2024 IEEE International Conference on Robotics and Automation (ICRA 2024)
May 13-17, 2024. Yokohama, Japan

HPL-ViT: A Unified Perception Framework for Heterogeneous Parallel LiDARs in V2V

Yuhang Liu[1], Boyi Sun[2], Yuke Li[3], Yuzheng Hu[4], Fei-Yue Wang[5]

***Abstract*— To develop the next generation of intelligent LiDARs, we propose a novel framework of parallel LiDARs and construct a hardware prototype in our experimental platform, DAWN (Digital Artificial World for Natural). It emphasizes the tight integration of physical and digital space in LiDAR systems, with networking being one of its supported core features. In the context of autonomous driving, V2V (Vehicle-to-Vehicle) technology enables efficient information sharing between different agents which significantly promotes the development of LiDAR networks. However, current research operates under an ideal situation where all vehicles are equipped with identical LiDAR, ignoring the diversity of LiDAR categories and operating frequencies. In this paper, we first utilize OpenCDA and RLS (Realistic LiDAR Simulation) to construct a novel heterogeneous LiDAR dataset named OPV2V-HPL. Additionally, we present HPL-ViT, a pioneering architecture designed for robust feature fusion in heterogeneous and dynamic scenarios. It uses a graph-attention Transformer to extract domain-specific features for each agent, coupled with a cross-attention mechanism for the final fusion. Extensive experiments on OPV2V-HPL demonstrate that HPL-ViT achieves SOTA (state-of-the-art) performance in all settings and exhibits outstanding generalization capabilities.**

I. INTRODUCTION

LiDAR sensor plays a crucial role in vehicle perception systems which enable the acquisition of 3D structural information. Fueled by advancements in both artificial intelligence and communication technologies, LiDAR systems are rapidly evolving towards digitization, networking, and increased intelligence [1]. In this vein, we propose a novel parallel LiDAR framework inspired by the principles of parallel intelligence [2], [3]. It leverages software systems to efficiently enhance the sensing capabilities of physical LiDARs, highlighting the transformative potential of the digital space within LiDAR systems. Notably, we have developed a hardware prototype of parallel LiDARs based on the DAWN parallel sensing platform [4]. DAWN, short for "Digital Artificial World for Natural", serves as a comprehensive platform for developing next-generation intelligent sensors. This paper delves into the networking capabilities of parallel LiDARs, specifically focusing on their application in autonomous driving scenarios.

This work was supported by the Joint Development of Multi-modal Parallel LiDARs with Waytous Inc..

[1]Yuhang Liu, and Fei-Yue Wang are with the Institute of Automation, Chinese Academy of Science, Beijing, China. liuyuhang2021@ia.ac.cn, feiyue.wang@ia.ac.cn

[2]Boyi Sun is with the Department of Computer Science and Technology, University of Chinese Academy of Sciences, Beijing, China. sunboyi20@mails.ucas.ac.cn

[3]Yuke Li is with Waytous Inc., Qingdao, China. liyuke14@mails.ucas.ac.cn

[4]Yuzheng Hu is with the Department of Computer Science, University of Illinois, Urbana-Champaign, IL, USA. yh46@illinois.edu

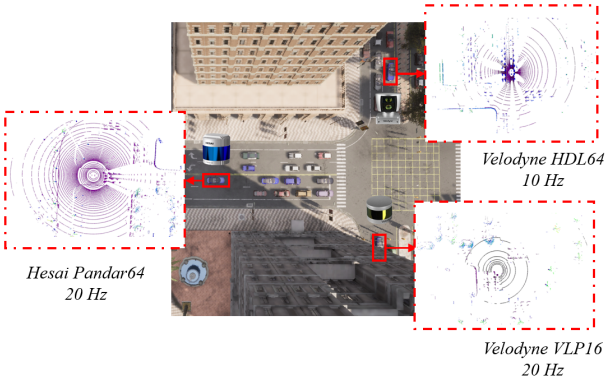


Fig. 1. A heterogeneous scenario with different LiDARs.

V2V (Vehicle-to-Vehicle) technology fosters efficient data exchange between autonomous vehicles, demonstrably enhancing their perception capabilities [5], [6], [7]. While sharing raw data [8] or detection results [9], [10] has been explored, intermediate feature sharing has emerged as the dominant approach in V2V cooperative perception [11], [12], [13]. It strikes a balance between achieving high perception accuracy and preserving valuable communication bandwidth. Currently, several datasets have been established for evaluating fusion methods in V2V, such as OPV2V [14] collected in the CARLA simulator [15] and V2V4Real [16] gathered from real-world scenarios. However, these datasets make the simplifying assumption that all agents are equipped with identical LiDARs, which deviates significantly from the diversity encountered in real-world deployments. As illustrated in Figure 1, different LiDAR systems are frequently employed within the same scenario. Moreover, individual vehicles often possess the flexibility to adjust their LiDARs' operating frequency, further introducing frequency heterogeneity. This disparity is expected to be amplified by the advent of software-defined adaptive parallel LiDARs.

