Doppler-Aware LiDAR-RADAR Fusion for Weather-Robust 3D Detection

Supplementary Material

1. Implementation Details

Following the official paper of K-RADAR dataset, point clouds within the driving corridor were used for the experiment, with the x-axis range (0, 72m), y-axis range (-6.4, 6.4m), and z-axis range (-2, 6m). The voxel size was set to 0.4 × 0.4 × 0.4m, and the preprocessing bin size with Δd was set to 0.2. The number of MPII and stages was set to b=3. Training was conducted with a batch size of 4 using the Adam optimizer with lr=1e-3, β_1=0.9, β_2=0.999 until performance convergence. A single A6000 GPU is utilized for training and evaluation. The total training time is around 1$\sim$1.5 days with full K-RADAR trainset. The current inference speed is around 20ms per sample, which is below real-time. However, this is primarily due to the unoptimized NumPy-based neighbor search. Replacing it with a PyTorch implementation or using TensorRT is expected to significantly improve the speed, potentially beyond real-time.

2. Additional Quantitative Comparison

We report quantitative comparison with additional baselines [2], student model of [1], and [4]. Those were not included in the main comparison for the following reasons: [1] fuses camera data in addition to LiDAR and radar, and [2] uses a LiDAR–4D radar fusion model as the teacher, but explores a different KD strategy for the student. We believe DLRFusion could also benefit from such a KD, and thus did not consider it a direct comparison. [3] is a recent work utilizing the denoising strategy. Since we believe that a broader comparison is valuable, therefore the results were summarized in the extended Table 1. As shown, our DLRFusion shows comparable results to the models that use camera input or KD, and even surpasses both the teacher and student models of [1] under IoU=0.5. While [4] benefits from denoising strategies, our method also shows superior performance in several aspects such as AP_{BEV} at IoU=0.3.

Table 1. Quantitative comparison with more baselines.

	IoU=0.3		IoU=0.5	
	AP_{BEV}	AP_{3D}	AP_{BEV}	AP_{3D}
3D-LRF [2]	84.0	74.8	73.6	45.2
Teacher of [1]	82.1	73.2	71.3	40.4
Student of [1]	83.3	75.6	72.4	43.3
L4DR [4]	79.5	78.0	77.5	53.5
DLRFusion	82.9	74.8	73.2	45.7

3. Further Qualitative Analysis

We provide more qualitative detection results on K-RADAR dataset under various weather conditions. The qualitative detection results of RTNH [7], RTNH† [7], PointPillars [5], VoxelNext [3], InterFusion [8], BEVFusion† [6], and DLRFusion (ours) are reported. Fig. 1 shows overall qualitative results extended from the Fig. 3 in the main paper. Fig. 2 $\sim$ 8 shows results under normal, overcast, fog, rain, sleet, light snow and heavy snow, respectively. As shown in the figures, our DLRFusion achieves the most accurate classification and precise regression across various weather conditions. This demonstrates the effective fusion of LiDAR and 4D RADAR's power and Doppler. Specifically, LiDAR-based methods experience misdetections and incorrect detections due to adverse weather noise and occlusion. In environments such as fog, rain, sleet, and snow, LiDAR points significantly decrease, and in rainy conditions, the LiDAR point distribution becomes scattered. In contrast, RADAR-based methods struggle with accurately predicting bounding boxes due to RADAR's weaker structural boundary information. In some cases, RADAR can confuse structures with objects, especially when power information alone is present in areas where no objects should exist. Other RADAR-LiDAR fusion methods often face difficulties when one modality is unclear, making it challenging to determine which modality's results to trust. In contrast, our method can discern non-object areas even when confusion arises between modalities. Through multi-path interaction, it recognizes when dynamic information is insufficient, even when power is strong. As demonstrated in the sample in Fig. 7 on the right, our method shows robust performance, accurately aligning the direction of vehicles, even when their trajectories are not perfectly linear.

References

[1] Yujeong Chae, Hyeonseong Kim, Changgyoon Oh, Minseok Kim, and Kuk-Jin Yoon. Lidar-based all-weather 3d object detection via prompting and distilling 4d radar. In *European Conference on Computer Vision*, pages 368–385. Springer, 2024. 1

[2] Yujeong Chae, Hyeonseong Kim, and Kuk-Jin Yoon. Towards robust 3d object detection with lidar and 4d radar fusion in various weather conditions. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 15162–15172, 2024. 1

[3] Yukang Chen, Jianhui Liu, Xiangyu Zhang, Xiaojuan Qi, and Jiaya Jia. Voxelnext: Fully sparse voxelnet for 3d object detection and tracking. In *Proceedings of the IEEE/CVF Con-*

