RobuRCDet: Enhancing Robustness of Radar-Camera Fusion in Bird's Eye View for 3D Object Detection

ICLR 2025

Contact us: yuejingtong137@gmail.com

Jingtong Yue[1*] Zhiwei Lin[1*] Xin Lin[3] Xiaoyu Zhou[1]

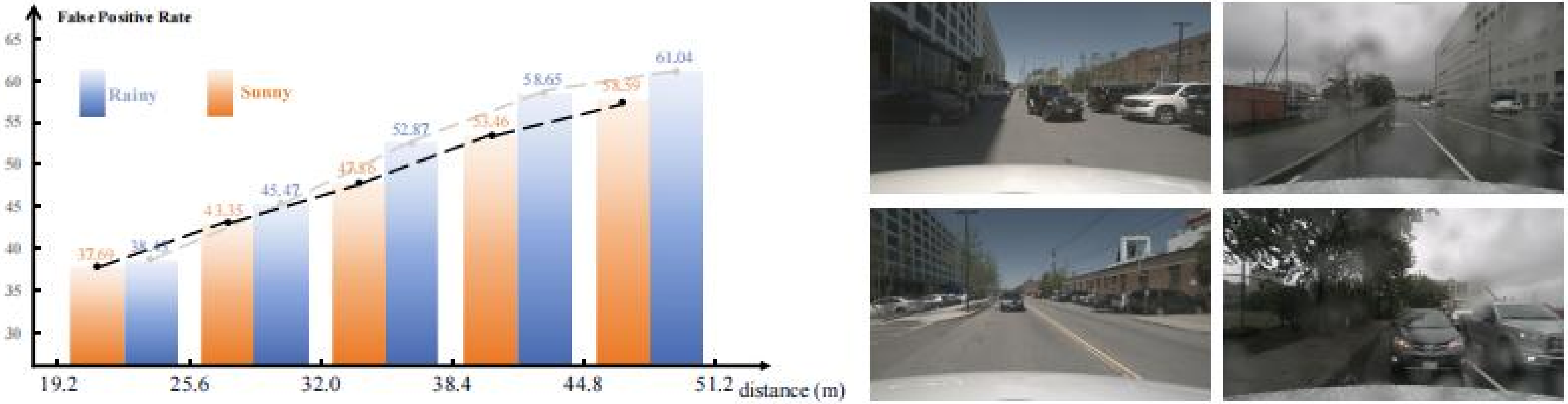
Xiangtai Li[4] Lu Qi[5] Yongtao Wang[1] Ming-Hsuan Yang[2]

[1] Peking University [2] University of California, Merced [3] Sichuan University

[4] Nanyang Technological University [5] Insta360 *Equal advising

Motivation

- Both the inherent noise of the radar itself and the noise in the image signals will have a significant impact on the robustness of the model.
- Among all the external noises, addressing the noise caused by adverse weather conditions is the most intuitive and pressing issue.



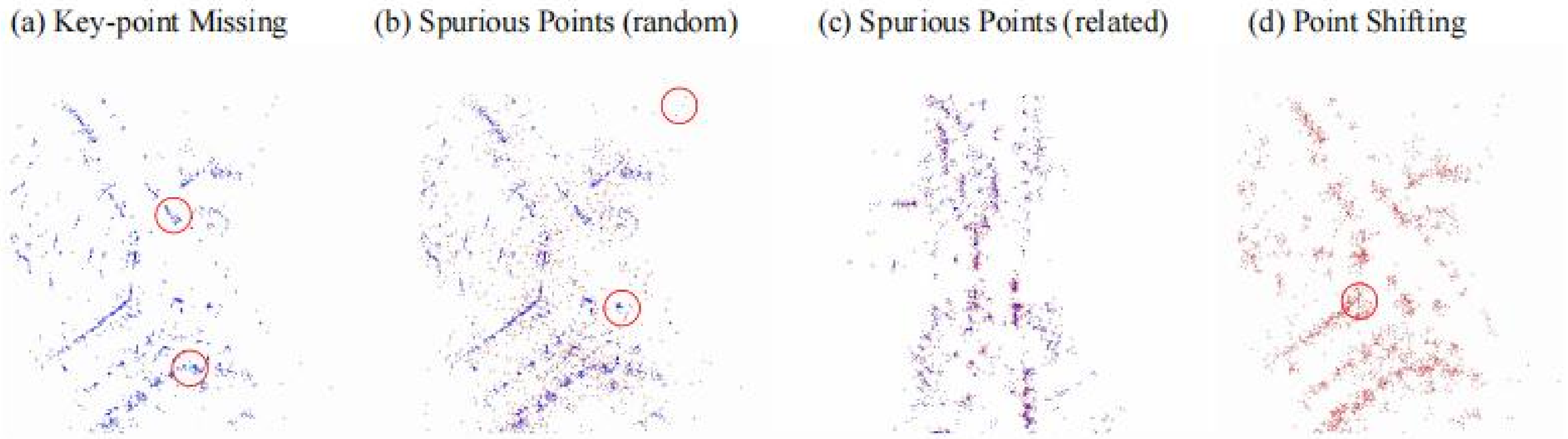
(a) Comparasion of radar point false positive rate on rainy and sunny days.