This work tackles the challenge of achieving robust cooperative perception among heterogeneous parallel LiDARs in V2V, aiming to minimize model performance degradation in diverse and complex scenarios. To facilitate the investigation, we introduce OPV2V-HPL, a new dataset constructed in the CARLA simulator using OpenCDA [17] and RLS (Realistic LiDAR Simulation) [18]. It overcomes the limitations of existing options by incorporating diverse LiDAR sensor models operating at different frequencies, reflecting the complexity of real-world environments. Specifically, we replay OPV2V scenes and collect data using four high-fidelity LiDAR

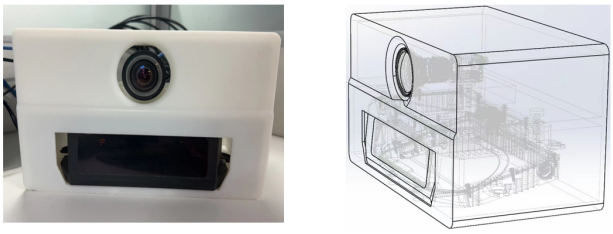


Fig. 2. Hardware prototype of parallel LiDARs.

models from RLS at two distinct operating frequencies. Furthermore, we present HPL-ViT (**H**eterogeneous **P**arallel **L**iDARs-**Vi**sion **T**ransformer), a novel framework for feature fusion among various LiDARs. It effectively incorporates prior information regarding LiDAR categories and operating frequencies to optimize its performance. Vehicles initially generate BEV (Bird's Eye View) feature maps and share them with connected agents. Received feature maps are then fused through HPL-ViT, which leverages multi-scale graph-attention and bi-directional cross-attention architectures for robust information aggregation. Extensive experiments on OPV2V-HPL demonstrate HPL-ViT's consistent achievement of SOTA (State-Of-The-Art) results across various experimental settings. Notably, it exhibits significant performance improvements in scenarios with greater heterogeneity, showcasing its ability to effectively handle diverse sensor configurations. Besides, HPL-ViT presents impressive generalizability in dynamic scenarios and with varying ego LiDARs, delivering a performance improvement of at least 3.3% over other fusion methods. The main contribution of this paper can be summarized as follows:

- To the best of our knowledge, this is the pioneering work that explores the impact of heterogeneous LiDAR systems on V2V. We utilize the combined power of OpenCDA and RLS to construct OPV2V-HPL, a next-generation dataset specifically designed to reflect real-world sensor diversity.
- We propose the innovative HPL-ViT framework to improve feature interaction among diverse LiDARs. Our approach incorporates both category and frequency information into graph attention computations without imposing a significant increase in communication bandwidth. Then a bi-directional cross-attention mechanism is introduced to merge features from both category and frequency branches.
- Each module in HPL-ViT can be seamlessly integrated into other methods to further improve perception performance. We will soon release our dataset and codes.

II. Related Work

A. Parallel LiDARs

Parallel LiDARs represent a novel category of intelligent LiDARs founded on parallel intelligence [2]. Parallel intelligence is a pioneering methodological framework introduced by Prof. Wang, which emphasizes the tight integration of physical and digital realms [19], [20]. It has already found extensive applications across various domains, including control [21], [22], [23], sensing [24], [25], [26], and autonomous driving [27], [28]. Within the realm of parallel sensing, the DAWN platform [4] facilitates the development of next-generation intelligent sensors like LiDARs and light field systems [29], [30]. [2] proposes the framework of parallel LiDARs which consists of three primary parts: descriptive, predictive, and prescriptive LiDARs. Descriptive LiDARs aim to construct complete digital LiDAR systems; predictive LiDARs highlight the significance of computational experiments in cyberspace; while prescriptive LiDARs enable real-time interaction between physical and digital worlds. As shown in Figure 2, a hardware prototype of parallel LiDAR within DAWN exemplifies this concept [4]. It leverages software systems to reconfigure hardware operations, enabling real-time optimization of perceptual resource allocation. This article focuses on the robust cooperative perception of heterogeneous parallel LiDARs in autonomous driving.

B. LiDAR-based 3D Object Detection

LiDAR sensors can provide accurate depth and structural information in autonomous driving which can be used for scene understanding and object identification. According to different data representations, LiDAR-based 3D object detection methods fall into four main categories [31], [32]:

- Point-based methods are specifically designed for point clouds' unique structure which directly extract features at the individual point level. PointNet [33], [34] stands out as a pioneering work in point-based methods that utilizes MLP for efficient point-wise feature capture, while subsequent works explore optimizing feature extraction with graph operators [35] or Transformers [36].
- Grid-based methods first partition point clouds into regular 2D or 3D voxels, then extract features using standard 2D vision convolutional networks. VoxelNet [37] pioneered 3D CNN feature generation, while SECOND [38] introduced sparse convolution for efficient computation. PointPillars [39] projects point clouds into a BEV perspective and utilizes 2D CNNs, offering significant speed advantages.
- Point-voxel-based methods use a hybrid architecture to extract features at both point and voxel levels. PV-RCNN [40] is a typical model that can learn features from different data representations at each stage.
- Range-based methods convert point clouds into an image-like format which can be regarded as a sparse depth map recording range information [41]. Advanced 2D object detection approaches can be directly used for feature extraction, facilitating multimodal data fusion.

Prioritizing real-time performance, we opted for PointPillars for intermediate feature extraction in our study.

