2024 IEEE Intelligent Vehicles Symposium (IV)
June 2-5, 2024. Jeju Island, Korea

V2X-DSI: A Density-Sensitive Infrastructure LiDAR Benchmark for Economic Vehicle-to-Everything Cooperative Perception

Xinyu Liu†, Baolu Li†, Runsheng Xu, Jiaqi Ma, Xiaopeng Li, Jinlong Li*, Hongkai Yu*

2024 IEEE Intelligent Vehicle Symposium (IV) | 979-8-3503-4881-1/24/$31.00 ©2024 IEEE | DOI: 10.1109/IV55156.2024.10588684

***Abstract*—Recent research has demonstrated that the Vehicle-to-Everything (V2X) communication techniques can fundamentally improve the perception system for autonomous driving by collaborating between vehicle and infrastructure sensors. LiDAR is the commonly-used sensor for V2X autonomous driving due to its robustness in challenging scenarios. However, the LiDAR sensor is expensive, so the cost of equipping LiDAR sensors to a large number of infrastructures on the large-scale roadway network is extremely high, which has limited the wide deployment of the V2X cooperative perception system. How to discover an economic V2X cooperative perception system is never been well studied before. Inspired by the cost difference of the various point cloud densities of LiDAR, we propose the first Density-Sensitive Infrastructure LiDAR benchmark for economic V2X cooperative perception, named V2X-DSI, in this paper. Using the proposed V2X-DSI benchmark, we analyze the effect of cooperative perception performance under different beam infrastructure LiDAR. We specifically assess three state-of-the-art methods, *i.e.*, OPV2V, V2X-ViT, and CoBEVT, using our V2X-DSI dataset. The results indicate that varying beam infrastructure LiDAR sensors play a crucial role in influencing cooperative perception performance.**

***Index Terms*—deep learning, vehicle-to-infrastructure cooperative perception, 3D object detection,**

I. Introduction

Due to the rapid development of the V2X communication, *e.g.*, Vehicle-to-Infrastructure (V2I), the multi-view information of same driving scene in a larger perspective range can be shared from nearby infrastructures to significantly improve the perception of ego vehicle and the reconstruction of driving scene [1]–[8]. This technique will greatly improve the safety of autonomous driving, cars, and pedestrians in many urban scenes, *e.g.*, intersection, roundabout, highway ramp. There is only one real-world large-scale dataset for V2I cooperative detection currently: DAIR-V2X [9]. In DAIR-V2X, the V2I data was collected from 28 intersections in China, where each intersection is equipped with four high-beam LiDAR sensors. The LiDAR sensor is widely used in V2I perception due to its robust quality in different scenarios.

However, the high-beam LiDAR for each infrastructure is very expensive, leading to extremely high cost on the large-scale roadway network in the future. The high-cost issue has limited the wide deployment of the V2X cooperative perception system in the real world. Taking the New York City (NYC) as an example, NYC has 13,250 intersections with traffic signals [10]. If we plan to equip one Velodyne LiDAR sensor on each of these intersections in NYC, the cost is as extremely high as approximate $1,116M and $928M for 128-beam and 64-beam LiDAR respectively, while the cost can be reduced to approximately $260M for 32-beam LiDAR and $52M for 16-beam LiDAR, as shown in Fig. 1. The LiDAR sensors with lower beamline numbers (*i.e.*, lower point cloud density) are normally much cheaper, but the only concern is about the accuracy of using low-beam infrastructure LiDAR for V2I cooperative perception.

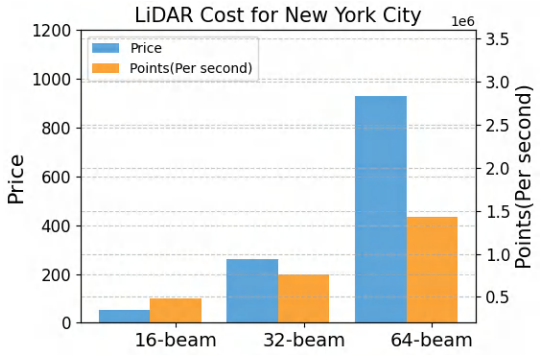


Fig. 1. **Comparison of costs and density in infrastructure LiDAR sensors.** Significant cost and density differences of various-beam infrastructure LiDAR sensors for Vehicle-to-Infrastructure (V2I) cooperative perception using New York City as an example.

From the perspective of economy theory, the large-scale deployments of sensors on infrastructures and the large-scale usages of Connected and Automated Vehicles (CAV) can benefit each other, leading to the overall cost reduction of each side. This motivates us to conduct novel research to discover an economic V2X cooperative perception system using low-beam infrastructure LiDAR sensors, which has never studied before. Specifically, inspired by the cost difference of the various point cloud densities of LiDAR, we propose the first Density-Sensitive Infrastructure LiDAR benchmark for economic V2X cooperative perception using the open-source CARLA simulation software [11], named V2X-DSI, in this paper. The proposed V2X-DSI benchmark has 128-beam, 64-beam, 32-beam, and 16-beam infrastructure LiDAR sensors for the V2I cooperative perception, including **56,984** frames in **57** diverse urban scenarios.

