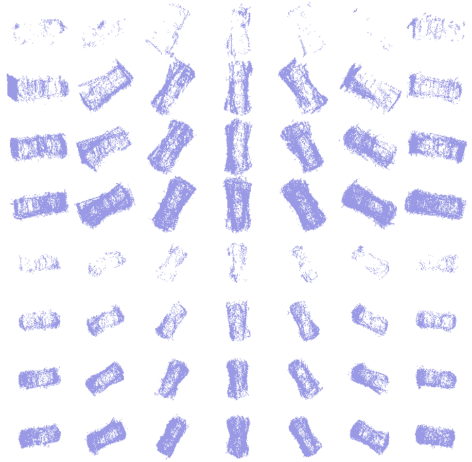


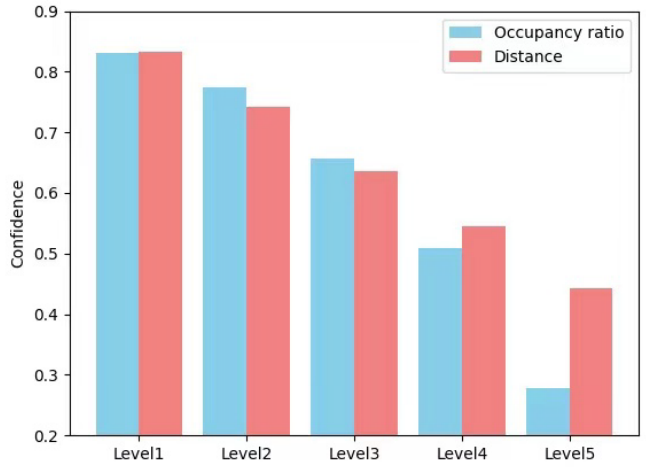
- Curriculum Learning : Learning in an easy to hard manner



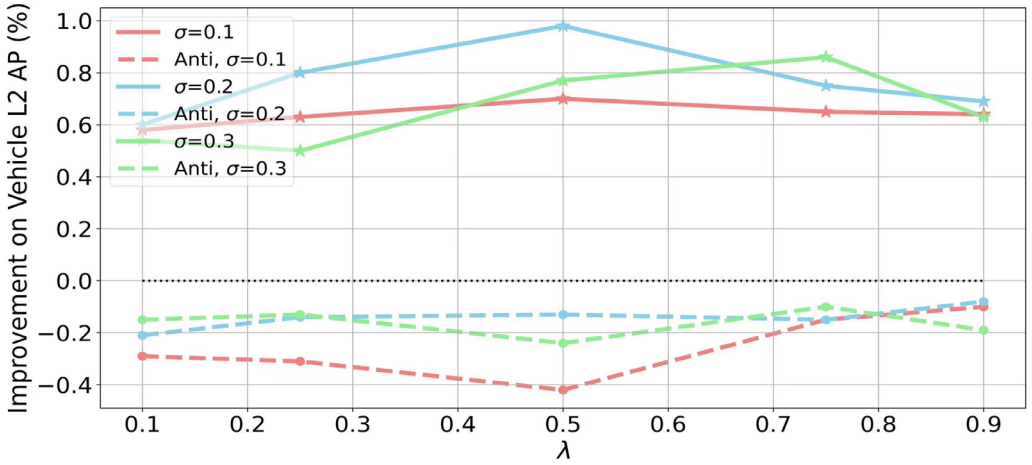
Observations

- The learning difficulties of objects vary greatly : 10% objects contain <= 5 points