C. Cooperative Perception

Cooperative perception utilizes V2X (Vehicle-to-Everything) technologies to enhance a vehicle's perception capabilities [42]. Recent datasets like OPV2V [14] and V2XSet [11] acquired in CARLA, as well as V2V4Real

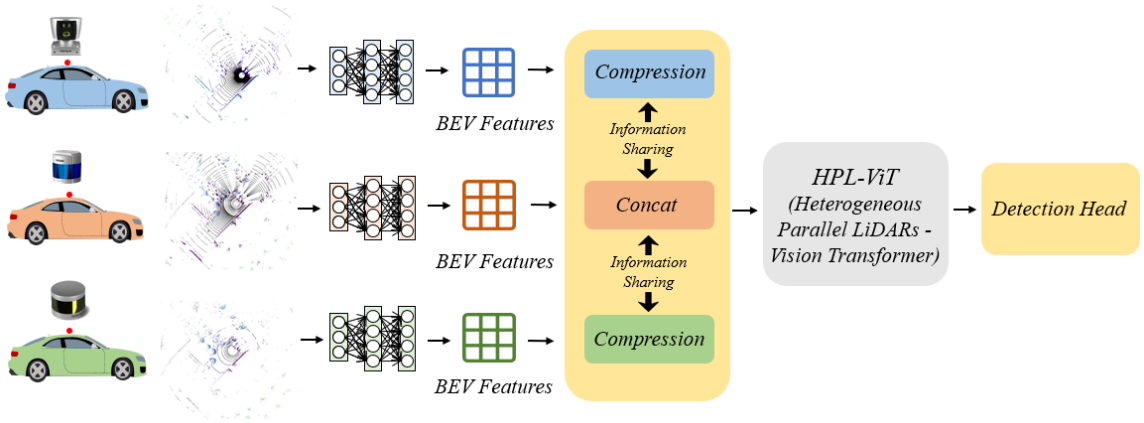


Fig. 3. Cooperative perception of heterogeneous parallel LiDARs in V2V.

[16] and DAIR-V2X [43] captured in physical settings, have significantly fueled progress in this field. Current V2V research focuses on LiDAR-based 3D object detection, which can be categorized into three main approaches: early, intermediate, and late fusion. Early fusion directly transmits raw point cloud data [8], while late fusion only shares generated 3D bounding boxes [9]. Intermediate fusion [11], [12], [13], [14] extracts and shares neural features for fusion, achieving a balance between bandwidth and model accuracy. V2VNet [44] is a pioneering work that proposes a spatial-aware mechanism for compressed feature map sharing. DiscoNet [45] adopts a graph-based structure to fuse features, and [14] suggests a novel single-head self-attention method. MPDA [46] investigates the impact of heterogeneous models, employing a learnable resizer module to align feature maps of varying sizes and a domain classifier for domain invariant feature extraction. Additionally, [47] considers the difference between LiDARs and cameras in cooperative perception. In contrast to previous works, we focus on the challenge of heterogeneous LiDARs in V2V and propose a novel HPL-ViT framework to improve feature fusion performance.

III. Methodology

This work tackles a more realistic V2V scenario in autonomous driving, where vehicles are equipped with diverse LiDAR sensors operating at various frequencies. Each vehicle communicates with its surroundings, and our focus lies on enhancing LiDAR-based 3D object detection through cooperative perception. As illustrated in Figure 3, our proposed framework consists of four key modules: feature extraction, data compression and sharing, HPL-ViT for fusion, and a detection head.

A. Main Architecture

1) Feature Extraction: For efficient inference, we leverage the grid-based PointPillars to extract intermediate features from the raw point cloud. It initially partitions the point cloud into individual pillars and generates a pseudo-image in the BEV perspective. Then a multi-scale CNN backbone extracts BEV feature maps denoted as $F(i) \in R^{H \times W \times C}$, where i represents the agent index and H, W, and C denote the height, width, and number of channels, respectively.

2) Data Compression and Sharing: To reduce communication bandwidth, we utilize 1×1 convolution kernels to compress the feature map into $F(i)^{'} \in R^{H \times W \times C_0}$ with $C_0 < C$. These compressed feature maps are then transmitted to the ego vehicle for feature fusion. Additionally, each agent shares information regarding their LiDAR category and operating frequency. Given our focus on LiDAR heterogeneity, this work does not consider factors like position errors or communication delays during data transmission.

3) HPL-ViT: HPL-ViT is a novel vision transformer architecture that takes concatenated feature maps as input. Figure 4 provides an overview of the HPL-ViT framework. Our approach begins by applying multi-scale heterogeneous graph-attention mechanisms to enhance feature interactions in both the category and frequency domains. After aligning features across scales, we feed them into a bidirectional cross-attention module for fusion. Despite its multi-scale nature, HPL-ViT maintains consistent dimensions for input and output feature maps, avoiding the loss of fine-grained details due to downsampling.

4) Detection Head: We utilize two 1×1 convolution layers as the detection head, with one responsible for bounding box regression and the other for classification.

B. HPL-ViT

1) Multi-scale Heterogeneous Graph-attention: Intermediate features extracted from different LiDARs exhibit distinctive characteristics due to their inherent variations in category and operating frequency. To address this challenge, we propose a multi-scale heterogeneous graph-attention to capture domain-specific information. Figure 5(a) illustrates a single heterogeneous graph-attention block which comprises a graph-attention and a local self-attention layer. The graph-attention layer enhances feature interactions at the same spatial location across different feature maps. Additionally, we introduce multi-scale feature computation to extract richer semantic information at different granularities.

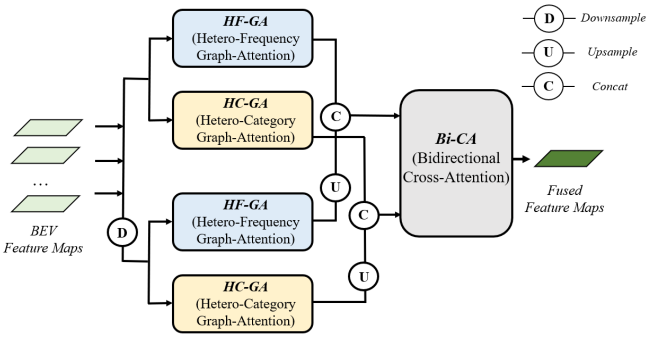


Fig. 4. The framework of HPL-ViT.