As shown in Table I, certain datasets exclusively focus on a singular type of cooperative perception, assuming uniformity in the level of beam infrastructure LiDAR. For instance, OPV2V [12] and V2V4Real [13] center around Vehicle-to-Vehicle (V2V) scenarios, while DAIR-V2X-C [14] solely addresses V2I scenarios. Furthermore, given that these datasets lack diversity in beam infrastructure LiDAR levels, they do not account for the impacts of diverse beamline numbers on cooperative perception performance. Our V2X-DSI benchmark

Jinlong Li, Baolu Li, Xinyu Liu, and Hongkai Yu are with Cleveland State University, Cleveland, OH 44115, USA. Runsheng Xu and Jiaqi Ma are with University of California, Los Angeles, CA 90024, USA. Xiaopeng Li is with the University of Wisconsin–Madison, Madison, WI 53706, USA.
† indicates co-first authors. *Co-corresponding authors.

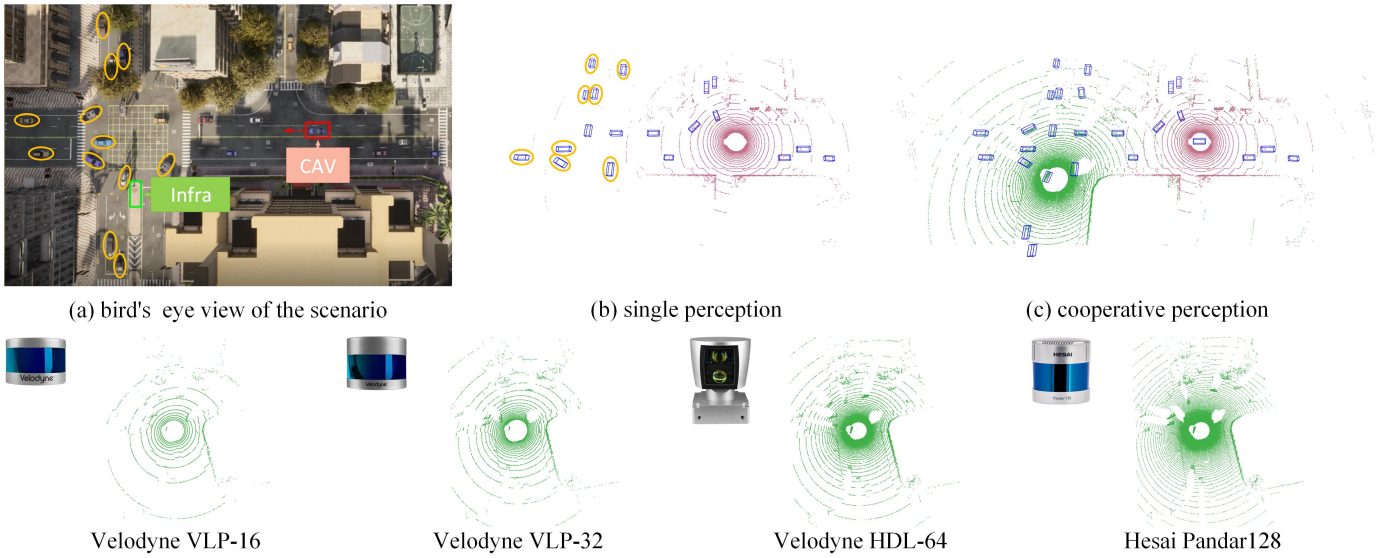


Fig. 2. **A data sample from the proposed V2X-DSI dataset.** (a) A simulated scenario in CARLA where a CAV and infrastructure are located at different sides of a busy intersection. (b) the LiDAR point cloud from the single CAV perception. (c) The aggregated LiDAR point clouds of CAV and infrastructure. The yellow circles represent the cars that are not detected for the ego vehicle due to the occlusion and long-distance but can be seen by infrastructure. The LiDAR sensor of infrastructure can deploy four types of LiDAR beams shown in the second line.

enables a comprehensive analysis of cooperative perception under different beam infrastructure LiDAR sensors.

Using the proposed V2X-DSI benchmark, we analyze the effect of cooperative perception performance under different beam infrastructure LiDAR sensors. Three state-of-the-art cooperative perception methods *i.e.*, OPV2V [12], V2X-ViT [1], and CoBEVT [15], are assessed in the our experiments. Training these models on the higher beam LiDAR from the V2X-DSI training set leads to improved performance. This suggests that the use of high-density LiDAR sensor information can offer more precise details, leading to enhanced perception performance. However, these models trained on high-beam data experience a performance drop when deployed in low-beam LiDAR scenarios. Here, we employed a fine-tuned approach to address these challenges in our experiments. The proposed V2X-DSI benchmark and related source code will be publicized after paper acceptance, then we expect that the to-be-publicized V2X-DSI benchmark could attract more researchers to investigate the economic V2X cooperative perception problem. The contributions of this paper are summarized as follows.

- To the best of our knowledge, this paper is the **first research** on the economic V2X cooperative perception by taking the infrastructure LiDAR point cloud density into consideration.
- This paper proposes a novel Density-Sensitive Infrastructure LiDAR benchmark named V2X-DSI for economic V2X cooperative perception, including 128-beam, 64-beam, 32-beam, 16-beam infrastructure LiDAR sensors of 56,984 frames in 57 diverse urban scenarios.
- This paper analyzes the effect of cooperative perception performance under different beam infrastructure LiDAR sensors by utilizing three cooperative perception methods *i.e.*, OPV2V [12], V2X-ViT [1], and CoBEVT [15].