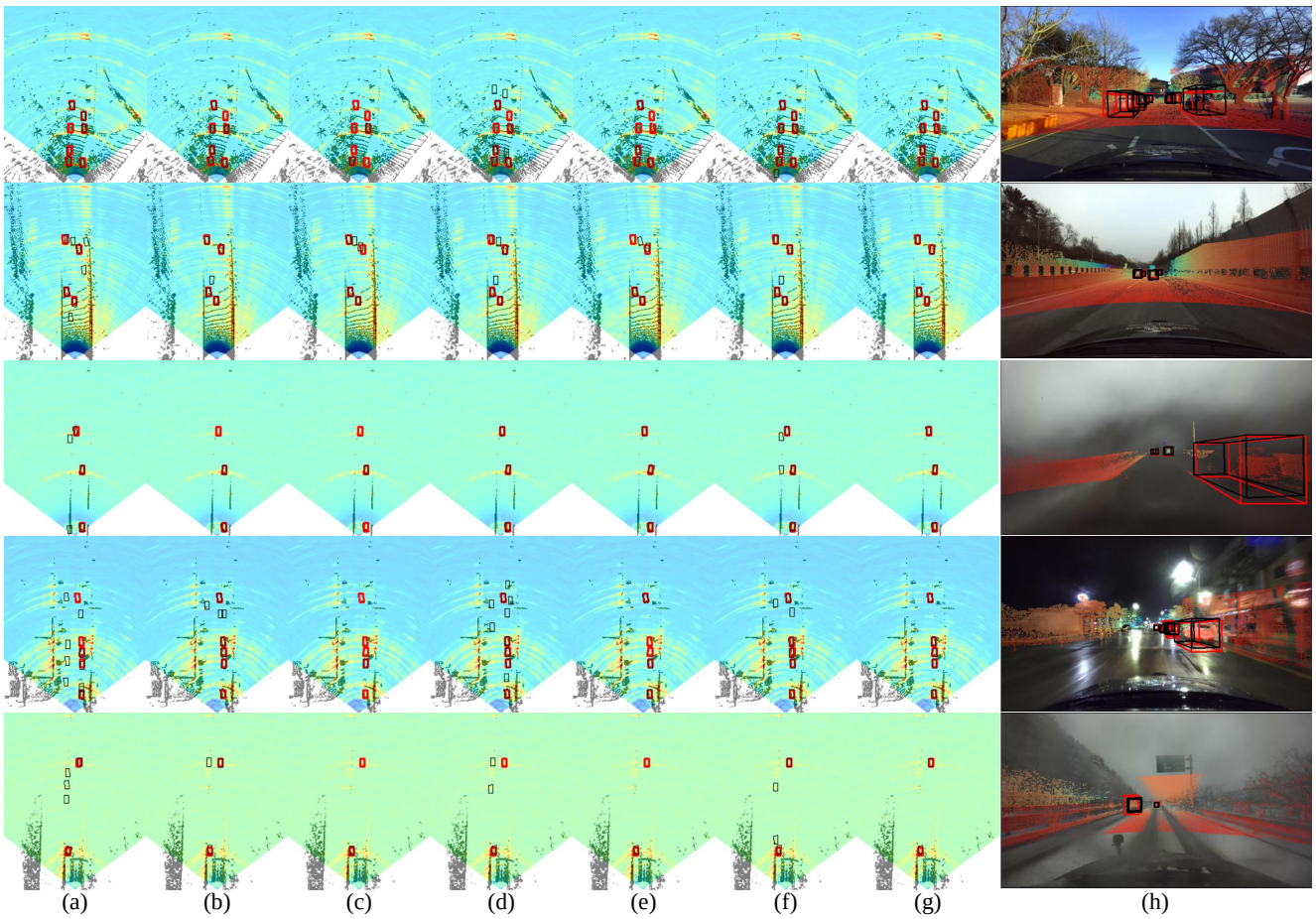

Figure 1. Visual comparison of (a) RTNH, (b) RTNH†, (c) PointPillars, (d) VoxelNext, (e) InterFusion, (f) BEVFusion†, and (g) DLRFusion (ours) on K-RADAR dataset. Each column shows a sample from normal, overcast, fog, rain, and heavy snow conditions. (a) $\sim$ (g) depict LiDAR and RADAR in BEV, while (h) shows our 3D results in range view with camera and LiDAR. The red represents the ground truth (GT), while the black represents the prediction.

ference on Computer Vision and Pattern Recognition, pages 21674–21683, 2023. 1, 3, 4, 5, 6, 7, 8, 9

[4] Xun Huang, Ziyu Xu, Hai Wu, Jinlong Wang, Qiming Xia, Yan Xia, Jonathan Li, Kyle Gao, Chenglu Wen, and Cheng Wang. L4dr: Lidar-4dradar fusion for weather-robust 3d object detection, 2024. 1

[5] Alex H Lang, Sourabh Vora, Holger Caesar, Lubing Zhou, Jiong Yang, and Oscar Beijbom. Pointpillars: Fast encoders for object detection from point clouds. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 12697–12705, 2019. 1, 3, 4, 5, 6, 7, 8, 9

[6] Zhijian Liu, Haotian Tang, Alexander Amini, Xingyu Yang, Huizi Mao, Daniela Rus, and Song Han. Bevfusion: Multi-task multi-sensor fusion with unified bird's-eye view representation. In *IEEE International Conference on Robotics and Automation (ICRA)*, 2023. 1, 3, 4, 5, 6, 7, 8, 9

[7] Dong-Hee Paek, Seung-Hyun Kong, and Kevin Tirta Wijaya. K-radar: 4d radar object detection for autonomous driving in various weather conditions. *Advances in Neural Information Processing Systems*, 35:3819–3829, 2022. 1, 3, 4, 5, 6, 7, 8, 9

