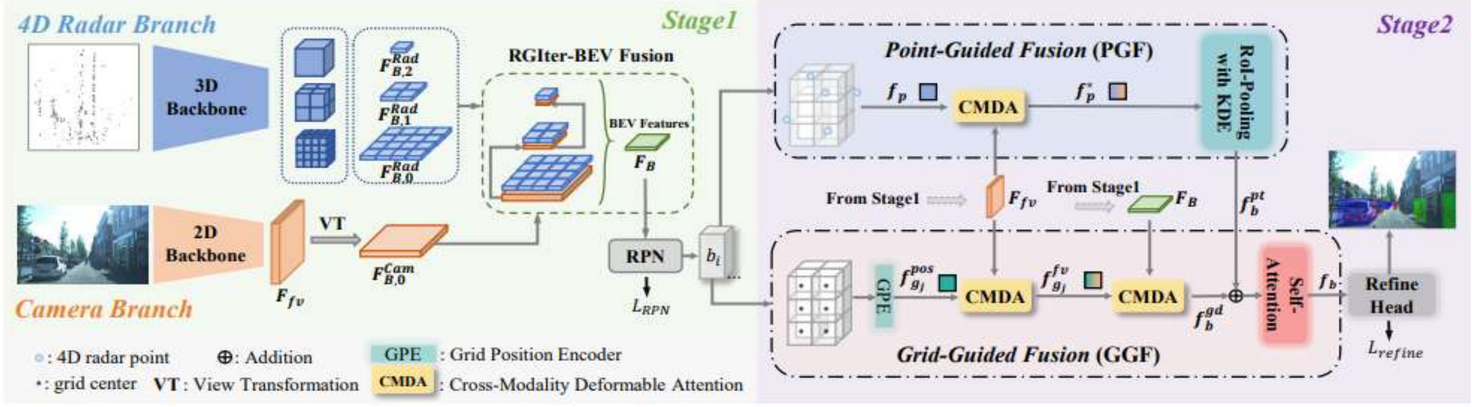
CVFusion: Cross-View Fusion of 4D Radar and Camera for 3D Object Detection

Hanzhi Zhong Zhiyu Xiang* Ruoyu Xu Jingyun Fu Peng Xu Shaohong Wang Zhihao Yang Tianyu Pu Eryun Liu

ZHEJIANG UNIVERSITY

ICCV OCT 19-23, 2025 HONOLULU HAWAII

CVFusion



Benchmark comparison

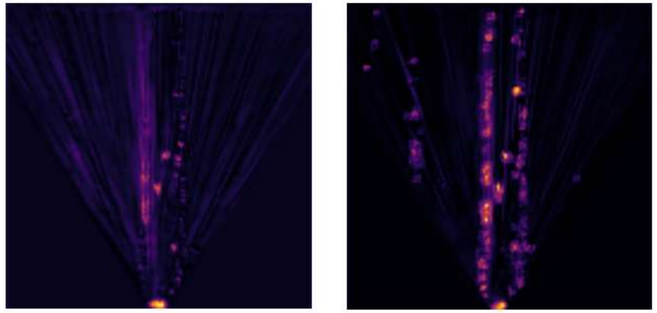
Method	Modality	Entire Annotated Area (%)				Driving Corridor Area (%)				FPS
		Car	Ped.	Cyclist	mAP	Car	Ped.	Cyclist	mAP	
PointPillars* [12]	L	61.11	49.77	72.58	61.16	90.89	54.63	94.29	79.93	43.7
MVFAN [34]	R	34.05	27.27	57.14	39.42	69.81	38.65	84.87	64.38	45.1
PointPillars* [12]	L→R	41.95	38.09	67.01	49.02	70.56	48.41	87.13	68.70	45.5
SECOND* [35]	L→R	39.53	36.23	65.68	47.14	71.85	46.95	86.68	68.49	51.3
PV-RCNN‡* [26]	L→R	40.90	41.89	73.61	52.13	72.11	51.78	85.89	69.93	22.6
SMIFormer [27]	R	39.53	41.88	64.91	48.77	77.04	53.40	82.95	71.13	16.4
SMURF [15]	R	42.31	39.09	71.50	50.97	71.74	50.54	86.87	69.72	\
FUTR3D* [2]	(R_{3D}→R)+C	46.01	35.11	65.98	49.03	78.66	43.10	86.19	69.32	7.3
BEVFusion* [18]	(R_{3D}→R)+C	37.85	40.96	68.95	49.25	70.21	45.86	**89.48**	68.52	7.1
LoGoNet‡* [13]	(L→R)+C	51.72	49.00	75.54	58.75	81.19	59.68	86.15	75.67	5.8
RCFusion [38]	R+C	41.70	38.95	68.31	49.65	71.87	47.50	88.33	69.23	\
RCBEVDet [14]	R+C	40.63	38.86	70.48	49.99	72.48	49.89	87.01	69.80	\
LXL [32]	R+C	42.33	49.48	77.12	56.31	72.18	58.30	88.31	72.93	6.1
CVFusion (Stage1-only)	R+C	52.53	50.76	75.80	59.70	81.30	61.44	87.23	76.66	6.9
CVFusion‡ (Ours)	R+C	**60.87**	**57.89**	**77.46**	**65.41**	**89.86**	**68.79**	88.62	**82.42**	5.4

- Our method has made significant improvements on all metrics and achieved new state-of-the-art, far exceeding the previous best 4D radar based method LXL.
- Compared with the two-stage network LoGoNet which generates proposal boxes from point-only BEV features, our method also performs much better thanks to the RGIter BEV fusion and elegant cross-view fusion.

Contributions

- We propose a novel cross-view fusion method for 3D object detection based on 4D radar and camera;
- A radar guided iterative BEV fusion module is proposed in stage1 to better align the BEV features and generate high-quality proposals;
- Point- and grid-guided fusion modules are proposed in the stage2 to effectively fuse multiple-view features;
- We achieve the new state-of-the-art performance on VoD and TJ4DRadSet.

Comparison of Visualization



the PGF-only (left) and the full stage2 (right)

Contacts

Email: zhonghanzhi@zju.edu.cn
Add: No. 38, Zheda Road, Lingyin Subdistrict, Xihu District, Hangzhou, Zhejiang Province, China, 310007