II. Related Work

A. Vehicle-to-Infrastructure Cooperative Perception

The utilization of the Vehicle-to-Infrastructure communication empowers the multi-agent setup to overcome the constraints of occlusions. This enables the system to attain superior large-range perception capabilities [18], [19]. To achieve a balance between precision and bandwidth demands, advanced techniques opt for intermediate neural features as shared information instead of utilizing raw sensing data or detected outputs. Certain studies employ the self-attention mechanism as the core fusion method to improve the intersection of shared V2I visual information [12], [20]. OPV2V [12] introduces a self-attention module with a single head to fuse the received intermediate features, V2VAM [20] leverages two intra-agent and inter-agent attention modules to enhance sharing information. Vision Transformers (ViT) have attained increasing attention and been applied to the cooperative perception. V2X-ViT [1] presents a unified ViT architecture for V2X perception, capturing the heterogeneous nature of V2X systems. SCOPE [21] introduces a deformable ViT for spatio-temporal feature integration to enhance perception performance. CoBEVT [15] proposes a local and sparsely global based ViT perception framework that collaboratively generates predictions for the efficient intersection.

B. V2X Cooperative Perception Benchmarks

Large-scale datasets are essential for training perception models. Similar to multi-agent perception systems, V2X/V2I datasets are crucial for training reliable cooperative perception models. However, acquiring V2X data from the real world poses significant challenges and expenses. Studies have employed simulators such as CARLA [22] to collect simulated multi-agent data. V2X-Sim [16] presents sensor recordings

TABLE I
COMPARISON OF THE PROPOSED DATASET AND EXISTING REPRESENTATIVE AUTONOMOUS DRIVING DATASETS.

Dataset	Pub. Year	Real/ Sim	V2X	Infra LiDAR Type	RGB images	LiDAR	3D boxes	Classes	Locations
V2X-Sim [16]	RAL 2022	Sim	V2V,V2I	1	60K	10K	26.6K	1	CARLA
OPV2V [12]	ICRA 2022	Sim	V2V	1	44K	11K	230K	1	CARLA
DAIR-V2X-C [14]	CVPR 2022	Real	V2I	1	39K	39K	464K	10	Beijing, CN
V2XSet [1]	ECCV 2022	Sim	V2V,V2I	1	44K	11K	230K	1	CARLA
V2V4Real [13]	CVPR 2023	Real	V2V	1	40K	20K	240K	5	Ohio, USA
V2X-Seq [17]	CVPR 2023	Real	V2V,V2I	1	15K	-	-	8	Beijing, CN
V2X-DSI	-	Sim	V2I	4	14K	57K	222K	1	CARLA

from multiple vehicles and the roadside unit (RSU) based on CARLA, facilitating collaborative perception in a multi-agent environment. V2XSet [1] gathers simulated data from both infrastructure and CAV sensors in CARLA. Infrastructure sensors were positioned on traffic light poles or street light poles. Within real V2X datasets, DAIR-V2X [14] is a dataset encompassing multi-modal, multi-view data collected from authentic V2I multi-agent driving scenarios. V2X-Seq [17] is introduced to alleviate the scarcity of real-world sequential data. V2X-Seq is specifically designed for tracking and trajectory forecasting tasks. However, in both DAIR-V2X and V2X-Seq, infrastructure LiDAR sensors employ high-beam technology, potentially posing real-world cost challenges. This paper introduces the first Density-Sensitive Infrastructure LiDAR benchmark for economic V2X cooperative perception, which is never studied before.

C. V2X Challenges of V2X Deployment

Leveraging V2X communication in multi-agent cooperation substantially improves the perception efficiency [23]–[28]. However, implementing a real-world multi-agent perception system demands resolution of various challenges [29]–[35], encompassing communication latency, model heterogeneity, localization errors, lossy communication, and the potential for adversarial attacks. V2X-ViT [1] utilizes the Delay-aware Positional Encoding module to withstand communication delay and the Multi-scale Window Attention Module to tackle GPS localization errors. [36] introduces a Feature Distribution-aware Aggregation framework to tackle the distribution gap in diverse independent private data used for training distinct agents in multi-agent perception. S2R-ViT [37] presents a Simulation-to-Reality transfer learning framework designed to mitigate the impacts of both deployment gap and feature gap on multi-agent cooperative perception between simulated and real-world data. [20] proposes the LC-aware Repair Network to overcome the missing data during the lossy communication.

III. V2X-DSI DATASET

LiDAR Beam Simulation. For analysis of large-scale deployments of LiDAR sensors on infrastructures, the real physical characteristics LiDAR sensors should be simulated in the powerful simulator. To achieve that, we use CARLA as our simulation framework, and utilize a latest open-source Realistic LiDAR Simulation (RLS) library [18] to simulate the LiDAR sensors. RLS library can simulate 14 popular LiDAR devices with their unique characteristics, including LiDAR beam, motion distortion, and ghosting object effect. Here, we select four frequently used type of LiDAR sensors from low-beam to high beam: Velodyne VLP-16, Velodyne VLP-32, Velodyne HDL-64, and Hesai Pandar128. We deploy these 4 LiDAR sensors on infrastructure on different scenario.