[8] Li Wang, Xinyu Zhang, Baowei Xv, Jinzhao Zhang, Rong Fu, Xiaoyu Wang, Lei Zhu, Haibing Ren, Pingping Lu, Jun Li, and Huaping Liu. Interfusion: Interaction-based 4d radar and lidar fusion for 3d object detection. In *2022 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 12247–12253, 2022. 1, 3, 4, 5, 6, 7, 8, 9

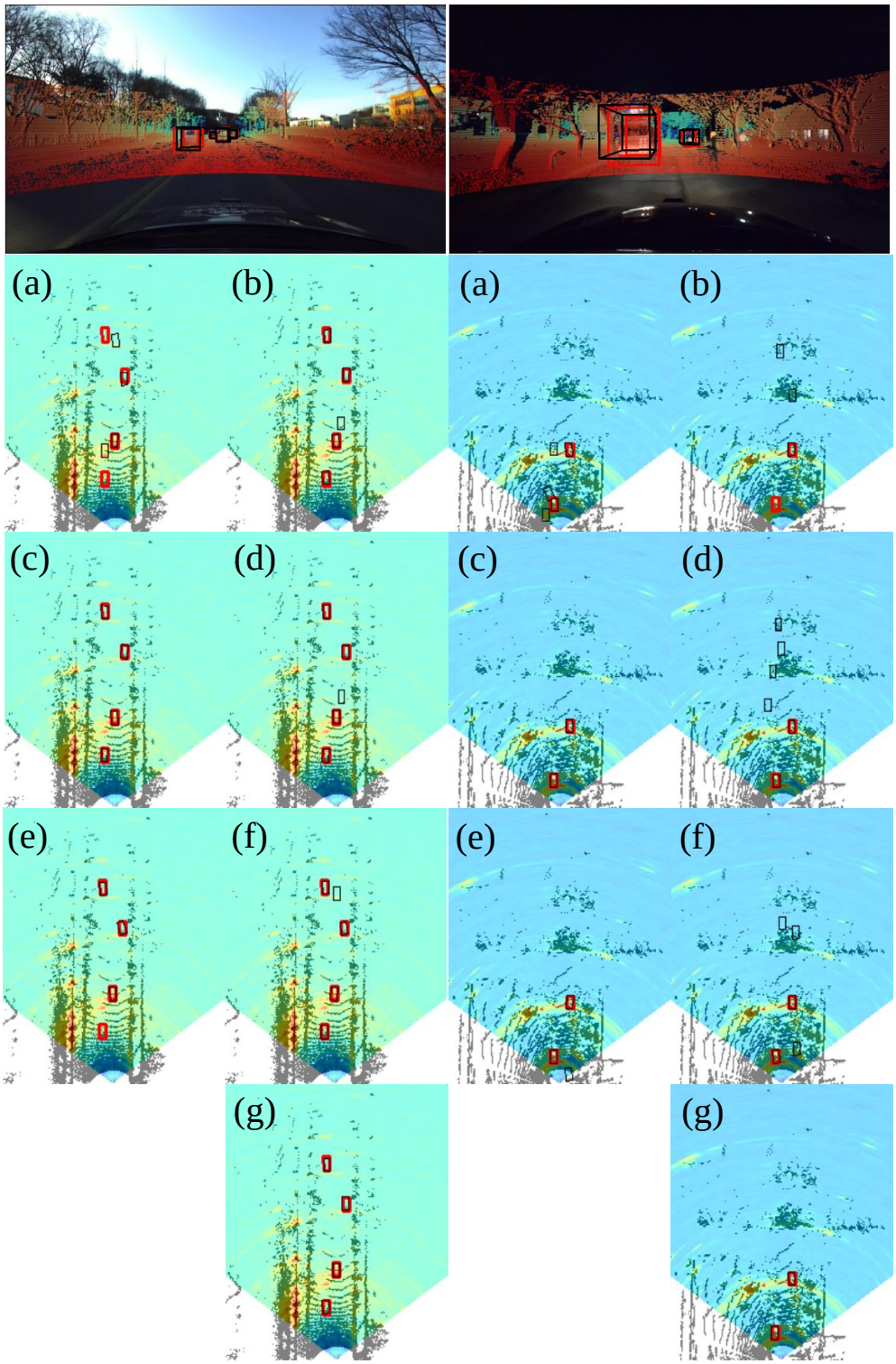

Figure 2. Additional results under *normal* conditions of (a) RTNH [7], (b) RTNH†[7], (c) PointPillars [5], (d) VoxelNext [3], (e) InterFusion [8], (f) BEVFusion† [6], and (g) DLRFusion (ours) with LiDAR and 4D RADAR in Bird's Eye View. The top image depicts the detection results of DLRFusion in 3D range view, with camera and LiDAR projection. The red represents the ground truth (GT), while the black represents the prediction.

