V2X-R: Cooperative LiDAR-4D Radar Fusion with Denoising Diffusion for 3D Object Detection

Supplementary Material

1. Additional Details of V2X-R Dataset

Here we will introduce some additional details about our V2X-R dataset, to help researchers using the V2X-R dataset get started quickly.

1.1. Calibration of sensors

We provide calibration information for each sensor (LiDAR, 4D radar, camera) of each agent for inter-sensor fusion. In particular, the exported 4D radar point cloud has been converted to the LiDAR coordinate system of the corresponding agent in advance to facilitate fusion, so the 4D radar point cloud is referenced to the LiDAR coordinate system. If necessary, it can be converted back by the LiDAR coordinate system to the 4D radar coordinate system.

Sensor	Data Structure	File Type	Attributes
LiDAR	Point Cloud	.pcd	$N \times 4$ $[x, y, z, intensity]$
4D Radar	Point Cloud	.pcd	$N \times 4$ $[x, y, z, velocity]$
Camera	Image	.png	$800 \times 600 \times 3$ $[R, G, B]$

Table 1. The detailed information of V2X-R data.

1.2. Information of data

The data attributes corresponding to LiDAR, 4D radar, and camera are shown in Table 1. Both LiDAR and 4D radar sensors provide $N \times 4$ point clouds, where N represents the number of points. It is worth noting that the 4D radar was originally exported in the CARLA simulator as an array of [azimuth, altitude, depth, velocity] in polar coordinates, which we converted to a Cartesian coordinate system. In addition, an 800x600x3 RGB image is obtained for each of the 4 cameras of each agent.

1.3. Information of data collection

We saved the critical collection details of each scene sequence by splits (can be found in "data_protocol.yaml" of the released dataset), which includes agent information and scene sources for each sequence. Additional detailed acquisition information, such as the configuration of each agent's trajectory, can be queried in our V2X-R documentation.

2. Additional Results Analysis

Modality	Communication volumes (4B as unit, log-scale)/ Latency time (ms as unit, 27Mbps as transmission speed)				
	AttFuse	V2XViT	CoBEVT	SICP	L4DR
LiDAR-based	18.2/357.7	16.0/75.2	17.6/226.1	14.9/35.2	-
4D radar-based	14.6/28.4	14.0/18.6	15.0/37.4	13.6/14.7	-
LiDAR+4D radar	18.4/402.1	16.9/142.0	18.1/317.6	15.2/42.7	17.7/253.1

Table 2. Analysis of Communication Costs and Bandwidth for Different Models (SM2MM,SA2MA).

2.1. Transmission cost and bandwidth.

We have calculated transmission cost (count of non-zero elements in the feature map) and latency for different modalities. As shown in Table 2, 4D radar has the advantage of sparse feature transmission, and its fusion with LiDAR brings acceptable transmission cost and latency time.

Method	Std of localization error (m)/(Val-AP@50)						
	0.0	0.1	0.2	0.3	0.4	0.5	0.6
AttFuse	84.30	81.28	79.87	77.84	73.09	68.30	62.84
CoBEVT	87.02	85.26	83.30	81.33	77.83	70.79	63.47
AdaFusion	87.31	87.15	86.25	84.22	79.79	72.89	64.15
Where2comm	85.78	85.58	84.78	82.91	79.57	76.23	71.77

Table 3. Performance of different models under varying degrees of localization error.