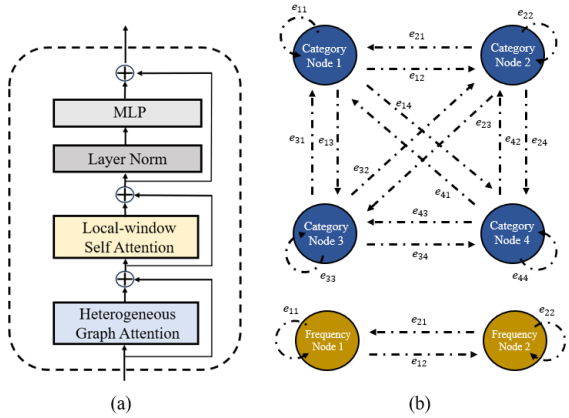


Fig. 5. (a). A heterogeneous graph-attention block. (b). The constructed heterogeneous graphs.

Figure 5(b) depicts two heterogeneous graphs, each with nodes representing individual vehicles. We focus on the category-based graph for explanation. Each node i stores its compressed BEV feature map $F(i)^{'}$ and installed LiDAR category $C(i)$. E_{ij} represents the edge connecting nodes i and j. Although we use a multi-head attention mechanism, we only present a single-head attention scenario in the description for clarity. Multi-head attention can be easily achieved by concatenating features from multiple heads. Firstly, we utilize liner layers to generate query $Q(i)$, key $K(i)$, and value $V(i)$ vectors from $F(i)^{'}$. Furthermore, we define two sets of learnable parameters, $W_{e_{ij}}^{Att}$ and $W_{e_{ij}}^{v}$, which are used for weighting attention maps and values:

$$Att_x(i,j) = softmax(Q_x(i)W_{e_{ij}}^{Att}K_x(j)), \quad (1)$$

$$Msg_x(i,j) = V_x(j)W_{e_{ij}}^{v}, \quad (2)$$

$x \in R^2$ denotes the spatial location in feature maps. Following that, we finalize the feature update for node i:

$$G_x(i) = \sum_{j \in N(i)} Att_x(i,j)Msg_x(i,j), \quad (3)$$

$N(i)$ denotes the neighborhood of node i, encompassing all other vehicles participating in data interaction with it. We then incorporate a local self-attention layer to promote further feature interaction within each node. Additionally,

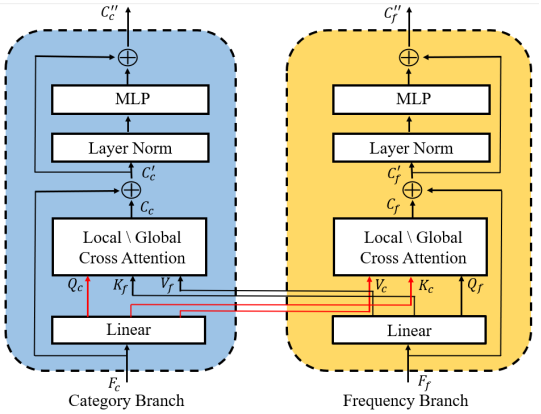


Fig. 6. A bidirectional cross-attention module.

we introduce multi-scale features to capture information at different granularity levels. This is achieved by downsampling the input features and aligning them through inverse convolution. We independently concatenate output features from the frequency and category branches to generate $F_c(i)$ and $F_f(i)$, respectively.

2) Bidirectional Cross-Attention: A novel bidirectional architecture stacked with multiple local and global cross-attention blocks is proposed to integrate features from different branches. Figure 6 presents the structure of a single cross-attention module.

We begin by aggregating multi-vehicle features into F_c and $F_f \in R^{N \times H \times W \times C}$, where N represents the count of connected vehicles. These tensors are then partitioned into $(\frac{H}{m}, \frac{W}{m}, N \times m^2, C)$, where each window includes Nm^2 tokens for feature interaction. m denotes the window size. The cross-attention module adopts distinct linear layers to generate queries, keys, and values for each branch. The query in the current branch is designed to access the key and value components of another branch. Following this, the fused feature is processed through an MLP layer with skip connections, ensuring robust gradient backpropagation:

$$C_a = softmax(\frac{Q_a K_b}{\sqrt{d_c}})V_b, \quad (4)$$

$$C_a^{'} = C_a + F_a, \quad (5)$$

$$C_a^{"} = MLP(LN(C_a^{'})) + C_a^{'}, \quad (6)$$

The pairs (a, b) corresponds to either $(category, frequency)$ or $(frequency, category)$, with d_c representing the feature length. Q, K, and V with subscripts denote the query, key, and value for their respective branches. To enhance the receptive field, we adopt an axis-swapping strategy [48], [49] for extracting global semantic information without introducing additional complexity. Local attention efficiently captures object-level details, while global attention uses discrete sampling to aggregate scene-level global information. We also utilize multi-head attention in our actual implementation.