- Simply sampling hard objects at early stage has no benefit



Questions

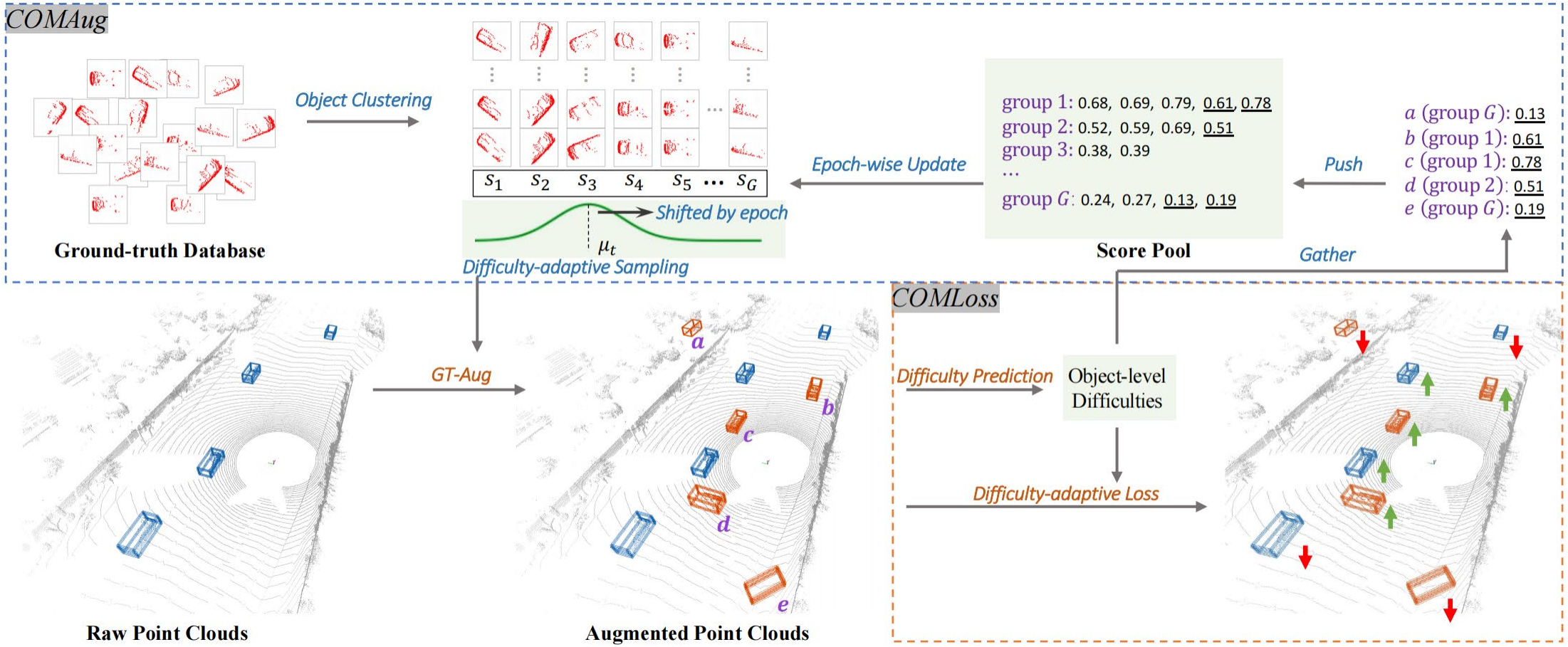
- How to adjust the sampling strategies according to different training stages ?

- How to properly handle augmented objects as well as **original objects** ?

Method

- Curriculum learning: we embed the easy-to-hard training strategy into both **loss design** and **augmentation** process

Our framework



Functions

- COMLoss

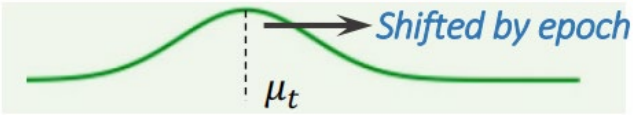
$$w = 1 + h_t \cdot (1 - e^{\beta \cdot \tilde{s}})/(1 + e^{\beta \cdot \tilde{s}}),$$
$$\text{where } h_t = H \cdot (t_r - t)/T.$$

$$\mathcal{L} = \frac{1}{N}\left(\mathcal{L}_n + \sum_{p=1}^{P} w_p \cdot (\mathcal{L}_c^p + \mathcal{L}_r^p) + \sum_{q=1}^{Q} w_q \cdot (\mathcal{L}_c^q + \mathcal{L}_r^q)\right)$$