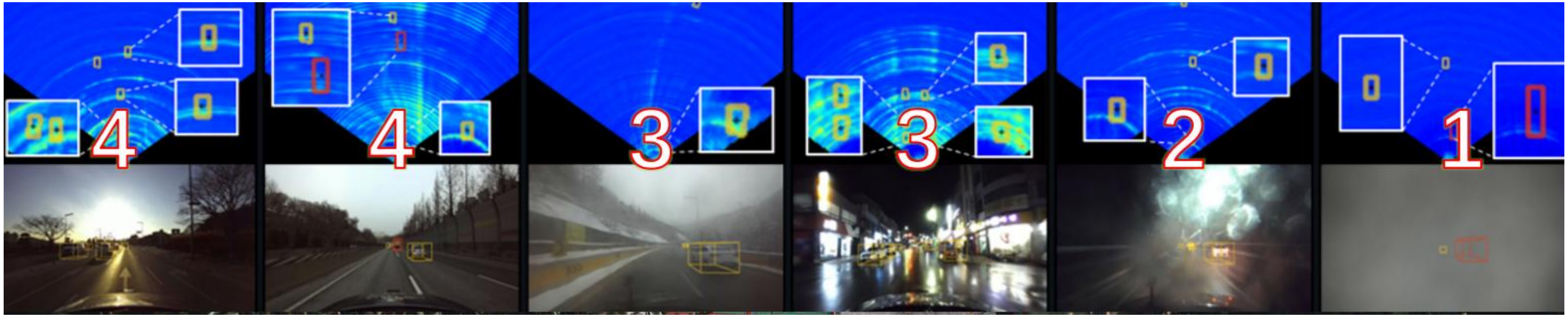
(b) Comparasion of image visibility on rainy and sunny days.

Figure 1: **Illustration of radar and camera noise on sunny and rainy days.** Radar noise increases with distance from the radar sensor and is greater in rainy weather.

Motivation

- The impact of noise on the radar pattern is mainly reflected in the density and distribution of the point cloud.