Sensor Configuration. To achieve the V2I communication and cooperative driving functionalities, we employ OpenCDA [38] to generate our dataset. It is featured with easy control of multiple CAVs, embedded vehicle network communication protocols, and more convenient and realistic traffic management. The data from eight existing towns in CARLA account is primary source of our dataset. On average, each frame contains around 3 interconnected vehicles, with a range from 2 to 7. Following [12], every CAV is outfitted with four cameras that collectively provide a 360-degree view, a 32-channel LiDAR, and GPS/IMU sensors.

Data Analysis. In total, we gathered 56,984 frames (time steps) comprising LiDAR point clouds. Our dataset incorporates six distinct road types, namely 4-way Intersection, T Intersection, Straight Segment, Curvy Segment, Roundabout, and Entrance Ramp. This selection aims to simulate the most prevalent driving scenarios encountered in real-life situations. Among them, T Intersection accounts for the highest proportion with 22 sets of data, including more than 20k LiDAR frames and 95k bounding boxes. The 4-way Intersection and Curvy Segment scenarios nearly share an equivalent number of data sets, comprising 11k and 15k LiDAR frames, respectively. Their respective bounding boxes exceed 52k and 47k. The Entrance Ramp combines partial data from the Straight Segment, Curvy Segment, and T Intersection, comprising 6.5k LiDAR frames and 20k bounding boxes.

Cooperative Perceptive System with Infrastructure. In this paper, we focus on the task of cooperative LIDAR-based 3D object detection on V2I perception system, considering a distribution gap among various levels of beam infrastructure LiDAR in real-world scenarios. The overall architecture of cooperative perceptive system with infrastructure comprises three modules: 1) V2I meta information exchanging, 2) LiDAR-based feature extraction, and 3) Infrastructure sensing feature sharing. In the stage of 1), V2I meta information such as relative poses, extrinsics, and agent type, are exchanged by each CAV in the communication network with other infrastructures. One of agents is selected as the ego vehicle to establish a spatial graph. In this graph, each node represents a CAV or an infrastructure, and edges denote directional V2X

communication channels among pairs of nodes. In the stage of 2), after all other infrastructures in proximity then project their LiDAR point clouds into the ego-vehicle's coordinate frame, LiDAR-based features are extracted from projected point clouds. In this scenario, the feature extractor can be utilized as the backbones of established 3D object detectors. In the stage of 3), the ego vehicle acquires infrastructure sensing feature maps from infrastructures nearby after feature extraction. These received intermediate features are then input into the remaining detection networks within the ego vehicle. In real-world settings, like challenging occlusions, inevitable interruptions resulting in degraded communication quality will be faced during the transmission of feature maps.

IV. Experiment

A. Experiments Setup

Evaluation metrics. Following [1], [12], [15], the performance evaluation primarily focuses on the 3D vehicle detection accuracy. The evaluation range is set as $x \in [-140, 140]$ meters, $y \in [-40, 40]$ meters, encompassing all CAVs within this spatial range throughout the experiment. We employ Average Precisions (AP) at Intersection-over-Union (IoU) thresholds of both 0.5 and 0.7 to measure the accuracy.

Experiment details. In this work, we aim to analyze the cooperative perception performance under low-beam infrastructure LiDAR. Our proposed V2X-DSI dataset contains four different type of infrastructure LiDAR: *1) 16-beam*, *2) 32-beam*, *3) 64-beam*, and *4) 128-beam*. Three state-of-the-art cooperative perception methods, *i.e.*, OPV2V [12], V2X-ViT [1], and CoBEVT [15] are utilized to train on these four different types of beam training set. Based on these, we assess the performance of models under a 16-beam LiDAR testing set with two settings: 1) These models are trained on three different high-beam LiDAR; 2) These high-beam trained models are finetuned based on the 16-beam LiDAR training set. We denote these setting that models are trained on a high-beam LiDAR training set and tested on a low-beam LiDAR testing set as *high-beam* $\rightarrow$ *low-beam* in our experiments.

Implementation Details. We employ the Adam optimizer with an initial learning rate of 10^{-3}, gradually decaying it every 10 epochs using a factor of 0.1. The hyperparameters align with those [1]. All models are trained using two RTX 3090 GPUs.

TABLE II
3D detection performance for four different beams LiDAR on V2X-DSI testing set.

