R4Det: 4D Radar-Camera Fusion for High-Performance 3D Object Detection

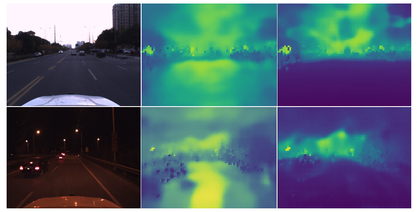
Zhongyu Xia[1] Yousen Tang[1] Yongtao Wang[1]✉ Zhifeng Wang[2] Weijun Qin[2]
[1] Wangxuan Institute of Computer Technology, Peking University [2] EBTech Co. Ltd

CVPR JUNE 3-7, 2026 DENVER COLORADO

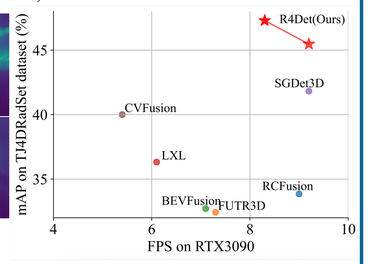
Introduction

Motivation

- Challenge 1: Inaccurate panoramic absolute depth estimation.
- Challenge 2: Ego-pose deficiency limits temporal fusion.
- Challenge 3: Small objects lack radar points, hard to detect.



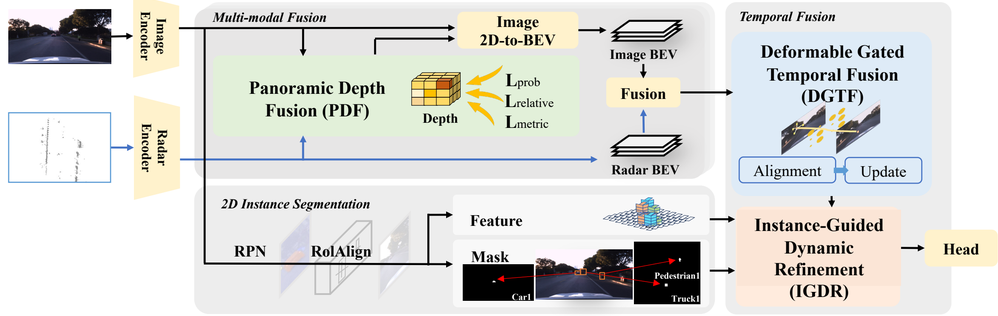
Input SGDet3D Ours

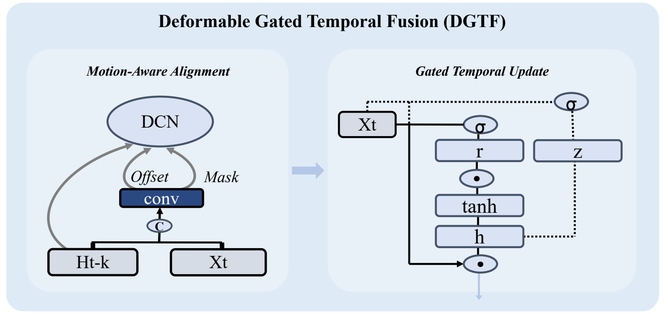


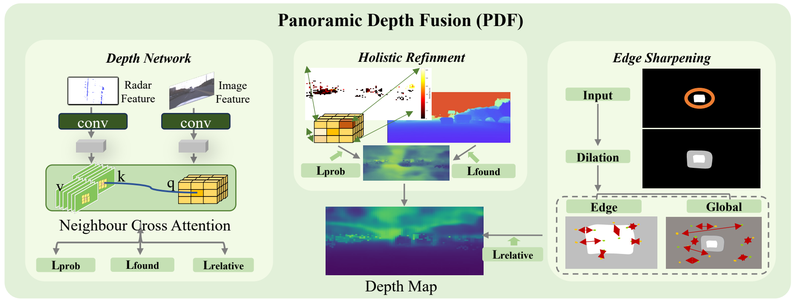
Contribution

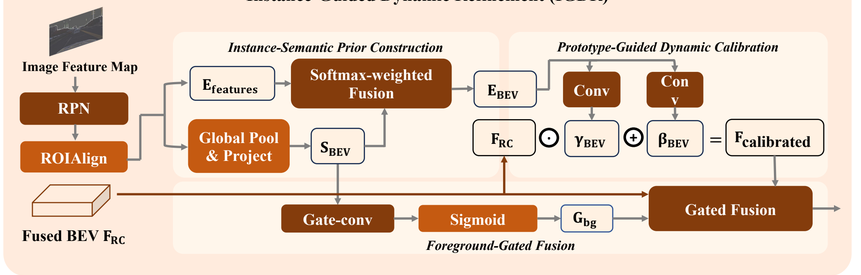
○ We propose R4Det, a novel 4D radar-camera fusion framework for highly accurate, efficient, and robust 3D object detection.
○ A Panoramic Depth Fusion (PDF) module is proposed to learn high-fidelity, edge-sharp, and structurally continuous geometry.
○ A Deformable Gated Temporal Fusion (DGTF) module is designed to achieve stable temporal feature aggregation without requiring ego-pose.
○ An Instance-Guided Dynamic Refinement (IGDR) module is introduced to purify features and restore clarity for small, distant, or occluded objects.
○ R4Det achieves state-of-the-art performance on VoD and TJ4DRadSet datasets.