2.2. localization noise

Following Where2comm [3], we added different degrees of localization error to our V2X-R dataset to conduct experiments. As shown in Table 3, all methods experience performance degradation to varying degrees as localization error

Methods	Modality	Class	Metric	Total	Normal	Overcast	Fog	Rain	Sleet	Lightsnow	Heavysnow
AttFuse [16]	L+4DR	Sedan	AP_{BEV}	70.3	68.0	**89.4**	90.5	79.5	66.9	88.3	60.4
			AP_{3D}	69.0	66.8	79.4	88.6	70.7	59.2	**86.2**	58.6
		Bus	AP_{BEV}	**64.3**	**59.4**	**75.7**	-	0.4	**66.2**	80.7	70.8
			AP_{3D}	51.0	**53.3**	**75.6**	-	0.2	**65.6**	76.8	36.3
AttFuse w/ MDD	L+4DR	Sedan	AP_{BEV}	**76.8**	**73.8**	88.9	**90.8**	**79.7**	**68.7**	**88.4**	**61.5**
			AP_{3D}	**74.0**	**67.2**	**85.5**	**89.6**	**75.7**	**64.6**	84.5	**59.8**
		Bus	AP_{BEV}	64.1	55.3	72.0	-	**15.1**	62.7	**97.5**	**73.9**
			AP_{3D}	**54.6**	52.8	71.0	-	**15.0**	61.3	**85.2**	**42.6**

Table 4. Quantitative results of different 3D object detection methods on K-Radar dataset. We present the modality of each method (L+4DR: LiDAR-4D radar fusion) and detailed performance for each weather condition. Best in **bold**.

Modality	**Method**	**Epoch**	**Batch_size**	**Max_Agents**	**Learning_Rate**	**LR_Scheduler**
LiDAR	V2XViT [15]	20	2	5	0.001	Multistep
	AttFuse [16]	30	4	5	0.002	Multistep
	Where2comm [3]	50	1	5	0.0002	Cosineannealwarm
	SCOPE [18]	30	2	5	0.002	Multistep
	CoBEVT [17]	30	2	5	0.001	Cosineannealwarm
	CoAlign [6]	15	2	5	0.002	Multistep
	AdaFusion [5]	30	2	5	0.0005	Multistep
	SICP [11]	20	1	5	0.001	Multistep
	MACP [7]	20	4	5	0.0002	Cosineannealwarm
4D Radar	PFA-Net [14]	30	4	5	0.001	Multistep
	RTNH [9]	15	4	5	0.001	Multistep
	V2XViT [15]	20	2	5	0.001	Multistep
	AttFuse [16]	30	4	5	0.002	Multistep
	Where2comm [3]	15	1	5	0.0002	Cosineannealwarm
	SCOPE [18]	15	2	5	0.002	Multistep
	CoBEVT [17]	30	2	5	0.001	Cosineannealwarm
	CoAlign [6]	20	2	5	0.002	Multistep
	AdaFusion [5]	15	2	5	0.0005	Multistep
	SICP [11]	20	1	5	0.001	Multistep
LiDAR+4D Radar	InterFusion [13]	20	1	5	0.002	Multistep
	L4DR [4]	30	2	5	0.002	Multistep
	V2XViT [15]	30	2	5	0.001	Multistep
	AttFuse [16]	30	2	5	0.002	Multistep
	SCOPE [18]	40	2	5	0.002	Multistep
	Where2comm [3]	30	4	5	0.0002	Cosineannealwarm
	CoBEVT [17]	40	2	5	0.001	Cosineannealwarm
	CoAlign [6]	30	2	5	0.002	Multistep
	AdaFusion [5]	40	2	5	0.0005	Multistep
	SICP [11]	20	1	5	0.001	Multistep

Table 5. Experimental parameter settings (epoch, batch_size, max_agent, learing_rate, lr_scheduler) for different modalities and methods in our benchmark section.

increases. This helps to explore performance under real localization errors.

2.3. Performance of various weather on K-Radar.

To further demonstrate the performance improvement of our MDD module in various real-world adverse weather conditions. We provide more detailed results on the K-Radar real adverse weather dataset rather than just the average of adverse weather. As shown in Table 4, with the addition of our MDD module, AttFuse first of all got a big boost in Total basically (except for Bus's AP_{BEV}). Under the Sedan class, there are significant improvements in

Component	Parameter	Value
Denoiser (U-net)	input_channel	128
	mid_channel	128
	timestep_channel	64
	output_channel	64
	number_layers	2
	number_resblock	2
Diffusion Process	timesteps	3
	betas	[0.005,0.0275,0.05]