Methods	16-beam		32-beam		64-beam		128-beam	
	AP0.5	AP0.7	AP0.5	AP0.7	AP0.5	AP0.7	AP0.5	AP0.7
NO Fusion	58.6	40.5	58.6	40.5	58.6	40.5	58.6	40.5
OPV2V [12]	70.9	52.5	69.3	54.6	74.2	59.7	79.0	67.5
V2X-ViT [1]	67.6	48.1	69.8	48.9	75.8	57.8	76.4	60.7
CoBVET [15]	**72.1**	**54.0**	**71.3**	**58.4**	**76.5**	**61.3**	**80.1**	**67.7**

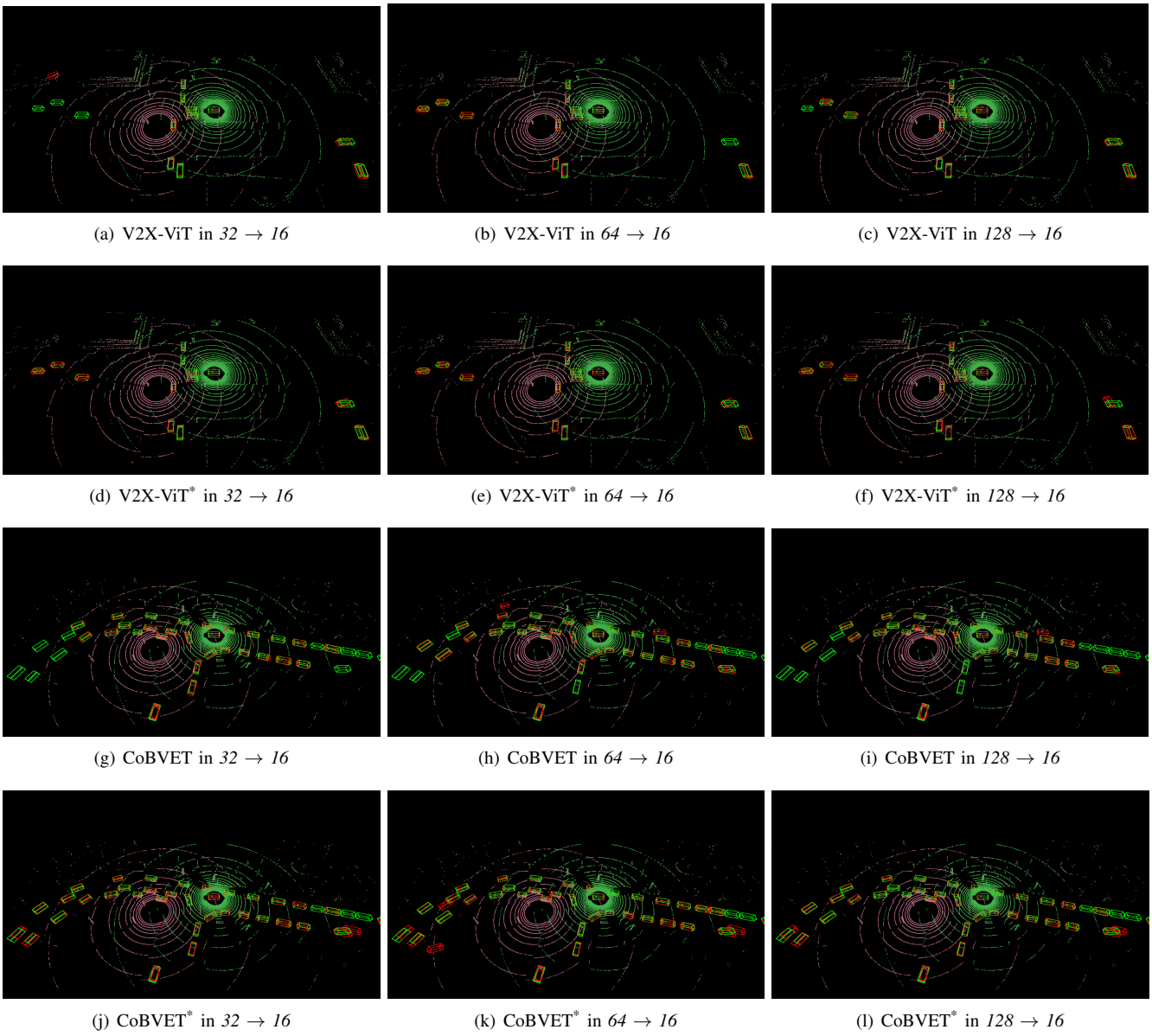
B. Quantitative Evaluation

Table II shows the 3D detection performance comparison on four different beam infrastructure LiDAR of V2X-DSI testing set, which contains three cooperative perception methods, *i.e.*, OPV2V [12], V2X-ViT [1] and CoBVET [15]. Without benefit of infrastructure LiDAR sensing, the *NO Fusion* only obtains 58.6%/40.5% for AP@0.5/0.7, While other three cooperative perception methods significantly surpass *NO Fusion* baseline. V2X-ViT, when tested on 64-beam infrastructure LiDAR, exhibits an 8.2%/9.7% higher accuracy for AP@0.5/0.7 compared to testing on 16-beam infrastructure LiDAR. Similarly, results on 128-beam infrastructure LiDAR also demonstrate a higher 8.8%/12.6% accuracy for AP@0.5/0.7 compared to 16-beam infrastructure LiDAR. Among the three models, CoBVET demonstrates the best performance, with significantly higher accuracy on both 64-beam and 128-beam infrastructure LiDAR compared to 16-beam infrastructure LiDAR. The observed trend indicates a substantial improvement in the accuracy of cooperative perception methods as the level of infrastructure LiDAR increases. Notably, CoBVET exhibits the highest increment from 16-beam to 128-beam infrastructure LiDAR with a 13.7% increase for AP@0.7, while V2X-ViT achieves a 12.6% increment for AP@0.7 from 16-beam to 128-beam infrastructure LiDAR. This highlights the clear correlation between the level of beam infrastructure LiDAR and the enhancement of available perception information, ultimately resulting in improved performance.

