

# Supplementary Materials: Joint Homophily and Heterophily Relational Knowledge Distillation for Efficient and Compact 3D Object Detection

Anonymous Author(s)

This supplementary material shows more visualization results to prove the effectiveness of joint Homophily and Heterophily Relational Knowledge Distillation (H2RKD) method for efficient and compact 3D Object Detection.

First of all, we compare the disparities in global and local feature relationships captured from both the student and teacher in Figure 1 and Figure 2 respectively. Specifically, in Figure 1, we calculate the mean square distance between feature maps of teacher PointPillars and student 4x compressed PointPillars. It's intuitive that a larger mean square value indicates a greater representation disparity, and the higher the frequency, the more values correspondingly fall within the interval. It can be seen from Figure 1(a) that RDD[3] most effectively imitates the teacher's feature map, but the small distance one needs to be further reduced. At the same time, PointDistiller[4] performs better at small distances, but the maximum distance is 3 times that of RDD in Figure 1(b). As shown in Figure 1(c). We optimize the transfer of similarity from teacher feature maps by exploring global second-order and third-order relationships. Regarding the transfer of local relationships, we visualized the local area voxel feature similarity matrix. In Figure 2 (a), we supplement the diagram in the introduction. It is evident that RDD failed to capture feature relationships effectively due to its neglect of relationship transfer. (b) primarily focuses on homophily, thus it does not perform well in low-similarity regions. In contrast, our method emphasizes simultaneously distilling both relations to effectively imitate the teacher by separate local distillation. In summary, our method effectively replicates the teacher's characteristics while also capturing both homophily and heterophily between the features.

Then, we provided a visualization of significant points detected on the KITTI[2] and nuScene[1] datasets. The CenterPoint is 2x compressed, and the PointPillars is 4x compressed. As depicted in Figure 3, the saliency of foreground objects (such as cars and pedestrians) is notably higher, while the background points are suppressed. This observation suggests that our method effectively localizes important points.

Finally, we give some detection results with 2x compressed CenterPoint on nuScene datasets. As depicted in Figure 4, superior detection results are attained in both night and day scenarios. Additionally, in Figure 5, we can observe that the model without distillation exhibits poorer detection performance for pedestrians at a distance and vehicles at corners.

Overall, our method demonstrates outstanding performance on KITTI and nuScene datasets, effectively identifying distant pedestrians and vehicles even in highly sparse point clouds.

## REFERENCES

- [1] Holger Caesar, Varun Bankiti, Alex H Lang, Sourabh Vora, Venice Erin Liong, Qiang Xu, Anush Krishnan, Yu Pan, Giancarlo Baldan, and Oscar Beijbom. 2020.

- nuscenes: A multimodal dataset for autonomous driving. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 11621–11631.
- [2] 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.
- [3] Yanjing Li, Sheng Xu, Mingbao Lin, Jihao Yin, Baochang Zhang, and Xianbin Cao. 2023. Representation Disparity-aware Distillation for 3D Object Detection. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 6715–6724.
- [4] Linfeng Zhang, Runpei Dong, Hung-Shuo Tai, and Kaisheng Ma. 2023. Pointdistiller: Structured knowledge distillation towards efficient and compact 3d detection. In *Proceedings of the IEEE/CVF conference on Computer Vision and Pattern Recognition*. 21791–21801.

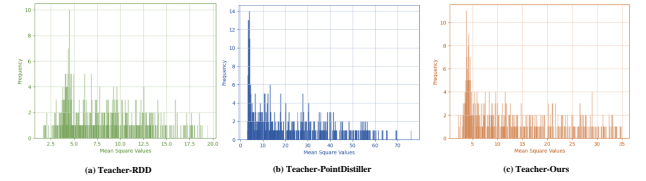


Figure 1: Histogram of the mean square distance between feature maps of the teacher and the student distilled by RDD, PointDistiller, and Ours on KITTI.