Table 6. The implementation details of our MDD module.

every weather except Overcast which is similar, especially in 6.1 AP_{3D}@Overcast and 5.4 AP_{3D}@Sleet performance improvements. In addition, the results of the Bus class are well worth exploring. We find significant decreases in Normal, Overcast, and Sleet, but very significant increases in Rain, Lightsnow, and Heavysnow. We assert this is due to the nature of the larger 3D bounding boxes of the Bus class, which is particularly sensitive to the denoising module, causing a drop in some weather and a significant rise in others. Overall, however, the Total performance on the Bus class remains suggestive. These detailed analyses further validate the effectiveness of our MDD module under real-world conditions.

3. Training Detail

3.1. Benchmark

We also provide details about the training of all the benchmark models in the main text for researchers to refer to, as shown in Table 5. In addition, we will disclose the training profiles of all models and the pre-trained models. This can help researchers efficiently use our well-trained models on the V2X-R dataset or reproduce the same results.

3.2. Implementation of MDD

Here, we will provide a concrete implementation of the MDD module. It comprises a denoiser network with a U-net [12] structure and a diffusion process. Some important parameter settings are shown in Table 6.

3.3. Configuration of weather simulation

To help the reader gain a deeper understanding of the severe weather portion of the study in our V2X-R work. As shown in Table 7, we list here some important configurations for fog and snow simulations, mainly parameters used to adjust the level of adverse weather. Most of the other configurations implemented refer to the default configurations available on their official open-source code [1] [2]. In addition, the simulation code we implemented will be included in the publicly released code in the future.

[1] Fog Simulation Code
[2] Snow Simulation Code

Simulation	Parameter	Value
Fog Simulation [1]	gamma	0.000001
	alpha	0.06
	noise_variant	v2
	noise	10
	r_noise	random(1, 20)
	max_intensity	255
Snow Simulation [2]	num_intervals	64
	interval_index	random(1,64)
	snowfall_rate	0.5
	terminal_velocity	0.2
	noise_floor	0.7
	beam_divergence	0.003
	max_intensity	255

Table 7. The detailed configuration of weather simulation. The parameter names refer to the naming of the official source code and the exact meanings can be found in [1, 2].

3.4. K-Radar dataset and evaluation metrics

The K-Radar dataset [8] contains 58 sequences with 34944 frames of 64-line LiDAR, camera, and 4D radar data in various weather conditions. According to the official K-Radar split, we used 17458 frames for training and 17536 frames for testing. We adopt two evaluation metrics for 3D object detection: AP_{3D} and AP_{BEV} of the class "Sedan" and "Bus" at IoU = 0.3. We use the newest version (v2.1) of the label.

4. Visualization of V2X-R Dataset

Finally, in order to visually intuitively verify the realism of the simulated LiDAR-4D radar data on our V2X-R dataset and the advantages of the cooperative LiDAR-4D radar point cloud. We have visualized and compared the simulated LiDAR-4D radar point cloud on our V2X-R dataset with the real LiDAR-4D radar point cloud on the VoD dataset [10]. As shown in Fig. 1, it can be found that our simulated LiDAR-4D radar point cloud has a certain degree of realism. This proves the value of conducting 4D radar-related research on our V2X-R dataset. Meanwhile, by comparing the real single-agent 4D radar with the multi-agent 4D radar, it can be found that the multi-agent collaborative 4D radar has a significantly higher resolution. As we introduced in the Introduction section of the main text, the multi-agent cooperative 4D radar has a certain independent perception ability.