TABLE III
3D detection performance on V2X-DSI 16-beam LiDAR testing set. $*$ notes that pre-trained high-beam based models are fine-tuned on V2X-DSI 16-beam LiDAR training set before testing.

Methods	*32* $\rightarrow$ *16-beam*		*64* $\rightarrow$ *16-beam*		*128* $\rightarrow$ *16-beam*	
	AP0.5	AP0.7	AP0.5	AP0.7	AP0.5	AP0.7
NO Fusion	58.6	40.5	58.6	40.5	58.6	40.5
OPV2V [12]	64.7	49.8	65.2	50.4	63.5	49.3
V2X-ViT [1]	65.8	44.2	69.9	50.1	65.7	45.7
CoBVET [15]	65.6	52.8	69.1	52.2	63.8	49.2
OPV2V [12]*	72.7	56.9	73.8	57.6	73.9	57.9
V2X-ViT [1]*	74.7	58.5	**75.8**	**59.4**	72.6	56.3
CoBVET [15]*	**78.9**	**62.1**	75.7	59.0	**77.1**	**60.2**

Table III presents 3D object detection performance on V2X-DSI 16-beam infrastructure LiDAR testing set. Notably, all three models experience a substantial decrease in performance compared to the results shown in Table II. CoBVET exhibits the most significant drop among the three models, decreasing by 8.3% for AP@0.5 in the transition from *128* $\rightarrow$ *16-beam* infrastructure LiDAR, followed by OPV2V with a 7.4% drop for AP@0.5. These outcomes underscore the presence of a distribution gap when high-beam trained models are tested on 16-beam infrastructure LiDAR, posing a challenge and resulting in perceptual deficiencies in real-world driving scenarios. Obviously, after these high-beam trained models are fine-tuned on the 16-beam LiDAR training set, their perception accuracy is effectively improved. CoBVET achieves the highest increment on both *32* $\rightarrow$ *16-beam* and *128* $\rightarrow$ *16-beam* infrastructure LiDAR, showing a 13.3% improvement for AP@0.5 compared to the outcomes before fine-tuning. Following closely is V2X-ViT, with an incremental improvement of 10.6% for AP@0.7 on *128* $\rightarrow$ *16-beam*. Meanwhile, these three fine-tuned models



(a) V2X-ViT in *32* → *16* (b) V2X-ViT in *64* → *16* (c) V2X-ViT in *128* → *16*

(d) V2X-ViT* in *32* → *16* (e) V2X-ViT* in *64* → *16* (f) V2X-ViT* in *128* → *16*

(g) CoBVET in *32* → *16* (h) CoBVET in *64* → *16* (i) CoBVET in *128* → *16*

(j) CoBVET* in *32* → *16* (k) CoBVET* in *64* → *16* (l) CoBVET* in *128* → *16*

Fig. 3. **3D object detection visualization on V2X-DSI 16-beam LiDAR testing set.** Red violet point cloud represents CAV, and the pastel green color of point clouds indicates infrastructure. Green and red 3D bounding boxes correspond to the ground truth and prediction, respectively. $*$ notes that pre-trained high-beam based models are fine-tuned on V2X-DSI 16-beam LiDAR training set before testing. Best viewed in color.

also exhibit better performance than the 16-beam trained and tested models shown in Table II.

C. 3D Detection Visualization

To evaluate the influence of the distribution gap stemming from lower beamline numbers, we visually compare various methods in the scenario of the 16-beam LiDAR testing set. This comparison illustrates their effects on OPV2V [12], V2X-ViT [1], and CoBVET [15], as shown in Fig. 3. Under the *32-beam* → *16-beam*, *64-beam* → *16-beam*, and *128-beam* → *16-beam* settings, noticeable disparities are observed in features before and after fine-tuning. Fine-tuning on 16-beam setting effectively mitigates this gap, rendering the features more comparable to those obtained with high-beam infrastructure LiDAR. For V2X-ViT, as shown in (d), (e) and (f) of Fig. 3, its performance undergoes significant enhancement compared to the results (a), (b) and (c) tested before fine-tuning on 16-beam LiDAR setting. In the case of CoBEVT, our method enhances local details, capturing previously missed objects in (g), (h), and (i) of Fig. 3, thereby narrowing the gap with high-beam infrastructure LiDAR, as evident in (j), (k) and (l). This visual evidence underscores the effectiveness of fine-tuning.

V. Conclusions

In this paper, we propose V2X-DSI benchmark with 128-beam, 64-beam, 32-beam, and 16-beam infrastructure LiDAR sensors of 56,984 frames in 57 diverse urban scenarios for the V2I cooperative perception, overcoming the challenge posed by the expensive deployment of infrastructure LiDAR sensors in widespread V2X cooperative perception systems. Utilizing the introduced V2X-DSI benchmark, we evaluate the performance of three cooperative perception methods: OPV2V,

V2X-ViT, and CoBEVT. These models are initially trained on higher beam LiDAR from the V2X-DSI training set with improved performance. We then employ a fine-tuned approach to address the challenge that these models trained on high-beam data experience a performance drop when deployed in low-beam LiDAR scenarios. We anticipate that the V2X-DSI benchmark could draw increasing attentions, encouraging to investigate the economic V2X cooperative perception.