- COMAug

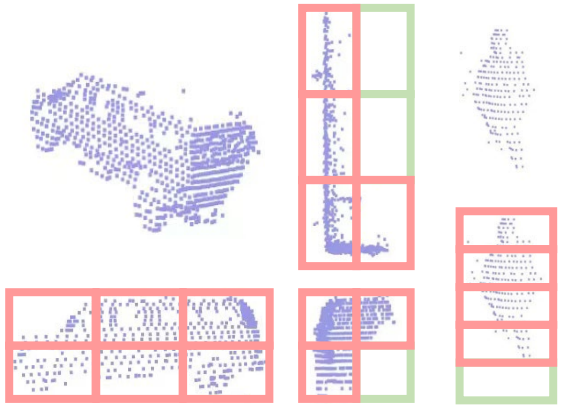
$$p_g = \exp\{-(\tilde{s}_g - \mu^t)^2/(2\sigma^2)\} \qquad \mu_t = s_g, \text{ where } g = \min(\lfloor(\lambda \cdot t/T) \cdot G\rfloor, G),$$

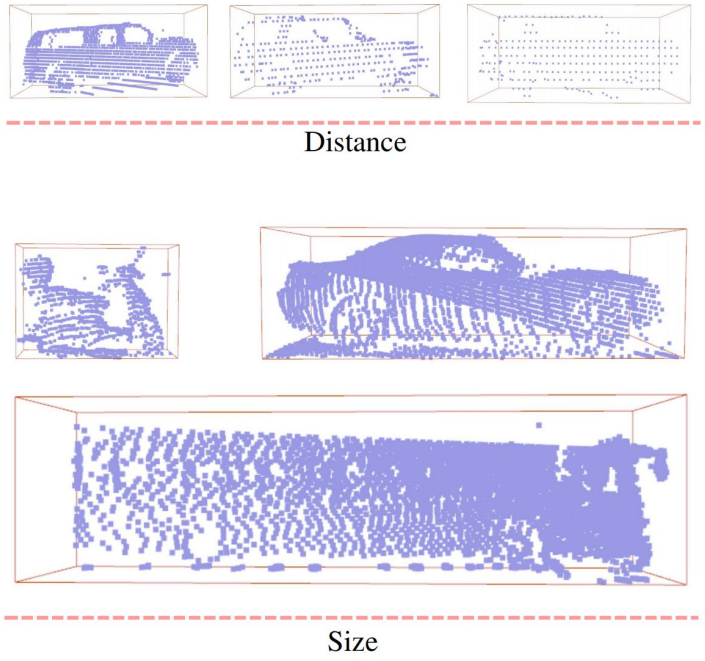
$$p^g = (p_g \cdot n_g)/(\sum_{i=1}^{G} p_i \cdot n_i).$$