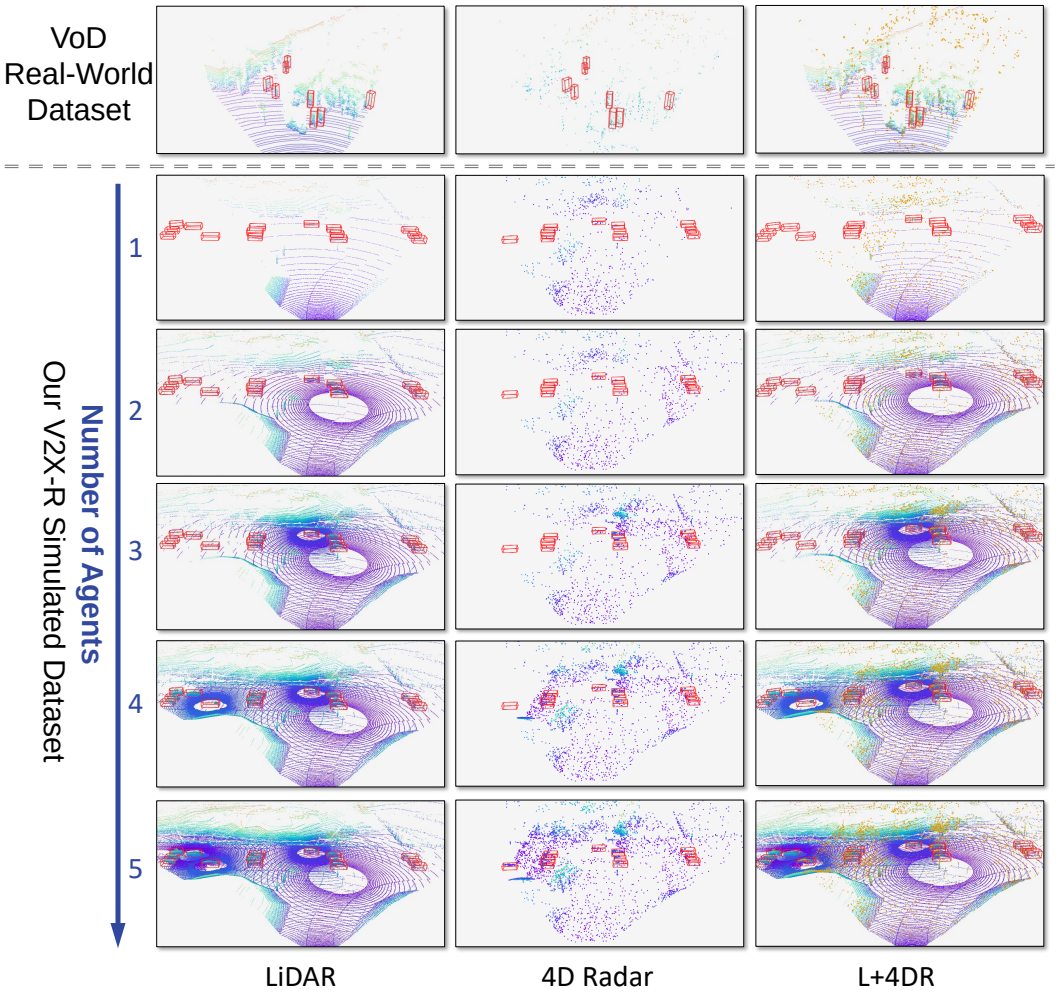

Figure 1. Visualization of our V2X-R dataset and VoD [10] real-world dataset. The L+4DR in the last column indicates that the LiDAR point cloud is visualized together with the 4D radar point cloud, where to distinguish between them, we use colored dots (slightly smaller) for the LiDAR point cloud and orange dots (slightly larger) for the 4D radar point cloud. Colored point clouds are assigned by z-axis values.