IV. Experiments

A. Dataset

Our experiments rely on OPV2V-HPL, which is an enhanced iteration of OPV2V. OPV2V [14] is the first large-scale V2V dataset collected in CARLA, comprising a total of 11,464 data frames. It's split into training, validation, and test sets with 6,764, 1,981, and 2,719 frames, respectively. Each frame incorporates a combination of point cloud and image data captured using CARLA's default sensor models. To explore the impact of heterogeneous LiDARs, we leverage RLS [18], a high-fidelity LiDAR library mirroring real-world sensor parameters. We select four different mechanical LiDARs from RLS and replay OPV2V scenarios in OpenCDA for generating OPV2V-HPL. These LiDARs include Hesai Pandar64, Velodyne HDL64, VLP32, and VLP16 models, each characterized by varying beam counts and pitch angle distributions. It allows for the accurate reconstruction of complex LiDAR interactions in autonomous driving. Furthermore, we consider two different operating frequencies, 10 Hz and 20 Hz, for even greater realism.

B. Experimental Setup

1) Evaluation Metrics: We use 3D object detection accuracy as the metric to evaluate different fusion methods. To be more specific, we compute the AP (Average Precision) at IoU (Intersection-over-Union) thresholds of 0.5 and 0.7. The detection range in our experiments is defined as x$\in [-140.8, 140.8]$, and y$\in [-38.4, 38.4]$.

2) Implementation Details: PointPillars is utilized to extract intermediate feature maps from raw point cloud data, which are then fed into HPL-ViT for fusion. We compare our approach with six other fusion methods, including no fusion baseline, OPV2V [14], V2VNet [44], F-Cooper [50], CoBEVT [49], and late fusion. All models are trained for 30 epochs using 8 Nvidia V100 GPUs with the Adam optimizer. The initial learning rate is configured at 0.001, and we employ the cosine annealing with a warm-up strategy to dynamically adjust the learning rate. Our loss function comprises two primary components: a smooth L1 regression loss with a coefficient of 1, and a focal classification loss with a coefficient of 2.

3) Experimental Scenarios: We evaluate the proposed method under three distinct scenarios, assuming all ego vehicles are equipped with 20Hz Pandar64 LiDARs:

- **Normal scenario**: In the normal scenario, all surrounding vehicles are equipped with 20Hz Pandar64 LiDARs, matching the LiDAR configuration of the ego vehicle.
- **Hetero scenario 1**: In hetero scenario 1, surrounding vehicles utilize diverse LiDAR types chosen randomly from four LiDAR models. While the LiDAR categories vary, all operate at a consistent frequency of 20 Hz.
- **Hetero scenario 2**: Hetero scenario 2 presents a more intricate setting by introducing both diversity in LiDAR types and variations in operating frequencies.

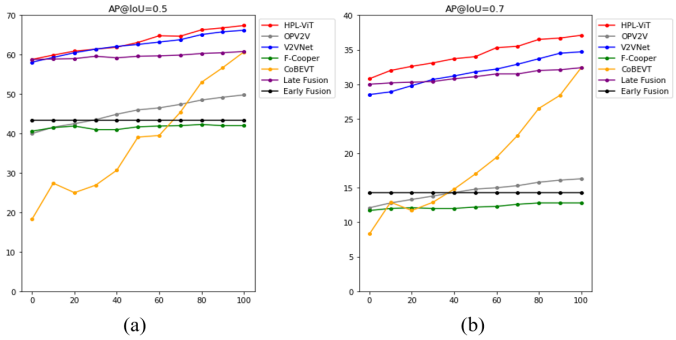


Fig. 7. Generalization performance in dynamic scenarios.

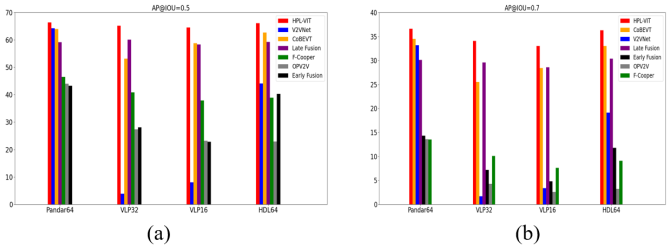
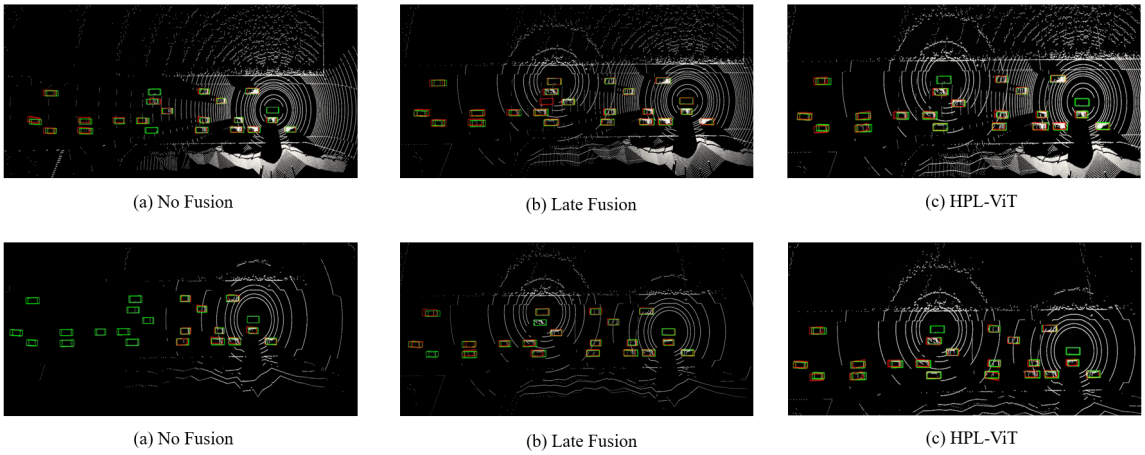


Fig. 8. Generalization performance with varying ego LiDARs.