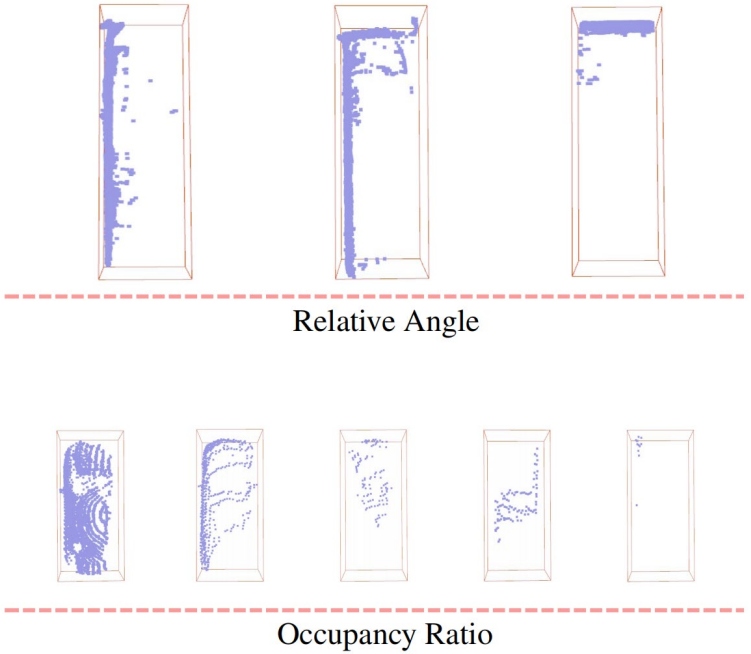


Clustering objects by well-designed heuristics

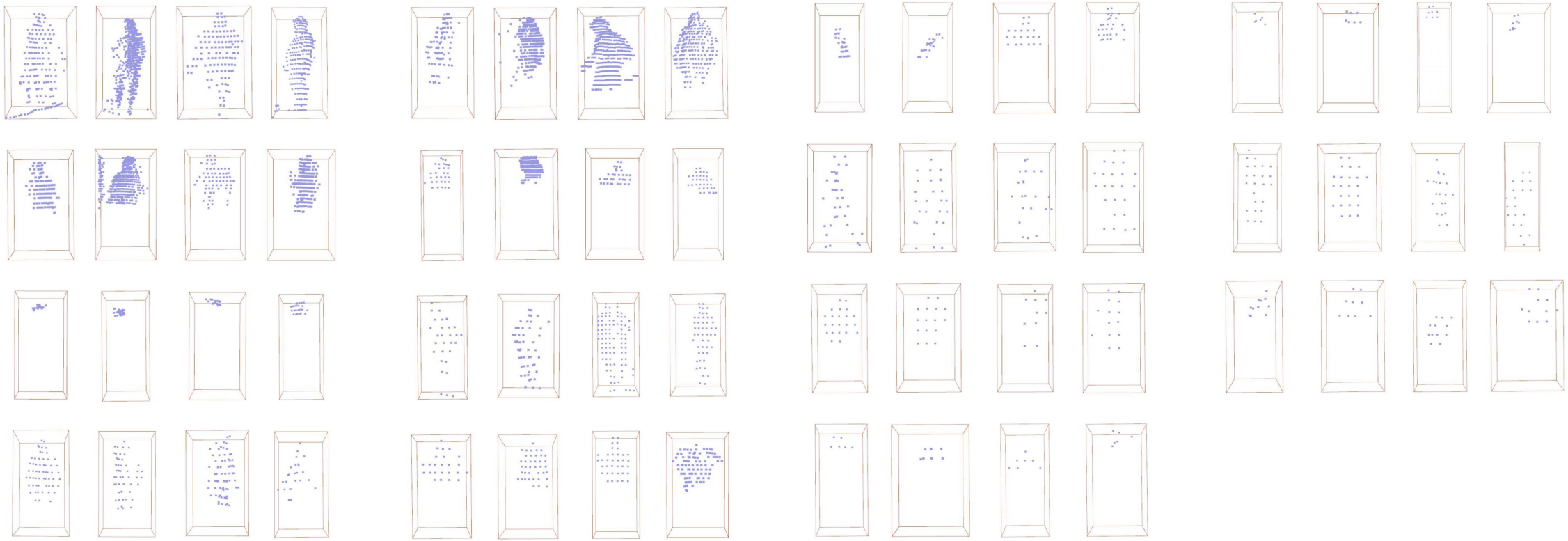
1.Distance 2.Object Size 3.Relative Angle 4.Occupancy Ratio







Visualization of pedestrian groups



Quantitative results on Waymo

Detector	Method	Vehicle	
		3D AP_{L1} (%)	3D AP_{L2} (%)
SECOND [51]	+ *None*	71.72	63.26
	+ *GT-Aug*	72.40 (+0.68)	63.95 (+0.69)
	+ *COMLoss*	72.82 (+1.10)	64.35 (+1.09)
	+ *COMAug*	73.07 (+1.35)	64.57 (+1.31)
	+ *COM*	73.32 (+1.60)	65.34 (+2.08)
PointPillars [21]	+ *None*	69.28	61.04
	+ *GT-Aug*	70.00 (+0.72)	61.70 (+0.66)
	+ *COMLoss*	70.78 (+1.50)	62.43 (+1.39)
	+ *COMAug*	71.01 (+1.73)	62.88 (+1.84)
	+ *COM*	71.38 (+2.10)	62.96 (+1.92)
CenterPoint [56]	+ *None*	70.97	62.54
	+ *GT-Aug*	71.05 (+0.08)	62.66 (+0.12)
	+ *COMLoss*	71.57 (+0.60)	63.13 (+0.59)
	+ *COMAug*	72.03 (+1.06)	63.61 (+1.07)
	+ *COM*	72.15 (+1.18)	64.29 (+1.75)

Table 2. Performances on class vehicle. In green are the gaps of at least +1.0 point.