Framework

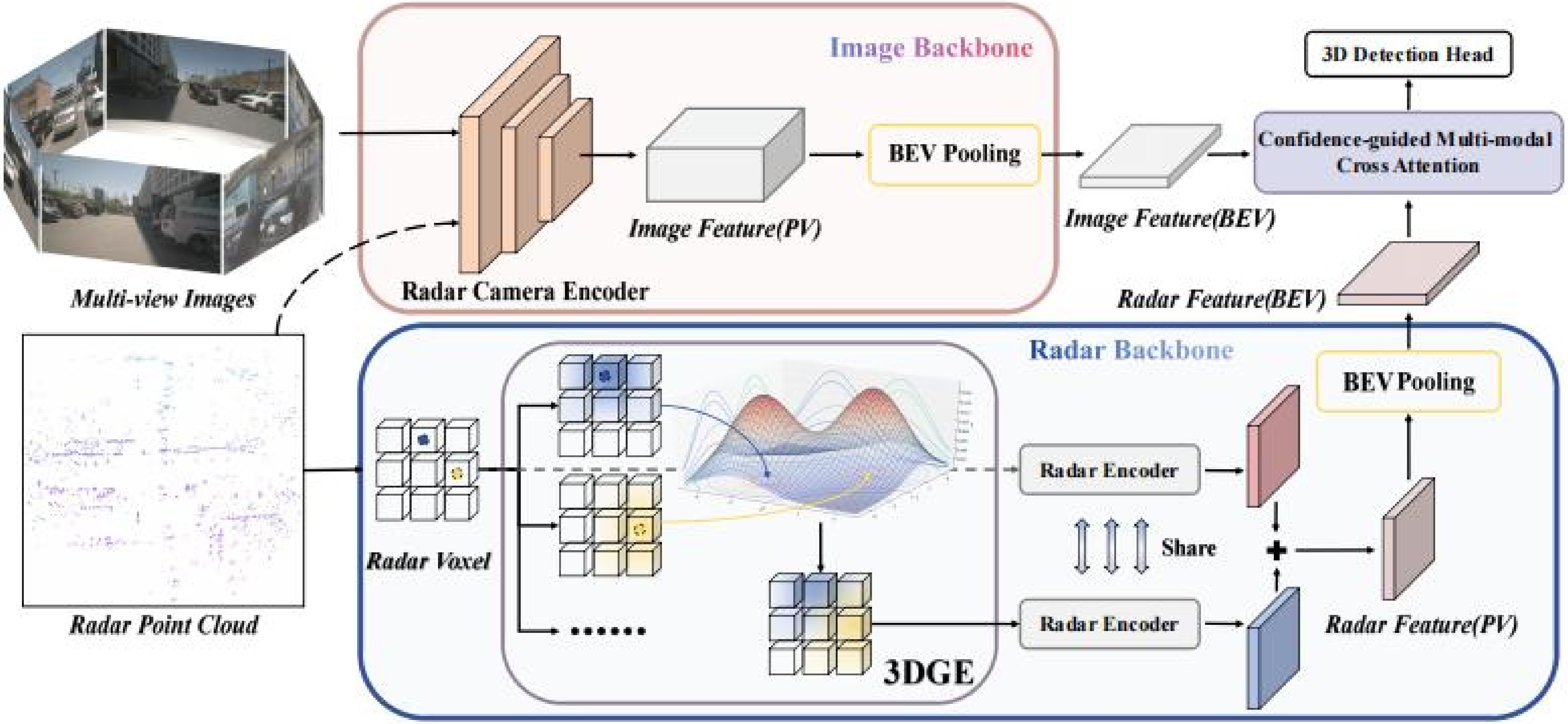
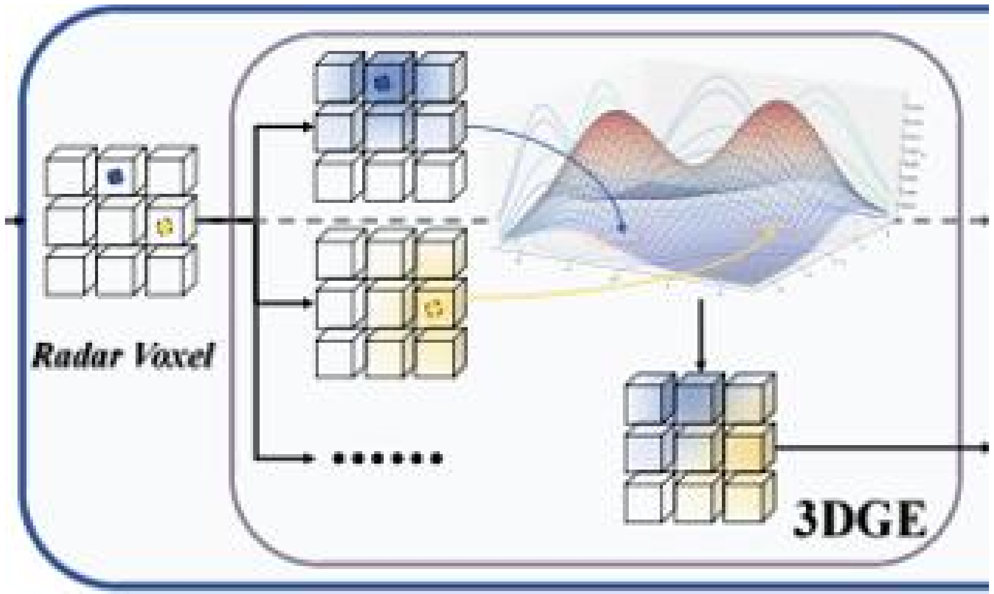