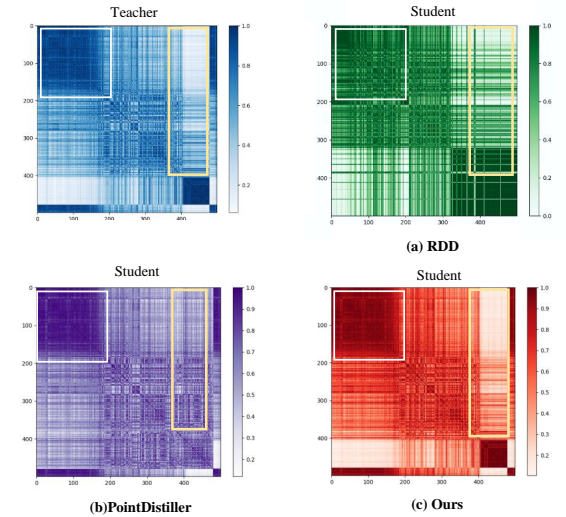
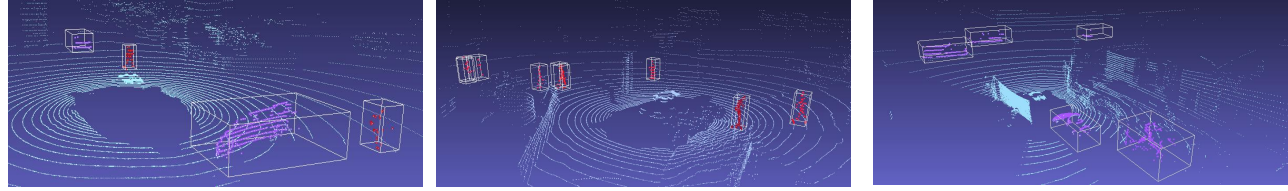
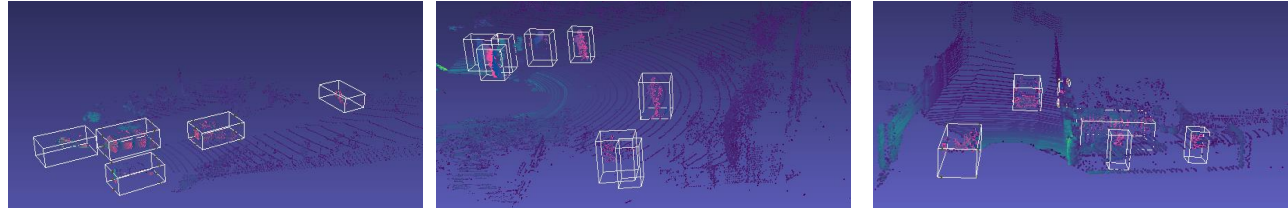


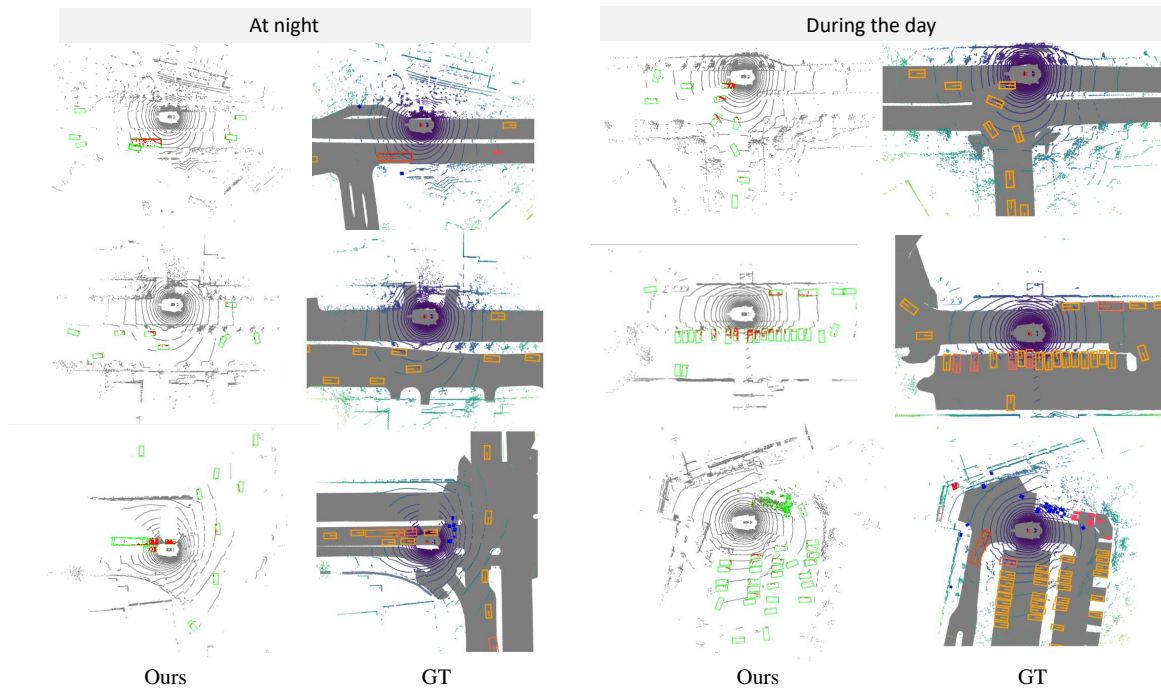
Figure 2: Visualization of the relations among voxel features from the teacher and the student distilled by (a), (b), and (c) on KITTI, respectively. The relations are demonstrated by the similarity matrix, where the white box represents high similarity, and the yellow box represents weak similarity.