C. Quantitative Evaluation

1) Main Performance Analysis: Table I presents the perceptual performance of all methods under different settings. We observe that most fusion methods outperform the no-fusion baseline, consistently demonstrating the benefits of collaborative perception. HPL-ViT achieves SOTA results in all scenarios, showcasing its effectiveness in diverse environments. It's crucial to note that introducing heterogeneous scenarios negatively impacts all models, with HPL-ViT exhibiting the least performance degradation. As LiDAR heterogeneity increases, we find a widening accuracy gap between HPL-ViT and the second-ranked method. It can reach a significant 2.1% improvement in hetero scenario 2, which effectively demonstrates its outstanding feature fusion capabilities in complex and challenging environments.

2) Generalization Analysis: To assess the generalizability of HPL-ViT, we test models trained in the most challenging hetero scenario 2 on a wider range of settings. First, we fix the ego LiDAR as a 20 Hz Pandar64 and vary the LiDAR configurations of surrounding vehicles. As illustrated in Figure 7, the horizontal axis represents the percentage of other vehicles equipped with the same 20 Hz Pandar64, and it is evident that HPL-ViT consistently achieves SOTA performance across all settings. It's worth noting that late fusion has outperformed most intermediate fusion methods in our experiments. Additionally, we introduce a change in the type of ego LiDAR to further assess models, with results plotted in Figure 8. While many methods experience a drop in accuracy, HPL-ViT maintains its top position, demonstrating remarkable adaptability. It exhibits an accuracy improvement of at least 3.3% when deployed on different LiDAR devices, highlighting its robustness to varying sensor configurations.



(a) No Fusion (b) Late Fusion (c) HPL-ViT

(a) No Fusion (b) Late Fusion (c) HPL-ViT

Fig. 9. 3D object detection visualization. Green and red bounding boxes represent the ground truth and predicted, respectively. (a)-(c) are visualization results in Hetero scenario 2 with Pandar64 as the ego LiDAR, while (d)-(e) use VLP16 as the ego LiDAR.

TABLE I

3D OBJECT DETECTION ACCURACY IN DIFFERENT SCENARIOS.

	Normal scenario		Hetero scenario 1		Hetero scenario 2	
Methods	**AP@0.5**	**AP@0.7**	**AP@0.5**	**AP@0.7**	**AP@0.5**	**AP@0.7**
No Fusion	43.3	14.3	43.3	14.3	43.3	14.3
Late Fusion	64.6	37.3	57.5	29.2	59.2	30.1
OPV2V [14]	51.1	18.2	48.1	16.0	44.0	13.6
V2VNet [44]	66.3	37.3	64.8	35.3	64.3	33.2
F-Cooper [50]	49.3	19.3	48.3	15.4	46.5	13.5
CoBEVT [49]	66.7	37.8	64.4	35.4	64.0	34.5
HPL-ViT	**67.3 (+0.6)**	**38.3 (+0.5)**	**65.9 (+1.1)**	**36.4 (+1.0)**	**66.4 (+2.1)**	**36.6 (+2.1)**

3) Ablation Study: We conduct comprehensive ablation studies to evaluate the individual contributions of HPL-ViT's components. As depicted in Table II, every component contributes positively to object detection accuracy. The absence of HC-GA or HF-GA leads to a similar performance decrease, while the omission of multi-scale strategy results in the most significant 2.3% drop in AP@0.7.

TABLE II

ABLATION STUDIES OF EACH COMPONENT IN HPL-VIT.

HC-GA	HF-GA	Bi-CA	Multi-scale	AP@0.5 / AP@0.7
	✓	✓	✓	65.2 / 36.2
✓		✓	✓	65.2 / 35.7
✓	✓		✓	65.4 / 35.4
✓	✓	✓		64.1 / 34.1
✓	✓	✓	✓	66.4 / 36.6

D. Qualitative Evaluation

In Figure 9 (a)-(c), we provide visualizations of no fusion, late fusion, and HPL-ViT in hetero scenario 2, respectively. Then we change the ego LiDAR to a VLP16 operating at 20Hz to present their generalization capabilities, and the testing results are plotted in Figure 9 (d)-(f). The no-fusion baseline displays numerous missed detections, highlighting the need for information sharing between vehicles. While late fusion improves detection performance, it still struggles with missed objects and false positives. In contrast, HPL-ViT demonstrates superior generalization. By effectively fusing diverse LiDAR data, it not only eliminates false positives but also achieves more precise bounding box positions.

V. CONCLUSIONS

This paper delves into cooperative perception for autonomous driving to facilitate the construction of parallel LiDAR networks. As the first work to tackle the challenge of heterogeneous LiDARs in V2V, we collect a novel dataset named OPV2V-HPL and present HPL-ViT, which is a vision Transformer architecture designed for robust feature fusion within LiDAR networks. Extensive experiments demonstrate that HPL-ViT can attain SOTA performance with strong generalizability across various scenarios. In future work, we aim to extend our research into real-world deployments and remain committed to the development of robust and reliable LiDAR networks.

REFERENCES

[1] R. Roriz, J. Cabral, and T. Gomes, "Automotive lidar technology: A survey," *IEEE Transactions on Intelligent Transportation Systems*, vol. 23, no. 7, pp. 6282–6297, 2022.