REFERENCES

[1] R. Xu, H. Xiang, Z. Tu, X. Xia, M.-H. Yang, and J. Ma, "V2x-vit: Vehicle-to-everything cooperative perception with vision transformer," in *European Conference on Computer Vision*, 2022, pp. 107–124.

[2] S. Han, F.-Y. Wang, G. Luo, L. Li, and F. Qu, "Parallel surface: Service-oriented v2x communications for autonomous vehicles," *IEEE Transactions on Intelligent Vehicles*, 2023.

[3] X. Liu, J. Li, J. Ma, H. Sun, Z. Xu, T. Zhang, and H. Yu, "Deep transfer learning for intelligent vehicle perception: a survey," *Green Energy and Intelligent Transportation*, p. 100125, 2023.

[4] C. Chang, J. Zhang, K. Zhang, W. Zhong, X. Peng, S. Li, and L. Li, "Bev-v2x: cooperative birds-eye-view fusion and grid occupancy prediction via v2x-based data sharing," *IEEE Transactions on Intelligent Vehicles*, 2023.

[5] Z. Meng, X. Xia, R. Xu, W. Liu, and J. Ma, "Hydro-3d: Hybrid object detection and tracking for cooperative perception using 3d lidar," *IEEE Transactions on Intelligent Vehicles*, 2023.

[6] E. Thonhofer, S. Sigl, M. Fischer, F. Heuer, A. Kuhn, J. Erhart, M. Harrer, and W. Schildorfer, "Infrastructure-based digital twins for cooperative, connected, automated driving and smart road services," *IEEE Open Journal of Intelligent Transportation Systems*, 2023.

[7] P. Zhang, D. Tian, J. Zhou, X. Duan, Z. Sheng, D. Zhao, and D. Cao, "Joint optimization of platoon control and resource scheduling in cooperative vehicle-infrastructure system," *IEEE Transactions on Intelligent Vehicles*, 2023.

[8] H. Yu, X. Liu, Y. Tian, Y. Wang, C. Gou, and F.-Y. Wang, "Sora-based parallel vision for smart sensing of intelligent vehicles: from foundation models to foundation intelligence," *IEEE Transactions on Intelligent Vehicles*, 2024.

[9] H. Yu, Y. Luo, M. Shu, Y. Huo, Z. Yang, Y. Shi, Z. Guo, H. Li, X. Hu, J. Yuan, and Z. Nie, "Dair-v2x: A large-scale dataset for vehicle-infrastructure cooperative 3d object detection," in *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022, pp. 21 361–21 370.

[10] "New york city dot infrastructure," https://www.nyc.gov/html/dot/html/infrastructure/infrastructure.shtml, accessed: 2023-06-20.

[11] A. Dosovitskiy, G. Ros, F. Codevilla, A. Lopez, and V. Koltun, "CARLA: An open urban driving simulator," in *1st Annual Conference on Robot Learning*, 2017, pp. 1–16.

[12] R. Xu, H. Xiang, X. Xia, X. Han, J. Li, and J. Ma, "Opv2v: An open benchmark dataset and fusion pipeline for perception with vehicle-to-vehicle communication," in *International Conference on Robotics and Automation*. IEEE, 2022, pp. 2583–2589.

[13] R. Xu, X. Xia, J. Li, H. Li, S. Zhang, Z. Tu, Z. Meng, H. Xiang, X. Dong, R. Song *et al.*, "V2v4real: A real-world large-scale dataset for vehicle-to-vehicle cooperative perception," in *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2023, pp. 13 712–13 722.

[14] H. Yu, Y. Luo, M. Shu, Y. Huo, Z. Yang, Y. Shi, Z. Guo, H. Li, X. Hu, J. Yuan *et al.*, "Dair-v2x: A large-scale dataset for vehicle-infrastructure cooperative 3d object detection," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2022, pp. 21 361–21 370.

[15] R. Xu, Z. Tu, H. Xiang, W. Shao, B. Zhou, and J. Ma, "Cobevt: Cooperative bird's eye view semantic segmentation with sparse transformers," in *Conference on Robot Learning*. PMLR, 2023, pp. 989–1000.

[16] Y. Li, D. Ma, Z. An, Z. Wang, Y. Zhong, S. Chen, and C. Feng, "V2x-sim: Multi-agent collaborative perception dataset and benchmark for autonomous driving," *IEEE Robotics and Automation Letters*, vol. 7, no. 4, pp. 10 914–10 921, 2022.

[17] H. Yu, W. Yang, H. Ruan, Z. Yang, Y. Tang, X. Gao, X. Hao, Y. Shi, Y. Pan, N. Sun *et al.*, "V2x-seq: A large-scale sequential dataset for vehicle-infrastructure cooperative perception and forecasting," in *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2023, pp. 5486–5495.

[18] X. Cai, W. Jiang, R. Xu, W. Zhao, J. Ma, S. Liu, and Y. Li, "Analyzing infrastructure lidar placement with realistic lidar simulation library," in *IEEE International Conference on Robotics and Automation*. IEEE, 2023, pp. 5581–5587.