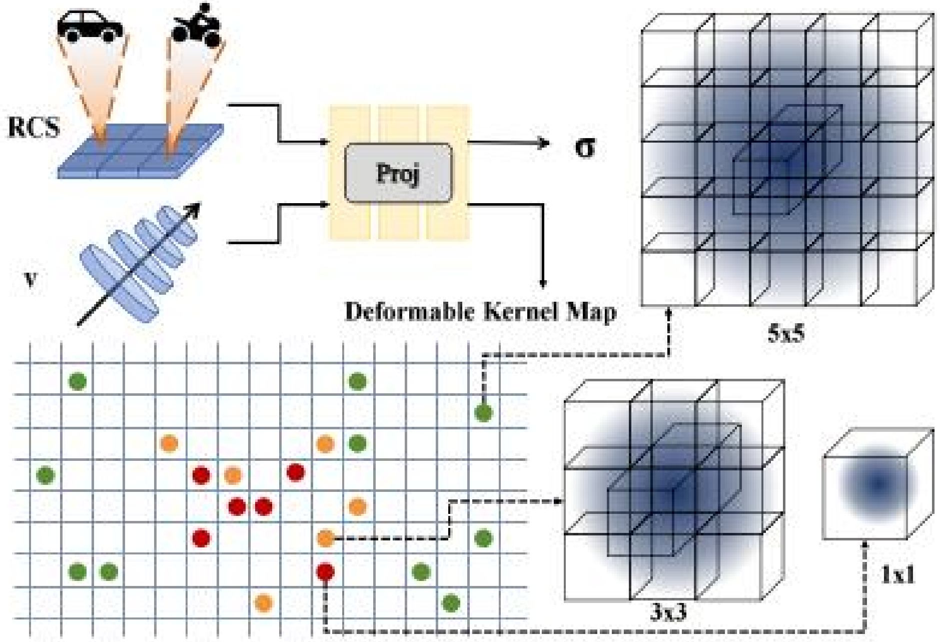
Image Backbone
3D Detection Head
BEV Pooling
Confidence-guided Multi-modal Cross Attention
Image Feature(PV)
Image Feature(BEV)
Multi-view Images
Radar Camera Encoder
Radar Feature(BEV)
Radar Backbone
BEV Pooling
Radar Encoder
Share
Radar Encoder
Radar Voxel
Radar Point Cloud
3DGE
Radar Feature(PV)

Methods

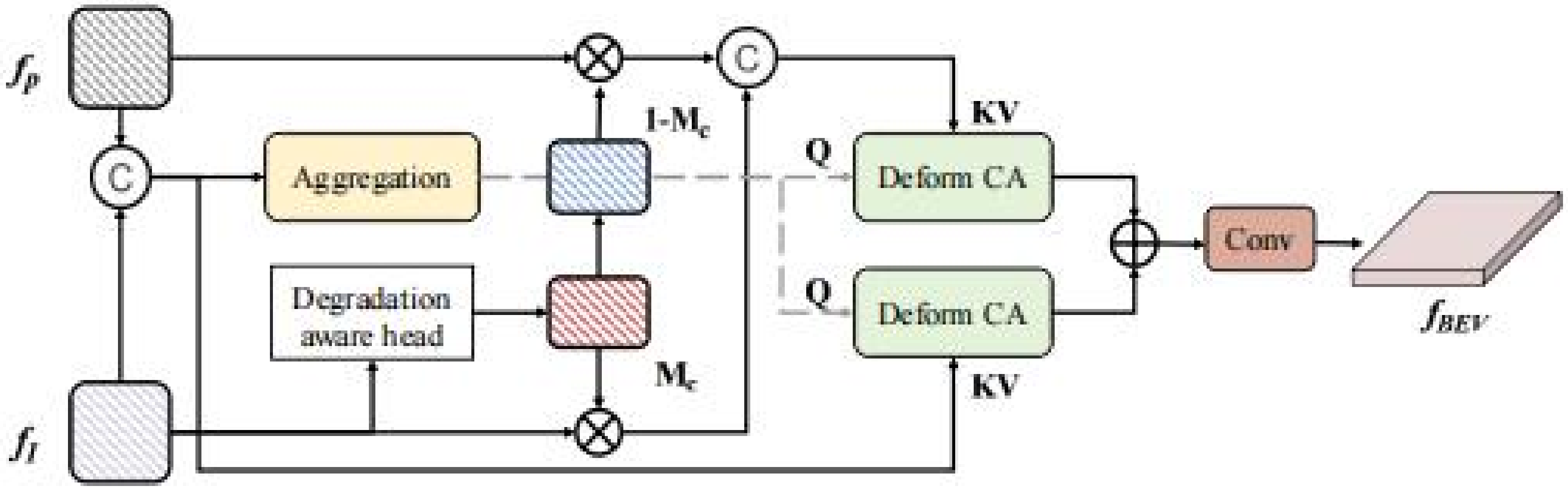
- 3DGE performs relative denoising of the radar signal by utilizing the difference in the density of radar points between the target area and the non-target area.
- σ and the Deformable Kernel Map determine the region and weights of the 3D Gaussian expansion, and they are key parameters that determine the effect.