[2] Y. Liu, Y. Shen, L. Fan, Y. Tian, Y. Ai, B. Tian, Z. Liu, and F.-Y. Wang, "Parallel radars: from digital twins to digital intelligence for smart radar systems," *Sensors*, vol. 22, no. 24, p. 9930, 2022.

[3] Y. Liu, Y. Shen, Y. Tian, Y. Ai, B. Tian, E. Wu, and L. Chen, "Radarverses in metaverses: A cpsi-based architecture for 6s radar systems in cpss," *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. 53, no. 4, pp. 2128–2137, 2022.

[4] Y. Liu, B. Sun, Y. Tian, X. Wang, Y. Zhu, R. Huai, and Y. Shen, "Software-defined active lidars for autonomous driving: A parallel intelligence-based adaptive model," *IEEE Transactions on Intelligent Vehicles*, pp. 1–10, 2023.

[5] A. Caillot, S. Ouerghi, P. Vasseur, R. Boutteau, and Y. Dupuis, "Survey on cooperative perception in an automotive context," *IEEE Transactions on Intelligent Transportation Systems*, vol. 23, no. 9, pp. 14 204–14 223, 2022.

[6] R. Xu, H. Xiang, X. Han, X. Xia, Z. Meng, C.-J. Chen, C. Correa-Jullian, and J. Ma, "The opencda open-source ecosystem for cooperative driving automation research," *IEEE Transactions on Intelligent Vehicles*, vol. 8, no. 4, pp. 2698–2711, 2023.

[7] Z. Zheng, X. Han, X. Xia, L. Gao, H. Xiang, and J. Ma, "Opencda-ros: Enabling seamless integration of simulation and real-world cooperative driving automation," *IEEE Transactions on Intelligent Vehicles*, vol. 8, no. 7, pp. 3775–3780, 2023.

[8] H. Gao, B. Cheng, J. Wang, K. Li, J. Zhao, and D. Li, "Object classification using cnn-based fusion of vision and lidar in autonomous vehicle environment," *IEEE Transactions on Industrial Informatics*, vol. 14, no. 9, pp. 4224–4231, 2018.

[9] Z. Song, F. Wen, H. Zhang, and J. Li, "A cooperative perception system robust to localization errors," in *2023 IEEE Intelligent Vehicles Symposium (IV)*. IEEE, 2023, pp. 1–6.

[10] R. Xu, W. Chen, H. Xiang, X. Xia, L. Liu, and J. Ma, "Model-agnostic multi-agent perception framework," in *2023 IEEE International Conference on Robotics and Automation (ICRA)*, 2023, pp. 1471–1478.

[11] 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*. Springer, 2022, pp. 107–124.

[12] Z. Lei, S. Ren, Y. Hu, W. Zhang, and S. Chen, "Latency-aware collaborative perception," in *European Conference on Computer Vision*. Springer, 2022, pp. 316–332.

[13] 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*, vol. 8, no. 4, pp. 2650–2660, 2023.

[14] 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 *2022 International Conference on Robotics and Automation (ICRA)*. IEEE, 2022, pp. 2583–2589.

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

[16] 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 *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2023, pp. 13 712–13 722.

[17] 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 *2021 IEEE International Intelligent Transportation Systems Conference (ITSC)*, 2021, pp. 1155–1162.

[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 *2023 IEEE International Conference on Robotics and Automation (ICRA)*, 2023, pp. 5581–5587.

[19] F.-Y. Wang, "Parallel system methods for management and control of complex systems," *CONTROL AND DECISION.*, vol. 19, pp. 485–489, 2004.

[20] P. Ye, X. Wang, W. Zheng, Q. Wei, and F.-Y. Wang, "Parallel cognition: Hybrid intelligence for human-machine interaction and management," *Frontiers of Information Technology & Electronic Engineering*, vol. 23, no. 12, pp. 1765–1779, 2022.

[21] Q. Wei, H. Li, and F.-Y. Wang, "Parallel control for continuous-time linear systems: A case study," *IEEE/CAA Journal of Automatica Sinica*, vol. 7, no. 4, pp. 919–928, 2020.

[22] F.-Y. Wang, "The dao to metacontrol for metasystems in metaverses: The system of parallel control systems for knowledge automation and control intelligence in cpss," *IEEE/CAA Journal of Automatica Sinica*, vol. 9, no. 11, pp. 1899–1908, 2022.

[23] J. Lu, L. Han, Q. Wei, X. Wang, X. Dai, and F.-Y. Wang, "Event-triggered deep reinforcement learning using parallel control: A case study in autonomous driving," *IEEE Transactions on Intelligent Vehicles*, vol. 8, no. 4, pp. 2821–2831, 2023.

[24] Y. Shen, Y. Liu, Y. Tian, and X. Na, "Parallel sensing in metaverses: Virtual-real interactive smart systems for "6s" sensing," *IEEE/CAA Journal of Automatica Sinica*, vol. 9, no. 12, pp. 2047–2054, 2022.

[25] Y. Liu, Y. Shen, C. Guo, Y. Tian, X. Wang, Y. Zhu, and F.-Y. Wang, "Metasensing in metaverses: See there, be there, and know there," *IEEE Intelligent Systems*, vol. 37, no. 6, pp. 7–12, 2022.