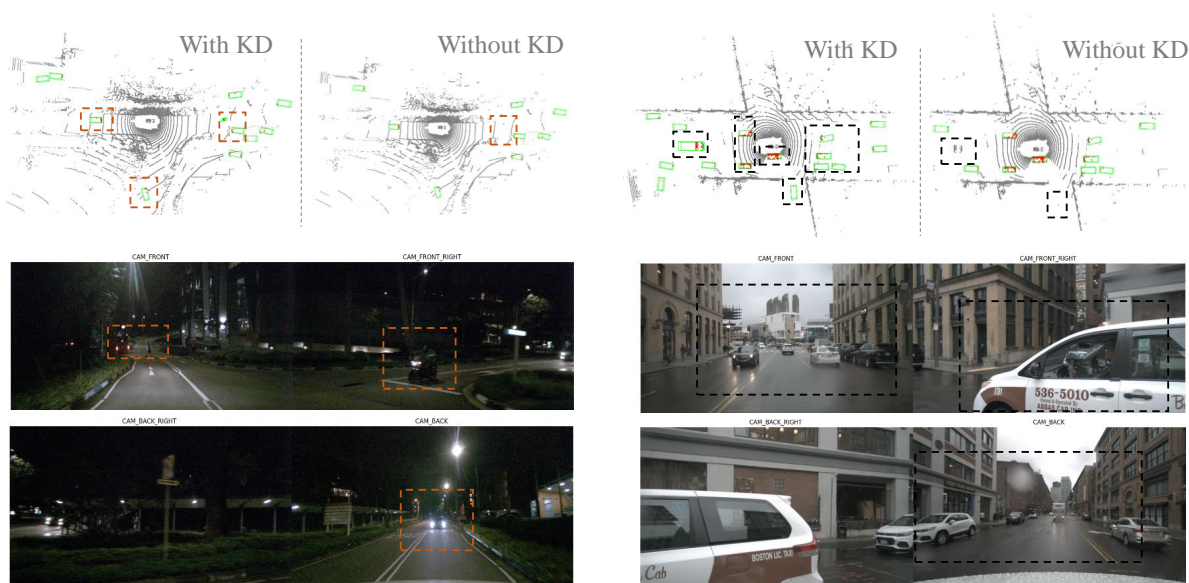


(a) Feature distribution with CenterPoint on the nuScenes dataset



(b) Feature distribution with PointPillars on the KITTI dataset

**Figure 3: Visualization of Important Points in Detectors on KITTI and nuScense datasets.****Figure 4: Visualization of the detection results in day and night with CenterPoint on the nuScense Validation set.**



**Figure 5: Comparison between the detection results of students trained with and without KD on the nuScense test set. The top are detection results, and the bottom are the front and back views of images corresponding to the scene.**