Supplementary Materials: VRDistill: Vote Refinement Distillation for Efficient Indoor 3D Object Detection

Anonymous Authors

1 APPENDIX

1.1 Discussion on the differences between indoor and outdoor detectors.

Our VRDistill focuses on raw point cloud-based object detection, which is more common for indoor object detection. Indoor object detection often involves areas such as augmented reality and robot control, which require detection algorithms to pay more attention to object categories and heights. Hence, indoor object detection algorithms are often based on raw point clouds, which can better reflect these characteristics. On the other hand, outdoor object detection is often related to autonomous driving, which demands real-time performance from algorithms. Therefore, input data is often transformed into a BEV perspective, and algorithms are typically based on pillars. Considering these differences, we evaluate our method based on indoor datasets. To show the generalization ability of VRDistill, we applied our method to the outdoor point cloud-based method PointRCNN [2] on the KITTI [1] dataset, and the results are shown in Table 1 below, which shows VRDistill is still effective to improve the results of student network. Note that the channel of the student is half of the channel of the teacher.

1.2 Visualization on the qualities of bounding boxes

As shown in Fig. 1, we visualize the qualities of the bounding boxes of different methods and observe that our method can largely improve the quality of generated bounding boxes, which further demonstrates the effectiveness of our method.

REFERENCES

[1] Andreas Geiger, Philip Lenz, Christoph Stiller, and Raquel Urtasun. 2013. Vision meets robotics: The kitti dataset. *The International Journal of Robotics Research* 32, 11 (2013), 1231–1237.
[2] S Shi, X Wang, H Pointrcnn Li, et al. 2019. 3d object proposal generation and detection from point cloud. In *CVPR*. 16–20.
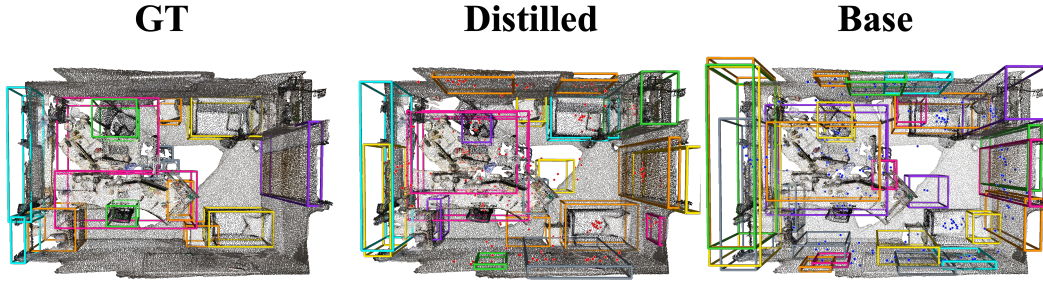
GT   **Distilled**   **Base**

Figure 1: From left to right, GT bounding boxes, bounding boxes from VoteNet trained by our VRDistill, and bounding boxes from VoteNet without distillation.

Setting	Car IoU=0.7			Pedestrian IoU=0.5			Cyclist IoU=0.5		
	Easy	Moderate	Hard	Easy	Moderate	Hard	Easy	Moderate	Hard
Teacher	86.7	75.8	73.2	63.7	59.2	52.1	86.1	69.0	64.7
Student(1/2)	78.6	68.6	67.4	55.1	50.0	44.9	85.9	65.6	62.3
VRDistill	85.2	72.8	69.7	60.4	57.4	49.1	85.3	68.0	63.2

Table 1: Results of VRDistill framework when using PointRCNN on the KITTI dataset.