$$V_{radar}^{3DGE}(x,y,z,RCS,v)=\frac{V(RCS,v)}{2\times\pi\times\sigma^2}\begin{cases}\exp^{\frac{(x-x_p)^2+(y-y_p)^2}{2\sigma}}, & |x-x_p|\in\lambda_p,\ |y-y_p|\in\lambda_p,\\ 0, & \text{otherwise},\end{cases}$$





Methods



$$M_c = \text{Softmax}(\text{MLP}(f_I)),$$

$$f_A = W(\text{Concat}(LN(f_I), LN(f_p))),$$
$$f_{mm} = \text{Concat}(LN(M_c \times f_I), LN((1 - M_c) \times f_p)),$$

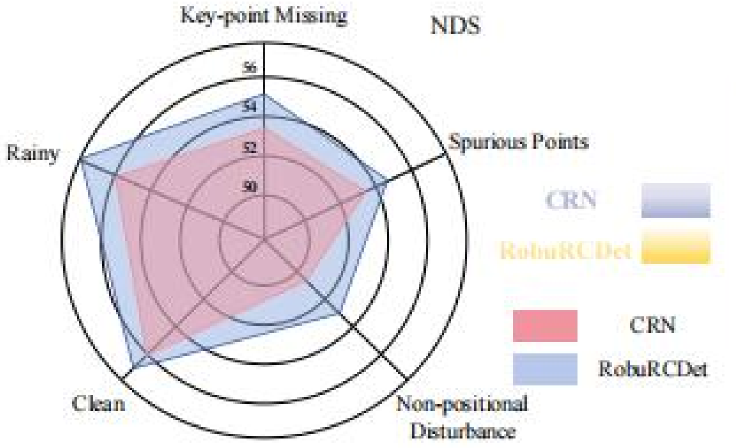
$$f_{BEV} = \text{Conv}(\text{Deform CA}(f_A, \text{Concat}(f_I, f_p)) + \text{Deform CA}(f_A, f_{mm})),$$

Results

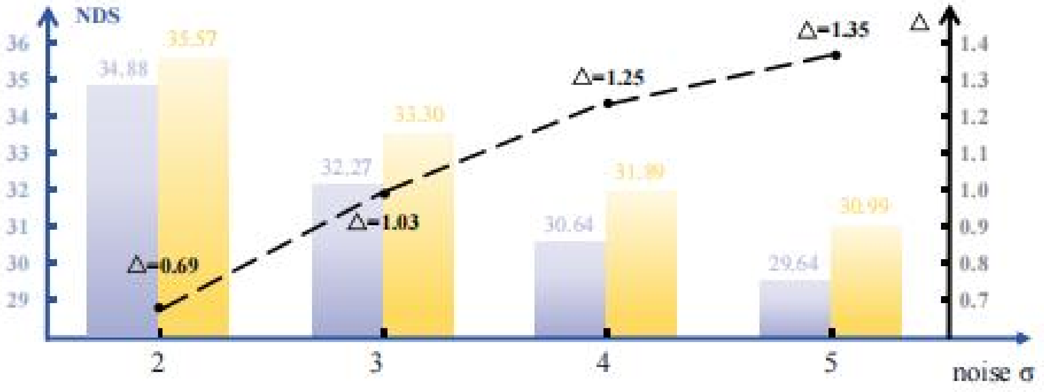
Table 1: **3D Object Detection on nuScenes `val` set.** 'C' and 'R' represent camera and radar, respectively. Some results are borrowed from RCBEVDet (Lin et al., 2024).