Detector	Method	Pedestrian	
		3D AP_{L1} (%)	3D AP_{L2} (%)
SECOND [51]	+ *None*	62.37	54.05
	+ *GT-Aug*	65.40 (+3.03)	57.07 (+3.02)
	+ *COMLoss*	66.23 (+3.86)	57.77 (+3.72)
	+ *COMAug*	66.46 (+4.09)	57.97 (+3.92)
	+ *COM*	66.66 (+4.29)	58.14 (+4.09)
PointPillars [21]	+ *None*	64.36	55.98
	+ *GT-Aug*	65.54 (+1.18)	57.20 (+1.22)
	+ *COMLoss*	66.19 (+1.83)	57.80 (+1.82)
	+ *COMAug*	66.75 (+2.39)	58.40 (+2.42)
	+ *COM*	66.81 (+2.45)	58.52 (+2.54)
CenterPoint [56]	+ *None*	72.79	65.02
	+ *GT-Aug*	73.62 (+0.83)	65.84 (+0.82)
	+ *COMLoss*	74.16 (+1.37)	66.36 (+1.34)
	+ *COMAug*	75.34 (+2.55)	67.66 (+2.64)
	+ *COM*	75.43 (+2.64)	67.76 (+2.74)

Table 1. Performances on class pedestrian. In green are the gaps of at least +2.0 point.

Ablation study for COMAug

- Different clustering strategies
- Visualization of our sampling probability

	Occup.	Dist.	Angle	Size	3D AP_{L1} (%)	3D AP_{L2} (%)
A0					71.05	62.66
A1	✓				71.36 (+0.31)	62.93 (+0.33)
A2		✓			71.51 (+0.46)	63.11 (+0.45)
A3			✓		71.33 (+0.28)	62.92 (+0.32)
A4				✓	71.38 (+0.33)	62.98 (+0.38)
A5	✓	✓			71.68 (+0.63)	63.24 (+0.31)
A6	✓	✓	✓		71.86 (+0.81)	63.41 (+0.75)
A7	✓	✓		✓	71.89 (+0.84)	63.44 (+0.78)
A8	✓	✓	✓	✓	72.03 (+0.98)	63.61 (+0.95)

Table 3. Performances with different clustering factors. Occup. and Dist. stand for occupancy ratio and distance, respectively.

