Bi-LRFusion: Bi-Directional LiDAR-Radar Fusion for 3D Dynamic Object Detection Supplementary Material

Yingjie Wang[1] Jiajun Deng[2*] Yao Li[1] Jinshui Hu[3] Cong Liu[3] Yu Zhang[1] Jianmin Ji[1]
Wanli Ouyang[4] Yanyong Zhang[1*]
[1]University of Science and Technology of China [2]University of Sydney [3]iFLYTEK [4]Shanghai AI Laboratory

A. Overview

The supplementary document is organized as follows:

- Section B presents the detailed model settings and training details.
- Section C presents additional qualitative results of two typical locations of Radar points on objects with different heights.
- Section D depicts the detailed effect of query-based height feature fusion (QHF) block (Section 4.4 of the main paper).
- Section E depicts the detailed effect of query-based BEV feature fusion (QBF) block (Section 4.4 of the main paper).
- Section F depicts the effect on objects at different ranges.
- Section H depicts the additional comparison with existing methods.

B. Model Settings and Training Details

Our implementation is based on the open-sourced code mmdetection3d [4]. We choose CenterPoint [7] and PointPillars [5] to serve as the baseline for the LiDAR-only detection on nuScenes and ORR datasets, respectively. We follow the network of PointPillars for the Radar feature stream. The radius of each ball query is $0.15m$. The Manhattan distance threshold is set as $(2,2)$ for BEV query. The final output channels of the PointNet module and the MLP layer are 32.

Our implementation is based on the open-sourced code mmdetection3d. We choose CenterPoint and PointPillars to serve as the baseline for the LiDAR-only detection on nuScenes and ORR datasets, respectively. We follow the network of PointPillars for the Radar feature stream. The radius of each ball query is $0.15m$. The Manhattan distance threshold is set as $(2,2)$ for BEV query. The final output channels of the PointNet module and the MLP layer are 32.

C. Qualitative Results of Radar Point Locations on Objects with Different Heights

As mentioned in the main paper, the currently available Radar is unable to capture the height information. The height value of a Radar point is assigned as the height of the ego Radar sensor, which is deployed on the top of the autonomous vehicle. As a result, Radar points are located at different part for objects with different heights. We illustrate the location of Radar points on object with different heights in Figure 1. This treatment leads to two typical types of Radar point locations: (1) For cars and motorcycles whose actual heights are similar to the data-collecting car, Radar points often fall on the top of their bounding boxes, and (2) for much taller objects such as trucks and buses, Radar points instead fall inside of their bounding boxes. As discussed in the Figure 1 from the main paper, the missing height information of Radar leads to unstable improvements for objects with different height.

D. Effect of QHF Block

We further report the mean AP of two dynamic object groups on nuScenes validation set to illustrate how QHF block affects the detection performance. As proposed in the main paper, we group the dynamic objects to (1) similar-height objects and (2) tall objects according to the Radar sensor's height. From Table 1, *Method (a)* is our LiDAR-only baseline CenterPoint [7]. *Method (b)* extends *(a)* by simply fusing the Radar feature via R2L fusion, which achieves 1.9% gain of AP for group (1). We notice that tall objects in group (2) do not enjoy as much performance gain. As such, *Method (c)* enriches Radar data by learning the corresponding height features from raw LiDAR points via QHF block, leading to steady AP improvement for both two

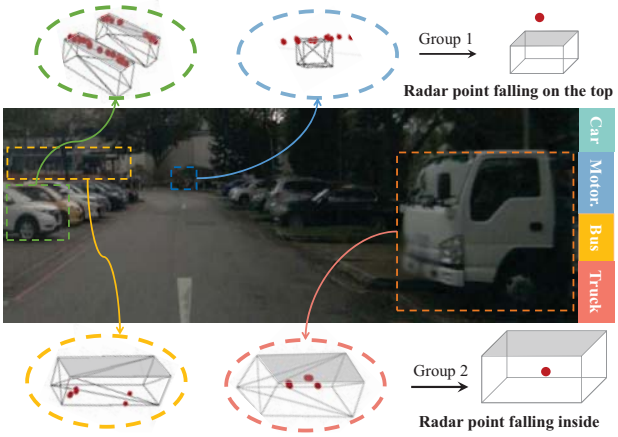

Figure 1. Due to the lack of height information, Radar points may fall at different locations on an object's bounding box depending upon the actual height of the object. The middle part of the figure is an 2D image from nuScenes dataset [2]; in dotted circles, we show the 3D ground-truth bounding boxes and Radar points using MeshLab [3] tool. Best viewed in color.

groups (2.3%/1.3%). It demonstrates that pseudo height information from QHF plays a more important role in improving the detection performance of large objects.

E. Effect of QBF Block

We also present the detailed performance gain of QBF block in Table 1. *Method (d)* extends *(b)* by learning from detailed LiDAR features on the BEV plane to enhance Radar features, which achieves +2.3% improvements in terms of AP for Group 1 and 0.4% for Group 2. The designed QBF block constructs a more fine-grained Radar BEV feature for more effective detection, benefiting from much-detailed clues, *e.g.* objects at long range or objects with small sizes. In addition, we can observe that *Method (d)* only brings small performance gains on tall objects. We speculate that this is mainly due to the height missing problem still exists when using the QBF block alone.

