Supplementary Material for Hierarchical Supervision and Shuffle Data Augmentation for 3D Semi-Supervised Object Detection

Chuandong Liu[1,2], Chenqiang Gao[1,2], Fangcen Liu[1,2], Pengcheng Li[1,2], Deyu Meng[3,4], Xinbo Gao[1]
[1]School of Communication and Information Engineering, Chongqing University of Posts and Telecommunications, Chongqing, China
[2]Chongqing Key Laboratory of Signal and Information Processing, Chongqing, China
[3]Xi'an Jiaotong University, Xi'an, China
[4]Macau University of Science and Technology, Taipa, Macau

1. Discuss on Other Augmentation Methods for Point Cloud

The Mixup-based [8] augmentation methods have been extensively studied in the field of image classification and widely applied in 2D semi-supervised object detection [6] task. Following this idea, there have been several explorations in point cloud tasks as well. PointMixup [1] first applied the idea of Mixup to point cloud and achieved linear interpolation through the optimal allocation. Mix3D [3] balances global contextual information and local geometric information to achieve high-performance models. In addition, PointCutMix [9] proposes two different ways of replacing points to mix two point clouds. The latest SageMix explores salient regions in two point clouds and smoothly combines them into a continuous shape. However, these methods mainly focus on point cloud classification and segmentation tasks. For outdoor 3D object detection task, objects are usually naturally separated [5], and merging two point cloud scenes will cause overlaps between objects (*e.g.*, two vehicles are rarely overlapped in 3D reality). Therefore, to the best of our knowledge, the above Mixup-based point cloud augmentation methods cannot be directly applied to detection tasks, which is the direction for our future research.

2. Visualization of Dynamic Dual-Threshold

To better understand the dual-threshold hierarchical supervision in intuitive, we visualize the dynamic threshold changes during the training process in Fig. 1, where a solid line of a certain color represents a high threshold, and the dotted line of the same color represents a low threshold.

3. Additional Experimental Results

(1) Additional experiments on the Waymo Dataset. We additionally test the Voxel-RCNN [2] on 1% of the Waymo [7] dataset, and the results in Tab. 1 still show the superiority of our method, which validates its generalization.

(2) If the shuffle data augmentation (SDA) strategy is also effective for full supervision training ? To verify the effect of the SDA on fully-supervised 3D object detector, we inset the SDA into the PV-RCNN [4] and the results are listed in Tab. 2, which shows that the superiority of SDA in the supervised framework is not as obvious as in the semi-supervised framework. This is due to that the design of the strong augmentation in the student branch module has two main purposes: (1) strong enough to make a significant difference with weakly augmented samples of the teacher branch and (2) not too strong to ensure effective supervision information transmission.

References

[1] Yunlu Chen, Vincent Tao Hu, Efstratios Gavves, Thomas Mensink, Pascal Mettes, Pengwan Yang, and Cees GM Snoek. Pointmixup: Augmentation for point clouds. In *ECCV*, pages 330–345. Springer, 2020. 1

[2] Jiajun Deng, Shaoshuai Shi, Peiwei Li, Wengang Zhou, Yanyong Zhang, and Houqiang Li. Voxel r-cnn: Towards high performance voxel-based 3d object detection. In *AAAI*, pages 1201–1209, 2021. 1, 2

[3] Alexey Nekrasov, Jonas Schult, Or Litany, Bastian Leibe, and Francis Engelmann. Mix3d: Out-of-context data augmentation for 3d scenes. In *3DV*, pages 116–125. IEEE, 2021. 1

[4] Shaoshuai Shi, Chaoxu Guo, Li Jiang, Zhe Wang, Jianping Shi, Xiaogang Wang, and Hongsheng Li. Pv-rcnn: Point-voxel feature set abstraction for 3d object detection. In *CVPR*, pages 10529–10538, 2020. 1, 2

1% Data (∼ 1.4k scenes)	Veh. (LEVEL 1)	Veh. (LEVEL 2)	Ped. (LEVEL 1)	Ped. (LEVEL 2)	Cyc. (LEVEL 1)	Cyc. (LEVEL 2)
Voxel-RCNN [2]	49.02/48.03	42.36/41.50	41.16/32.81	34.73/27.66	5.84/5.61	5.62/5.40
Ours (Voxel-RCNN-based)	**54.89/54.06**	**48.28/47.53**	**43.86/37.84**	**36.59/31.56**	**17.47/16.73**	**16.72/16.01**

Table 1. Results on the Waymo for the Voxel-RCNN detector

Model	Data	3D Detection (Car)			3D Detection (Ped.)			3D Detection (Cyc.)		
		Easy	Mod	Hard	Easy	Mod	Hard	Easy	Mod	Hard
PV-RCNN [4]	100%	**92.10**	84.36	**82.48**	**63.12**	54.84	**51.78**	89.10	70.38	66.01
PV-RCNN [4] with SDA	100%	91.91	**84.57**	82.31	62.83	**55.49**	51.04	**89.68**	**71.09**	**66.71**

Table 2. Ablation study of SDA in the fully supervised framework.

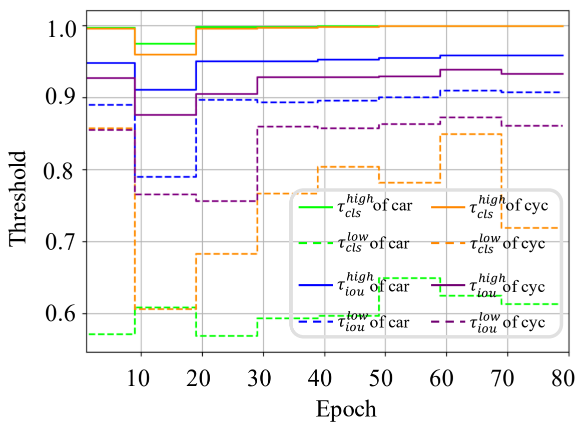

Figure 1. Visualization curve of the dynamic dual-threshold during training

[5] Shaoshuai Shi, Xiaogang Wang, and Hongsheng Li. Pointrcnn: 3d object proposal generation and detection from point cloud. In *CVPR*, pages 770–779, 2019. 1

[6] Kihyuk Sohn, Zizhao Zhang, Chun-Liang Li, Han Zhang, Chen-Yu Lee, and Tomas Pfister. A simple semi-supervised learning framework for object detection. *arXiv preprint arXiv:2005.04757*, 2020. 1

[7] Pei Sun, Henrik Kretzschmar, Xerxes Dotiwalla, Aurelien Chouard, Vijaysai Patnaik, Paul Tsui, James Guo, Yin Zhou, Yuning Chai, Benjamin Caine, et al. Scalability in perception for autonomous driving: Waymo open dataset. In *CVPR*, pages 2446–2454, 2020. 1

[8] Hongyi Zhang, Moustapha Cisse, Yann N Dauphin, and David Lopez-Paz. mixup: Beyond empirical risk minimization. *arXiv preprint arXiv:1710.09412*, 2017. 1

[9] Jinlai Zhang, Lyujie Chen, Bo Ouyang, Binbin Liu, Jihong Zhu, Yujin Chen, Yanmei Meng, and Danfeng Wu. Pointcutmix: Regularization strategy for point cloud classification. *Neurocomputing*, 505:58–67, 2022. 1