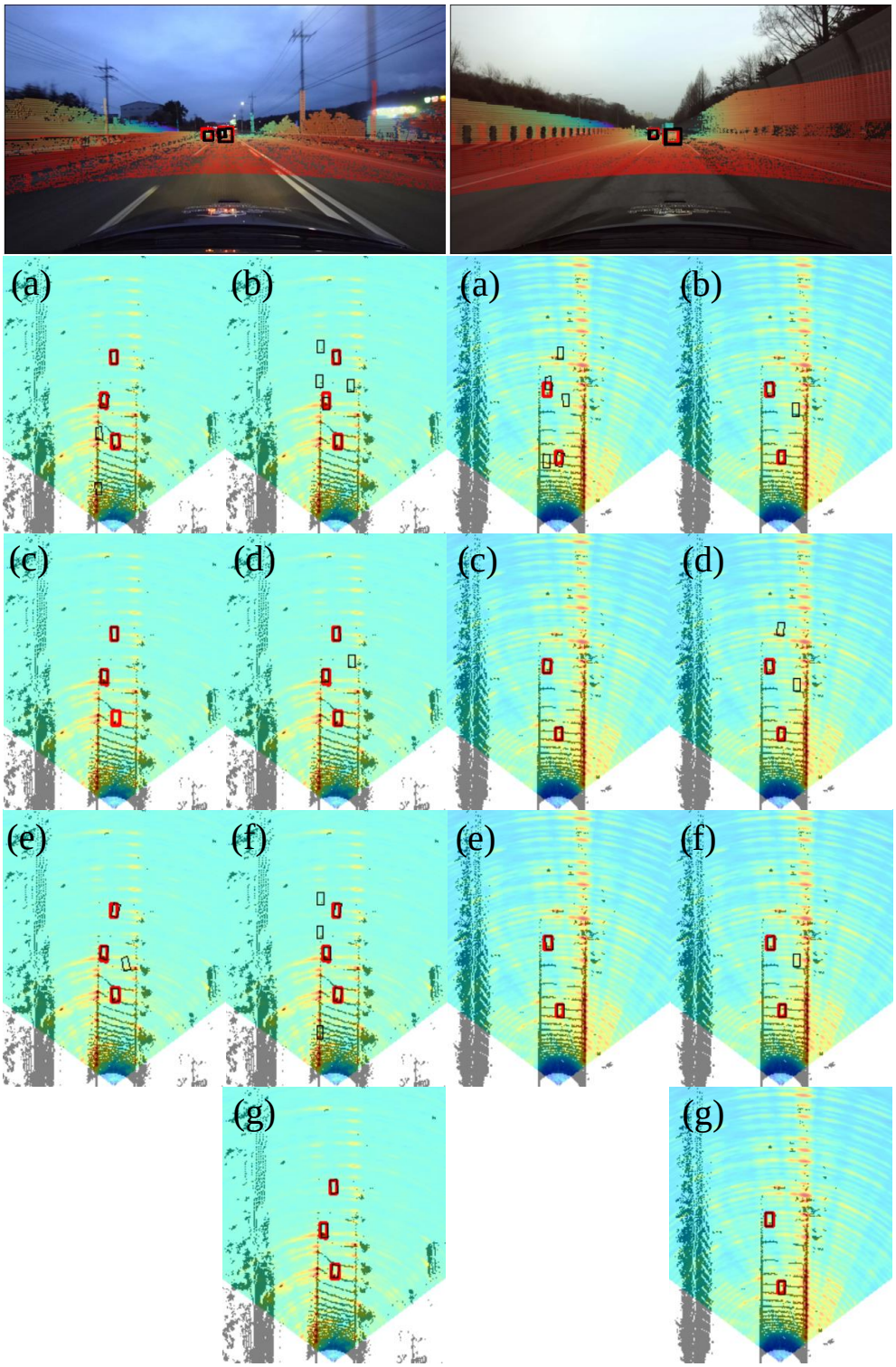

Figure 3. Additional results under *overcast* conditions of (a) RTNH [7], (b) RTNH†[7], (c) PointPillars [5], (d) VoxelNext [3], (e) InterFusion [8], (f) BEVFusion† [6], and (g) DLRFusion (ours) with LiDAR and 4D RADAR in Bird's Eye View. The top image depicts the detection results of DLRFusion in 3D range view, with camera and LiDAR projection. The red represents the ground truth (GT), while the black represents the prediction.