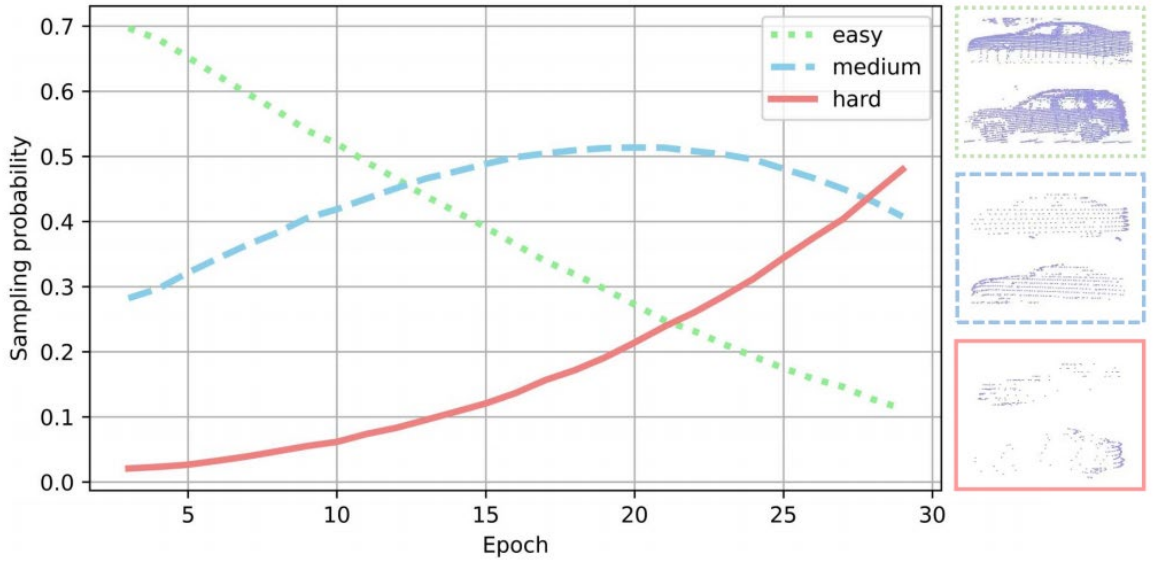
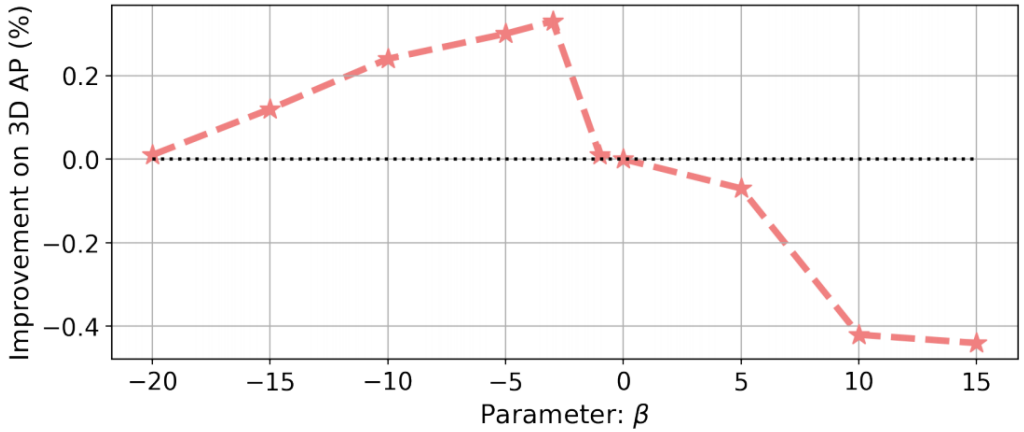


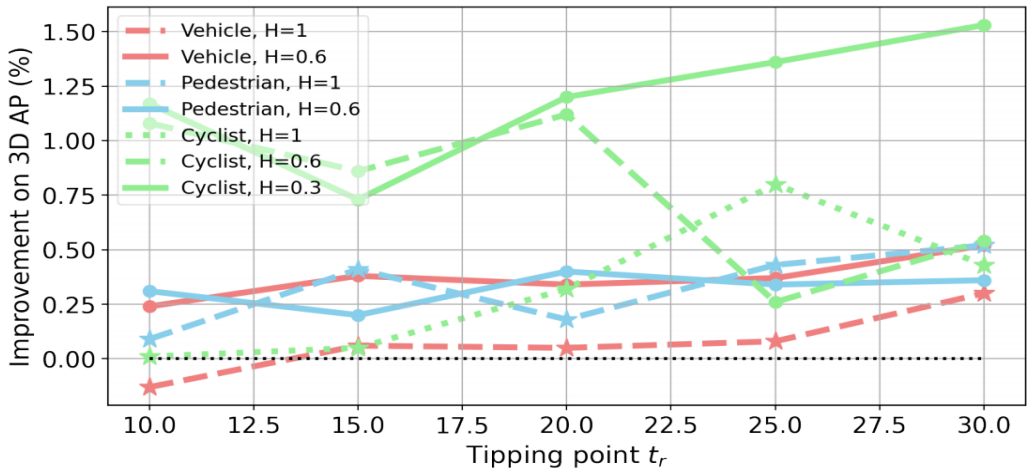
Figure 8. Visualization of sampling probability.

Ablation study for COMLoss

- Whether the anti-curriculum learning is effective ?



- When to start emphasizing difficult samples ?



TO DO

- Consider the regression difficulty
- Design more appropriate grouping strategies
- Extend to other LiDAR-related tasks, such as LiDAR segmentation and multi-modality 3D object detection

Thank You !