[19] J. Zhang, J. Ge, S. Li, S. Li, and L. Li, "A bi-level network-wide cooperative driving approach including deep reinforcement learning-based routing," *IEEE Transactions on Intelligent Vehicles*, 2023.

[20] J. Li, R. Xu, X. Liu, J. Ma, Z. Chi, J. Ma, and H. Yu, "Learning for vehicle-to-vehicle cooperative perception under lossy communication," *IEEE Transactions on Intelligent Vehicles*, 2023.

[21] K. Yang, D. Yang, J. Zhang, M. Li, Y. Liu, J. Liu, H. Wang, P. Sun, and L. Song, "Spatio-temporal domain awareness for multi-agent collaborative perception," in *IEEE/CVF International Conference on Computer Vision*, 2023, pp. 23 383–23 392.

[22] A. Dosovitskiy, G. Ros, F. Codevilla, A. Lopez, and V. Koltun, "Carla: An open urban driving simulator," in *Annual Conference on Robot Learning*. PMLR, 2017, pp. 1–16.

[23] R. Chen, J. Ning, Y. Lei, Y. Hui, and N. Cheng, "Mixed traffic flow state detection: A connected vehicles assisted roadside radar and video data fusion scheme," *IEEE Open Journal of Intelligent Transportation Systems*, 2023.

[24] E. Andreotti, M. Aramrattana *et al.*, "Cooperative merging strategy between connected autonomous vehicles in mixed traffic," *IEEE Open Journal of Intelligent Transportation Systems*, vol. 3, pp. 825–837, 2022.

[25] C. Pilz, P. Sammer, E. Piri, U. Grossschedl, G. Steinbauer-Wagner, L. Kuschnig, A. Steinberger, and M. Schratter, "Collective perception: A delay evaluation with a short discussion on channel load," *IEEE Open Journal of Intelligent Transportation Systems*, 2023.

[26] R. Valiente, B. Toghi, R. Pedarsani, and Y. P. Fallah, "Robustness and adaptability of reinforcement learning-based cooperative autonomous driving in mixed-autonomy traffic," *IEEE Open Journal of Intelligent Transportation Systems*, vol. 3, pp. 397–410, 2022.

[27] X. Dai, M. Vallati, R. Guo, Y. Wang, S. Han, and Y. Lin, "The road ahead: Dao-secured v2x infrastructures for safe and smart vehicular management," *IEEE Transactions on Intelligent Vehicles*, 2023.

[28] C. Liu, J. Chen, Y. Chen, R. Payton, M. Riley, and S.-H. Yang, "Self-supervised adaptive weighting for cooperative perception in v2v communications," *IEEE Transactions on Intelligent Vehicles*, 2023.

[29] R. Xu, J. Li, X. Dong, H. Yu, and J. Ma, "Bridging the domain gap for multi-agent perception," in *IEEE International Conference on Robotics and Automation*. IEEE, 2023, pp. 6035–6042.

[30] J. Li, R. Xu, J. Ma, Q. Zou, J. Ma, and H. Yu, "Domain adaptive object detection for autonomous driving under foggy weather," in *IEEE Winter Conference on Applications of Computer Vision*, 2023, pp. 612–622.

[31] G. Elghazaly, R. Frank, S. Harvey, and S. Safko, "High-definition maps: Comprehensive survey, challenges and future perspectives," *IEEE Open Journal of Intelligent Transportation Systems*, 2023.

[32] A. Kloukiniotis, A. Papandreou, A. Lalos, P. Kapsalas, D.-V. Nguyen, and K. Moustakas, "Countering adversarial attacks on autonomous vehicles using denoising techniques: A review," *IEEE Open Journal of Intelligent Transportation Systems*, vol. 3, pp. 61–80, 2022.

[33] J. Lee, G. Bang, T. Nishimori, K. Nakao, and S. Kamijo, "Radar translation network between sunny and rainy domains by combination of kp-convolution and cyclegan," *IEEE Open Journal of Intelligent Transportation Systems*, vol. 4, pp. 833–845, 2023.

[34] J. Li, B. Li, X. Liu, J. Fang, F. Juefei-Xu, Q. Guo, and H. Yu, "Advgps: Adversarial gps for multi-agent perception attack," in *IEEE International Conference on Robotics and Automation*, 2024.

[35] B. Li, J. Li, X. Liu, R. Xu, Z. Tu, J. Guo, X. Li, and H. Yu, "V2x-dgw: Domain generalization for multi-agent perception under adverse weather conditions," *arXiv preprint arXiv:2403.11371*, 2024.

[36] J. Li, B. Li, X. Liu, R. Xu, J. Ma, and H. Yu, "Breaking data silos: Cross-domain learning for multi-agent perception from independent private sources," in *IEEE International Conference on Robotics and Automation*, 2024.

[37] J. Li, R. Xu, X. Liu, B. Li, Q. Zou, J. Ma, and H. Yu, "S2r-vit for multi-agent cooperative perception: Bridging the gap from simulation to reality," in *IEEE International Conference on Robotics and Automation*, 2024.

[38] R. Xu, Y. Guo, X. Han, X. Xia, H. Xiang, and J. Ma, "Opencda: an open cooperative driving automation framework integrated with co-simulation," in *IEEE International Intelligent Transportation Systems Conference*. IEEE, 2021, pp. 1155–1162.