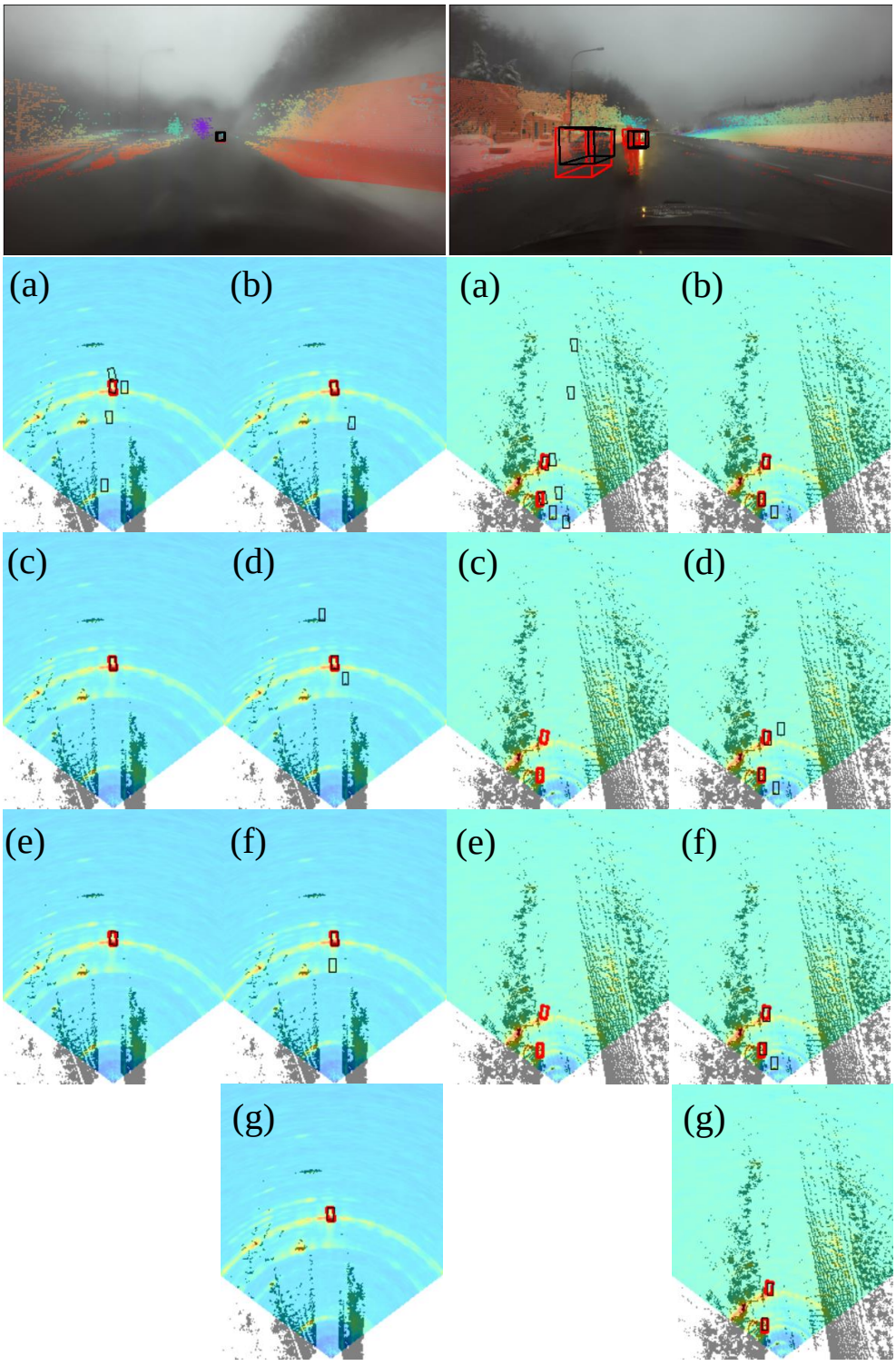

Figure 4. Additional results under *fog* conditions of (a) RTNH [7], (b) RTNH†[7], (c) PointPillars [5], (d) VoxelNext [3], (e) InterFusion [8], (f) BEVFusion† [6], and (g) DLRFusion (ours) with LiDAR and 4D RADAR in Bird's Eye View. The top image depicts the detection results of DLRFusion in 3D range view, with camera and LiDAR projection. The red represents the ground truth (GT), while the black represents the prediction.