Method: R4Det









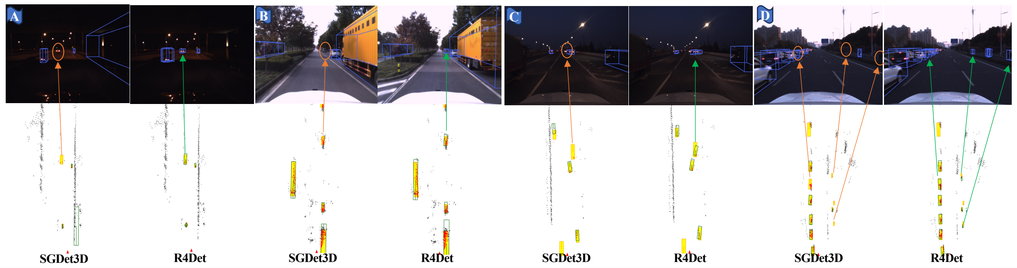
Experimental Results

Table 1. **Comparison of 3D object detection results on the `test` set of TJ4DRadSet [27].** In the modality column, R denotes 4D radar and C denotes camera. The best values are in **bold**.

Method	Modality	AP_{3D} (%)				mAP_{3D}	AP_{BEV} (%)				mAP_{BEV}
		Car	Pedestrian	Cyclist	Truck		Car	Pedestrian	Cyclist	Truck	
PointPillars [9]	R	21.26	28.33	52.47	11.18	28.31	38.34	32.26	56.11	18.19	36.23
CenterPoint [25]	R	22.03	25.02	53.32	15.92	29.07	33.03	27.87	58.74	25.09	36.18
RadarPillarNet [28]	R	28.45	26.24	51.57	15.20	30.37	45.72	29.19	56.89	25.17	39.24
SMURF [19]	R	28.47	26.22	54.61	22.64	32.99	43.13	29.19	58.81	32.80	40.98
RCFusion [28]	R+C	29.72	27.17	54.93	23.56	33.85	40.89	30.95	58.30	28.92	39.76
FUTR3D [3]	R+C	-	-	-	-	32.42	-	-	-	-	37.51
BEVFusion [13]	R+C	-	-	-	-	32.71	-	-	-	-	41.12
LXL [22]	R+C	-	-	-	-	36.32	-	-	-	-	41.20
SGDet3D [1]	R+C	59.43	26.57	51.30	30.00	41.82	66.38	29.18	53.72	39.36	47.16
CVFusion [29]	R+C	51.54	29.49	49.41	29.55	40.00	58.07	31.65	51.29	35.29	44.07
R4Det	**R+C**	**63.60**	**31.24**	**62.84**	**31.46**	**47.29**	**72.36**	**33.48**	**64.58**	**45.85**	**54.07**

Table 2. **Comparison of 3D object detection results on the `val` set of VoD [17].** R and C denote 4D radar and camera modalities, respectively. The best values are in **bold**.

Methods	Modality	Entire Annotated Area (AP_{EAA}, %)				Driving Corridor (AP_{DC}, %)				FPS
		Car	Pedestrian	Cyclist	mAP	Car	Pedestrian	Cyclist	mAP	
PointPillars [9]	R	37.06	35.04	63.44	45.18	70.15	47.22	85.07	67.48	113.9
CenterPoint [25]	R	32.74	38.00	65.51	45.42	62.01	48.18	84.98	65.06	-
RadarPillarNet [28]	R	39.30	35.10	63.63	46.01	71.65	42.80	83.14	65.86	98.8
SMURF [19]	R	42.31	39.09	71.50	50.97	71.74	50.54	86.87	69.72	-
RCFusion [28]	R+C	41.70	38.95	68.31	49.65	71.87	47.50	88.33	69.23	9.0
FUTR3D [3]	R+C	46.01	35.11	65.98	49.03	78.66	43.10	86.19	69.32	7.3
BEVFusion [13]	R+C	37.85	40.96	68.95	49.25	70.21	45.86	89.48	68.52	7.1
LXL [22]	R+C	42.33	49.48	77.12	56.31	72.18	58.30	88.31	72.93	6.1
SGDet3D [1]	R+C	53.16	49.98	76.11	59.75	81.13	60.91	90.22	77.42	9.2
CVFusion [29]	R+C	60.87	**57.89**	77.46	65.41	89.86	**68.79**	88.62	82.42	5.4
R4Det	R+C	**66.90**	55.42	**77.75**	**66.69**	**90.62**	66.47	**93.96**	**83.68**	8.3



Acknowledgments: This work was supported by the National Natural Science Foundation of China (Grant No. 62176007).