Methods	Input	Backbone	Image Size	NDS↑	mAP↑	mATE↓	mASE↓	mAOE↓	mAVE↓	mAAE↓
CenterFusion	C+R	DLA34	448×800	45.3	33.2	0.649	0.263	0.535	0.540	0.142
CRAFT	C+R	DLA34	448×800	51.7	41.1	0.494	0.276	0.454	0.486	0.176
RCBEV4d	C+R	Swin-T	256×704	49.7	38.1	0.526	0.272	0.445	0.465	0.185
CRN	C+R	R18	256×704	54.2	44.9	0.518	0.283	0.552	0.279	0.180
RCBEVDet	C+R	R18	256×704	54.8	42.9	0.502	0.291	0.432	0.210	0.178
RobuRCDet	C+R	R18	256×704	**55.0**	**45.5**	0.516	0.287	0.521	0.281	0.184
BEVDet	C	R50	256×704	39.2	31.2	0.691	0.272	0.523	0.909	0.247
BEVDepth	C	R50	256×704	47.5	35.1	0.639	0.267	0.479	0.428	0.198
SOLOFusion	C	R50	256×704	53.4	42.7	0.567	0.274	0.411	0.252	0.188
StreamPETR	C	R50	256×704	54.0	43.2	0.581	0.272	0.413	0.295	0.195
CRN	C+R	R50	256×704	56.0	49.0	0.487	0.277	0.542	0.344	0.197
RCBEVDet	C+R	R50	256×704	**56.8**	45.3	0.486	0.285	0.404	0.220	0.192
RobuRCDet	C+R	R50	256×704	56.7	**51.2**	**0.481**	**0.273**	0.499	0.317	0.193

Results



(a) Comparasion of CRN and RobuRCDet on five degradation types.



(b) Comparasion of CRN and RobuRCDet on Point Shifting with various σ.

IB	RB	3DGE	CMCA	Normal Condition				Corruption Condition			
				NDS↑	mAP↑	mATE↓	mAP (Car)↑	NDS↑	mAP↑	mATE↓	mAP (Car)↑
✓				43.9	33.2	0.716	50.4	-	-	-	-
✓	✓			54.3	42.4	0.536	68.4	28.5	23.9	0.709	39.5
✓	✓	✓		54.9↑0.6	**46.1**↑3.7	0.523↓0.013	**71.5**↑3.1	33.6 ↑5.1	29.4↑5.5	0.677↓0.032	47.3↑7.8
✓	✓		✓	**55.2**↑0.9	45.8↑3.4	0.531↓0.005	70.7↑2.3	33.1↑4.6	28.6↑4.7	0.681↓0.028	46.7↑7.2
✓	✓	✓	✓	55.0↑0.7	45.5↑3.1	**0.516**↓0.020	70.7↑2.3	34.1↑5.6	30.07↑6.14	0.635↓0.074	48.7↑9.2

Conclusion

1. RobuRCDet: A radar-camera fusion method designed to enhance the robustness of 3D object detection.

2. Two modules: Encourage the model to pay more attention to the target area and the characteristics of high-quality signal features. .

3. Outperforms previous methods in challenging degraded scenarios.

Limitations and Future works:

1. Task
(1). Synthetic low-quality scene
(2). Real-world low-quality scene
(3). Better accuracy with sparse encoder

2. Generalization and transfer on other models and tasks:
(1). Will 3DGE work on other models?
(2). Segmentation&tracing task version

THANKS!

Contact us: yuejingtong137@gmail.com