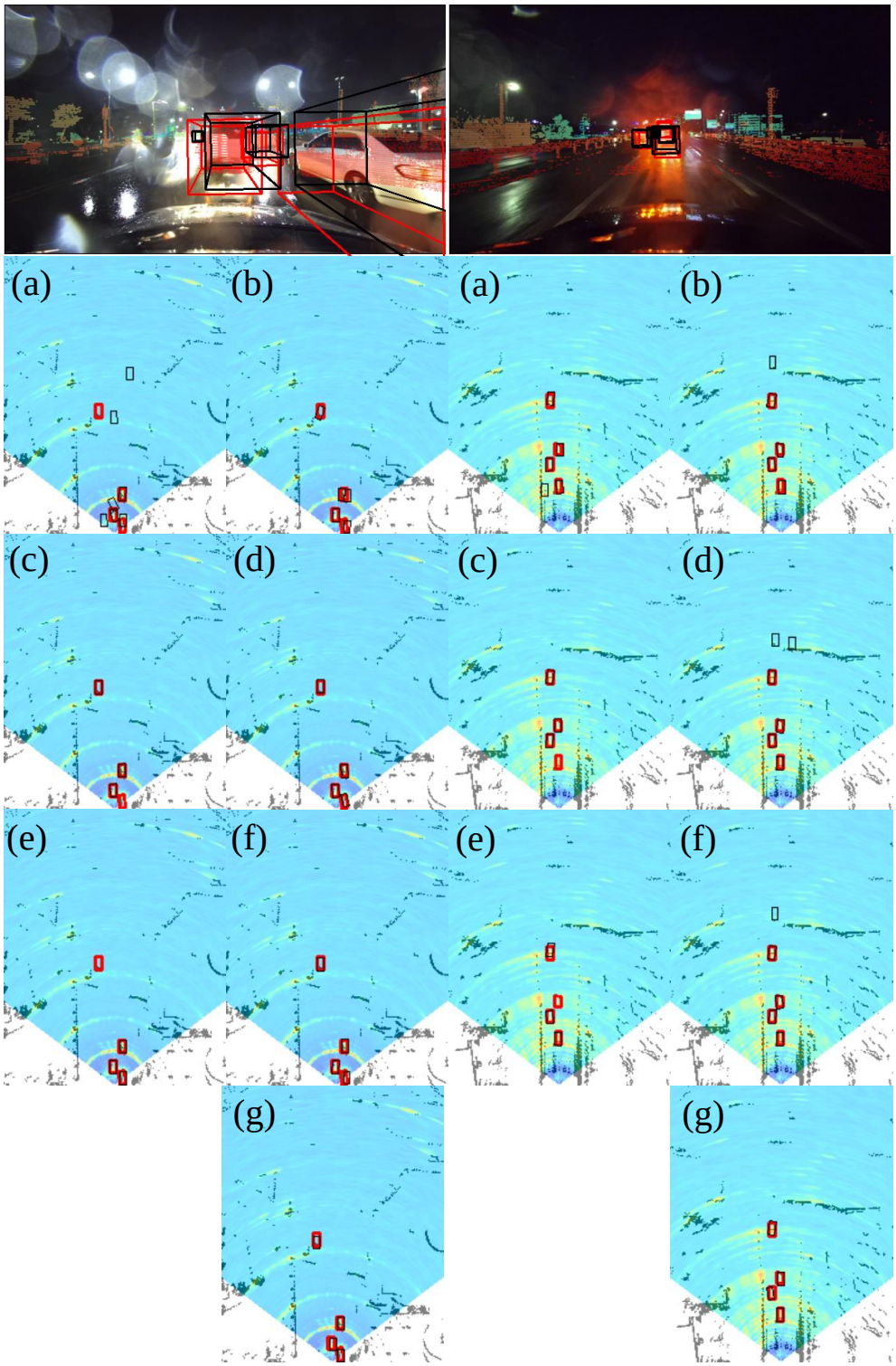

Figure 5. Additional results under *rain* conditions of (a) RTNH [7], (b) RTNH†[7], (c) PointPillars [5], (d) VoxelNext [3], (e) InterFusion [8], (f) BEVFusion† [6], and (g) DLRFusion (ours) with LiDAR and 4D RADAR in Bird's Eye View. The top image depicts the detection results of DLRFusion in 3D range view, with camera and LiDAR projection. The red represents the ground truth (GT), while the black represents the prediction.