References

[1] Martin Hahner, Christos Sakaridis, Dengxin Dai, and Luc Van Gool. Fog Simulation on Real LiDAR Point Clouds for 3D Object Detection in Adverse Weather. In *ICCV*, 2021. 3

[2] Martin Hahner, Christos Sakaridis, Mario Bijelic, Felix Heide, Fisher Yu, Dengxin Dai, and Luc Van Gool. LiDAR Snowfall Simulation for Robust 3D Object Detection. In *CVPR*, 2022. 3

[3] Yue Hu, Shaoheng Fang, Zixing Lei, Yiqi Zhong, and Siheng Chen. Where2comm: Communication-efficient collaborative perception via spatial confidence maps. *Advances in neural information processing systems*, 35:4874–4886, 2022. 1, 2

[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. *arXiv preprint arXiv:2408.03677*, 2024. 2

[5] Haowen Lai, Peng Yin, and Sebastian Scherer. Adafusion: Visual-lidar fusion with adaptive weights for place recognition. *IEEE Robotics and Automation Letters*, 7(4):12038–12045, 2022. 2

[6] Yifan Lu, Quanhao Li, Baoan Liu, Mehrdad Dianati, Chen Feng, Siheng Chen, and Yanfeng Wang. Robust collaborative 3d object detection in presence of pose errors. In *2023 IEEE International Conference on Robotics and Automation (ICRA)*, pages 4812–4818, 2023. 2

[7] Yunsheng Ma, Juanwu Lu, Can Cui, Sicheng Zhao, Xu Cao, Wenqian Ye, and Ziran Wang. Macp: Efficient model adaptation for cooperative perception. In *2024 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, pages 3361–3370, 2024. 2

[8] 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. 3

[9] Dong-Hee Paek, SEUNG-HYUN KONG, and Kevin Tirta Wijaya. K-radar: 4d radar object detection for autonomous driving in various weather conditions. In *Advances in Neural Information Processing Systems*, pages 3819–3829. Curran Associates, Inc., 2022. 2

[10] Andras Palffy, Ewoud Pool, Srimannarayana Baratam, Julian FP Kooij, and Dariu M Gavrila. Multi-class road user detection with 3+ 1d radar in the view-of-delft dataset. *IEEE Robotics and Automation Letters*, 7(2):4961–4968, 2022. 3, 4

[11] Deyuan Qu, Qi Chen, Tianyu Bai, Hongsheng Lu, Heng Fan, Hao Zhang, Song Fu, and Qing Yang. Sicp: Simultaneous individual and cooperative perception for 3d object detection in connected and automated vehicles, 2024. 2

[12] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In *Medical image computing and computer-assisted intervention–MICCAI 2015: 18th international conference, Munich, Germany, October 5-9, 2015, proceedings, part III 18*, pages 234–241. Springer, 2015. 3

[13] 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. 2

[14] Baowei Xu, Xinyu Zhang, Li Wang, Xiaomei Hu, Zhiwei Li, Shuyue Pan, Jun Li, and Yongqiang Deng. Rpfa-net: a 4d radar pillar feature attention network for 3d object detection. In *2021 IEEE International Intelligent Transportation Systems Conference (ITSC)*, pages 3061–3066, 2021. 2

[15] Runsheng Xu, Hao Xiang, Zhengzhong Tu, Xin Xia, Ming-Hsuan Yang, and Jiaqi Ma. V2x-vit: Vehicle-to-everything cooperative perception with vision transformer. In *European conference on computer vision*, pages 107–124. Springer, 2022. 2

[16] Runsheng Xu, Hao Xiang, Xin Xia, Xu Han, Jinlong Li, and Jiaqi Ma. Opv2v: An open benchmark dataset and fusion pipeline for perception with vehicle-to-vehicle communication. In *2022 International Conference on Robotics and Automation (ICRA)*, pages 2583–2589. IEEE, 2022. 2

[17] Runsheng Xu, Zhengzhong Tu, Hao Xiang, Wei Shao, Bolei Zhou, and Jiaqi Ma. Cobevt: Cooperative bird's eye view semantic segmentation with sparse transformers. In *Conference on Robot Learning*, pages 989–1000. PMLR, 2023. 2

[18] Kun Yang, Dingkang Yang, Jingyu Zhang, Mingcheng Li, Yang Liu, Jing Liu, Hanqi Wang, Peng Sun, and Liang Song. Spatio-temporal domain awareness for multi-agent collaborative perception. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 23383–23392, 2023. 2