F. Effect of Objects at Different Ranges

Apart from object parameters like heights and velocities, we also evaluate the improvements of Bi-LRFusion for objects with different distances compared with the LiDAR-centric baseline CenterPoint. Table 2 shows the mAP and mAVE of the LiDAR-only detector and Bi-LRFusion on the representative car class at different ranges. At the range closer than 20 m, CenterPoint performs better due to the precise and dense point cloud from LiDAR data. At the range farther than 20 m, the performance of our Bi-LRFusion exceeds the LiDAR-only method with the advantage of a long detection range from Radar data. The result demonstrates that our LiDAR-Radar fusion framework is less sensitive to the distance because the LiDAR points become sparse as the distance increases, while the Radar points are uniform compared to the LiDAR points. In conclusion, Radar data are of great help to locating and detecting targets especially at long distances.

Table 1. The detailed effect of each proposed component in Bi-LRFusion. We report the mean AP on two typical groups. Group (1) includes similar-height objects like cars, motorcycles, bicycles and pedestrians, while group (2) includes tall objects like trucks, buses and trailers. We also show the performance gain of different components compared with LiDAR-only method.

Method	R2L	L2R		Mean AP ↑ (%)		Gain (%)	
		QHF	QBF	Group 1	Group 2	Group 1	Group 2
(a)				64.5	52.6	-	-
(b)	✓			66.4	52.6	+1.9	+0.0
(c)	✓	✓		66.8	53.9	+2.3	+1.3
(d)	✓		✓	68.6	53.0	+4.1	+0.4

Table 2. mAP and mAVE results for LiDAR-only CenterPoint [7] and Bi-LRFusion for cars at different ranges.

Method	Range (m)	*AP*(%) on Car		
		0.5m	1.0m	2.0m
LiDAR-only	0-20	93.3	96.5	96.5
	20-40	69.8	82.2	85.5
	>40	26.1	40.1	45.8
Bi-LRFusion	0-20	92.6	95.5	**96.6** (+0.1)
	20-40	**71.0** (+1.2)	**83.5** (+1.3)	**86.7** (+1.2)
	>40	**27.9** (+1.8)	**42.6** (+2.5)	**48.0** (+2.2)

G. Effect of Sensor Misalignment levels

We first define misalignment levels followed BEVFormer [6]. For the i-th level, a translation noise sampled from a normal distribution $\sigma \sim (0, 2.5i)$ is introduced to each direction. Besides, a horizontal rotation noise sampled from a normal distribution $\sigma \sim (0, i)$ is also introduced. From Table 3, Bi-LRFusion is able to tolerate misalignment levels 1-3, thanks to the query-based mechanism that can effectively query features within a certain area (instead of a single point or a small number of points). However, when the misalignment level increases beyond the query range, the performance starts to drop significantly (level 4). We will elaborate on the misalignment tolerance issue in the revision.

Table 3. The mAP (%) under different misalignment levels.

Level	0	1	2	3	4
mAP (%)	62.0	61.8	61.5	61.4	60.8

H. Comparison with existing methods

Bi-LRFusion can be applied to different LiDAR-only detectors, not limited to CenterPoint. Therefore, we choose TransFusion-L [1], which is also a prevalent baseline in the community, to conduct further experiments. As shown in Table 4, our Bi-LRFusion improves TransFusion-L considerably, demonstrating its generalizability.

Table 4. Experimental Results with Transfusion-L.

Methods	Modality	mAVE ↓	mAP ↑
Transfusion-L	L	24.8	65.3
+ Bi-LRFusion	L+R	**23.7**	**67.5**

References

[1] Xuyang Bai, Zeyu Hu, Xinge Zhu, Qingqiu Huang, Yilun Chen, Hongbo Fu, and Chiew-Lan Tai. Transfusion: Robust lidar-camera fusion for 3d object detection with transformers. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 1090–1099, 2022. 3

[2] Holger Caesar, Varun Bankiti, Alex H Lang, Sourabh Vora, Venice Erin Liong, Qiang Xu, Anush Krishnan, Yu Pan, Giancarlo Baldan, and Oscar Beijbom. nuscenes: A multimodal dataset for autonomous driving. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 11621–11631, 2020. 2

[3] Paolo Cignoni, Marco Callieri, Massimiliano Corsini, Matteo Dellepiane, Fabio Ganovelli, Guido Ranzuglia, et al. Meshlab: an open-source mesh processing tool. In *Eurographics Italian chapter conference*, volume 2008, pages 129–136. Salerno, Italy, 2008. 2

[4] MMDetection3D Contributors. MMDetection3D: OpenMMLab next-generation platform for general 3D object detection. https://github.com/open-mmlab/mmdetection3d, 2020. 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

[6] Zhiqi Li, Wenhai Wang, Hongyang Li, Enze Xie, Chonghao Sima, Tong Lu, Qiao Yu, and Jifeng Dai. Bevformer: Learning bird's-eye-view representation from multi-camera images via spatiotemporal transformers. *arXiv preprint arXiv:2203.17270*, 2022. 2

[7] Tianwei Yin, Xingyi Zhou, and Philipp Krahenbuhl. Center-based 3d object detection and tracking. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 11784–11793, 2021. 1, 2