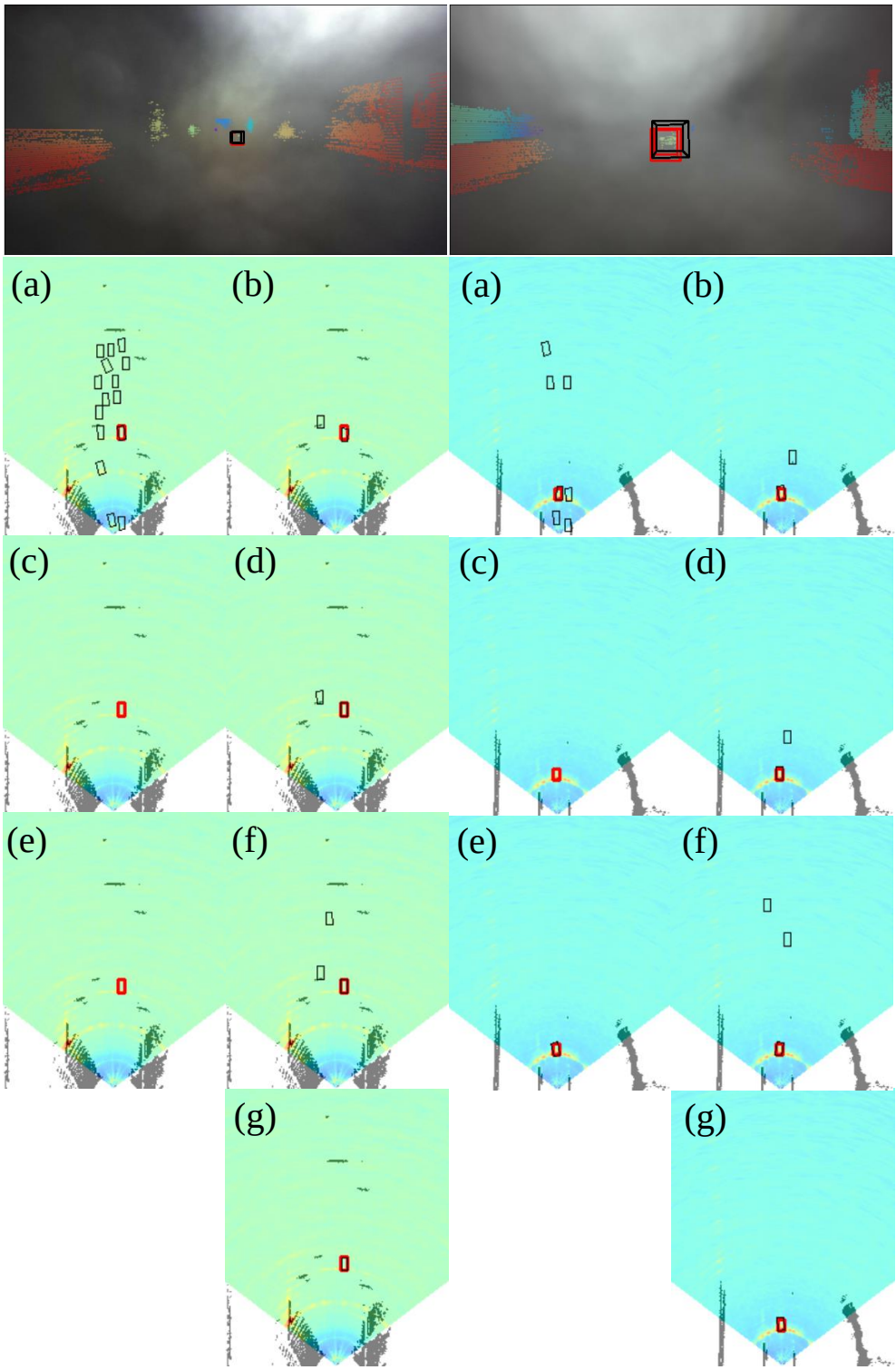

Figure 6. Additional results under *sleet* conditions of (a) RTNH [7], (b) RTNH†[7], (c) PointPillars [5], (d) VoxelNext [3], (e) InterFusion [8], (f) BEVFusion† [6], and (g) DLRFusion (ours) with LiDAR and 4D RADAR in Bird's Eye View. The top image depicts the detection results of DLRFusion in 3D range view, with camera and LiDAR projection. The red represents the ground truth (GT), while the black represents the prediction.