[26] Y. Liu, Y. Tian, B. Sun, Y. Wang, and F.-Y. Wang, "Parallel lidars meet the foggy weather," *IEEE Journal of Radio Frequency Identification*, vol. 6, pp. 867–870, 2022.

[27] F.-Y. Wang, N.-N. Zheng, D. Cao, C. M. Martinez, L. Li, and T. Liu, "Parallel driving in cpss: A unified approach for transport automation and vehicle intelligence," *IEEE/CAA Journal of Automatica Sinica*, vol. 4, no. 4, pp. 577–587, 2017.

[28] Z. Wang, C. Lv, and F.-Y. Wang, "A new era of intelligent vehicles and intelligent transportation systems: Digital twins and parallel intelligence," *IEEE Transactions on Intelligent Vehicles*, vol. 8, no. 4, pp. 2619–2627, 2023.

[29] F. Wang, X. Meng, S. Du, and Z. Geng, "Parallel light field: The framework and processes," *Chin. J. Intell. Sci. Technol*, vol. 3, no. 1, pp. 110–122, 2021.

[30] F.-Y. Wang and Y. Shen, "Parallel light fields: A perspective and a framework," *IEEE/CAA Journal of Automatica Sinica*, vol. 11, no. 2, pp. 542–544, 2024.

[31] K. Wang, T. Zhou, X. Li, and F. Ren, "Performance and challenges of 3d object detection methods in complex scenes for autonomous driving," *IEEE Transactions on Intelligent Vehicles*, vol. 8, no. 2, pp. 1699–1716, 2023.

[32] L. Wang, X. Zhang, Z. Song, J. Bi, G. Zhang, H. Wei, L. Tang, L. Yang, J. Li, C. Jia, and L. Zhao, "Multi-modal 3d object detection in autonomous driving: A survey and taxonomy," *IEEE Transactions on Intelligent Vehicles*, vol. 8, no. 7, pp. 3781–3798, 2023.

[33] C. R. Qi, H. Su, K. Mo, and L. J. Guibas, "Pointnet: Deep learning on point sets for 3d classification and segmentation," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 652–660.

[34] C. R. Qi, L. Yi, H. Su, and L. J. Guibas, "Pointnet++: Deep hierarchical feature learning on point sets in a metric space," *Advances in neural information processing systems*, vol. 30, 2017.

[35] W. Shi and R. Rajkumar, "Point-gnn: Graph neural network for 3d object detection in a point cloud," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2020, pp. 1711–1719.

[36] X. Pan, Z. Xia, S. Song, L. E. Li, and G. Huang, "3d object detection with pointformer," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2021, pp. 7463–7472.

[37] Y. Zhou and O. Tuzel, "Voxelnet: End-to-end learning for point cloud based 3d object detection," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2018, pp. 4490–4499.

[38] Y. Yan, Y. Mao, and B. Li, "Second: Sparsely embedded convolutional detection," *Sensors*, vol. 18, no. 10, p. 3337, 2018.

[39] A. H. Lang, S. Vora, H. Caesar, L. Zhou, J. Yang, and O. Beijbom, "Pointpillars: Fast encoders for object detection from point clouds," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2019, pp. 12 697–12 705.

[40] S. Shi, C. Guo, L. Jiang, Z. Wang, J. Shi, X. Wang, and H. Li, "Pv-rcnn: Point-voxel feature set abstraction for 3d object detection," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2020, pp. 10 529–10 538.

[41] G. P. Meyer, A. Laddha, E. Kee, C. Vallespi-Gonzalez, and C. K. Wellington, "Lasernet: An efficient probabilistic 3d object detector for autonomous driving," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2019, pp. 12 677–12 686.

[42] H. Pei, J. Zhang, Y. Zhang, X. Pei, S. Feng, and L. Li, "Fault-tolerant

cooperative driving at signal-free intersections," *IEEE Transactions on Intelligent Vehicles*, vol. 8, no. 1, pp. 121–134, 2023.

[43] 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 *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022, pp. 21 361–21 370.

[44] T.-H. Wang, S. Manivasagam, M. Liang, B. Yang, W. Zeng, and R. Urtasun, "V2vnet: Vehicle-to-vehicle communication for joint perception and prediction," in *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part II 16*. Springer, 2020, pp. 605–621.

[45] Y. Li, S. Ren, P. Wu, S. Chen, C. Feng, and W. Zhang, "Learning distilled collaboration graph for multi-agent perception," *Advances in Neural Information Processing Systems*, vol. 34, pp. 29 541–29 552, 2021.

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

[47] H. Xiang, R. Xu, and J. Ma, "Hm-vit: Hetero-modal vehicle-to-vehicle cooperative perception with vision transformer," *arXiv preprint arXiv:2304.10628*, 2023.

[48] Z. Tu, H. Talebi, H. Zhang, F. Yang, P. Milanfar, A. Bovik, and Y. Li, "Maxvit: Multi-axis vision transformer," in *European conference on computer vision*. Springer, 2022, pp. 459–479.

[49] H. X. W. S. B. Z. J. M. Runsheng Xu, Zhengzhong Tu, "Cobevt: Cooperative bird's eye view semantic segmentation with sparse transformers," in *Conference on Robot Learning (CoRL)*, 2022.

[50] Q. Chen, X. Ma, S. Tang, J. Guo, Q. Yang, and S. Fu, "F-cooper: Feature based cooperative perception for autonomous vehicle edge computing system using 3d point clouds," in *Proceedings of the 4th ACM/IEEE Symposium on Edge Computing*, 2019, pp. 88–100.