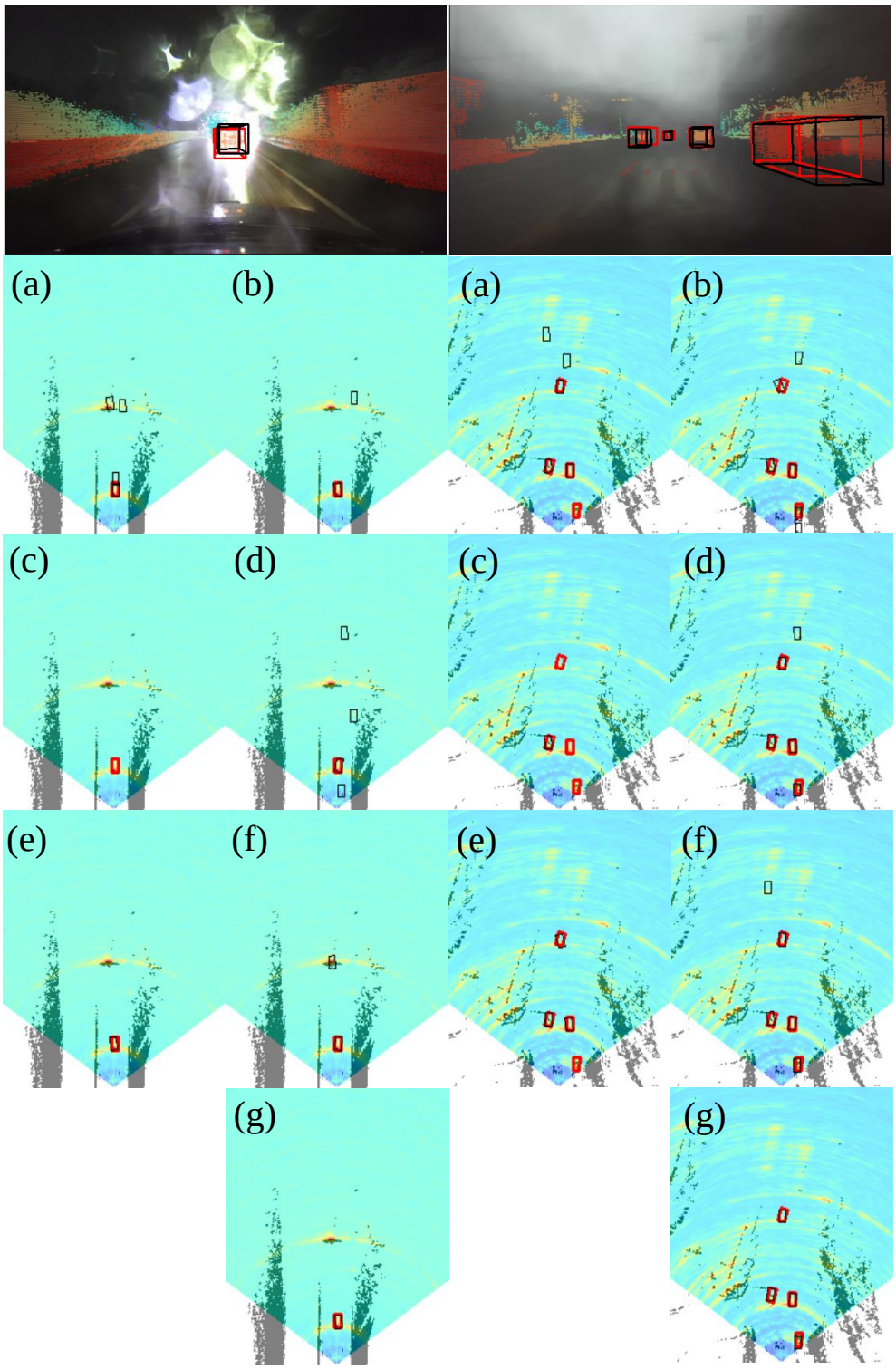

Figure 7. Additional results under *light snow* conditions of (a) RTNH [7], (b) RTNH† [7], (c) PointPillars [5], (d) VoxelNext [3], (e) InterFusion [8], (f) BEVFusion† [6], and (g) DLRFusion (ours) with LiDAR and 4D RADAR in Bird's Eye View. The top image depicts the 3D detection results of DLRFusion in range view, with camera and LiDAR projection. The red represents the ground truth (GT), while the black represents the prediction.

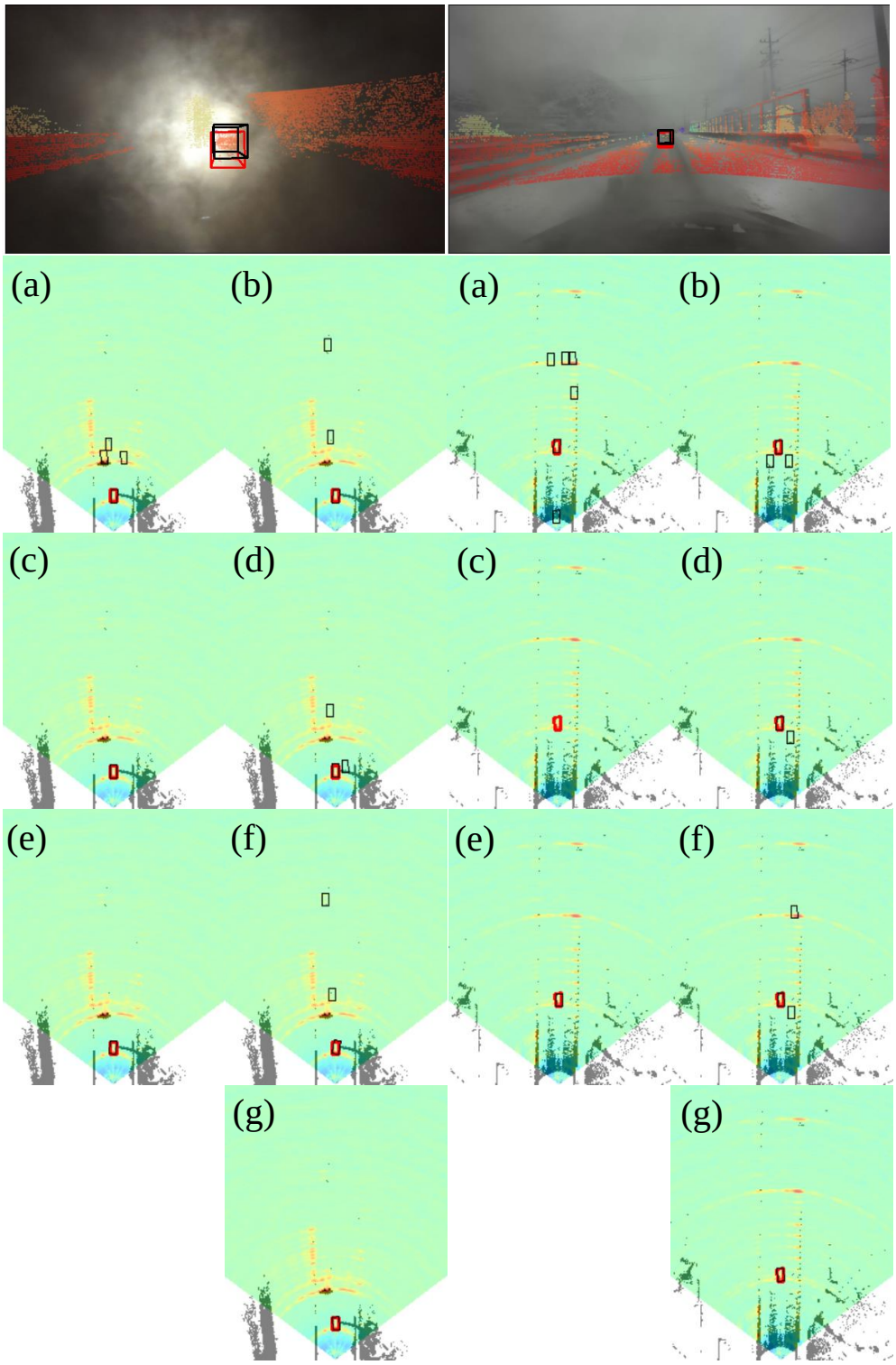

Figure 8. Additional results under *heavy snow* conditions of (a) RTNH [7], (b) RTNH†[7], (c) PointPillars [5], (d) VoxelNext [3], (e) InterFusion [8], (f) BEVFusion† [6], and (g) DLRFusion (ours) with LiDAR and 4D RADAR in Bird's Eye View. The top image depicts the detection results of DLRFusion in 3D range view, with camera and LiDAR projection. The red represents the ground truth (GT), while the black represents the prediction.