

Enhanced Cooperative Perception for Autonomous Vehicles Using Imperfect Communication

Ahmad Sarlak*, Hazim Alzorgan*, Sayed Pedram Haeri Boroujeni*, Abolfazl Razi*, Rahul Amin†

*School of Computing, Clemson University, Clemson, SC, USA

Email: {asarlak, halzorg, shaerib, arazi}@clemson.edu

†Lincoln Laboratory, Massachusetts Institute of Technology, Lexington, MA, USA

Email: rahul.amin@ll.mit.edu

Abstract—Sharing and joint processing of camera feeds and sensor measurements, known as Cooperative Perception (CP), has emerged as a new technique to achieve higher perception qualities. CP can enhance the safety of Autonomous Vehicles (AVs) where their individual visual perception quality is compromised by adverse weather conditions (haze as foggy weather), low illumination, winding roads, and crowded traffic. While previous CP methods have shown success in elevating perception quality, they often assume perfect communication conditions and unlimited transmission resources to share camera feeds, which may not hold in real-world scenarios. Also, they make no effort to select better helpers when multiple options are available.

To cover the limitations of former methods, in this paper, we propose a novel approach to realize an optimized CP under constrained communications. At the core of our approach is recruiting the best helper from the available list of front vehicles to augment the visual range and enhance the Object Detection (OD) accuracy of the ego vehicle. In this two-step process, we first select the helper vehicles that contribute the most to CP based on their visual range and lowest motion blur. Next, we implement a radio block optimization among the candidate vehicles to further improve communication efficiency. We specifically focus on pedestrian detection as an exemplary scenario. To validate our approach, we used the CARLA simulator to create a dataset of annotated videos for different driving scenarios where pedestrian detection is challenging for an AV with compromised vision. Our results demonstrate the efficacy of our two-step optimization process in improving the overall performance of cooperative perception in challenging scenarios, substantially improving driving safety under adverse conditions. Finally, we note that the networking assumptions are adopted from LTE Release 14 Mode 4 side-link communication, commonly used for Vehicle-to-Vehicle (V2V) communication. Nonetheless, our method is flexible and applicable to arbitrary V2V communications.

Index Terms—Cooperative perception, connected autonomous vehicles, 3d object detection, intermittent connectivity, vehicular communications.

I. INTRODUCTION

Cooperative Perception (CP) represents a new paradigm in Autonomous Vehicles (AVs) by mitigating the limitations of individual vehicle perception [1], [2]. CP involves sharing and the integrative processing of camera feeds and sensor readings by multiple AVs within a network, enabling them to collectively enhance their situational awareness. The use of CP stems from the fact that the reliable operation of AVs can

face significant challenges due to perception errors in adverse weather conditions and complex traffic scenarios, where individual vehicle perception systems may be compromised. CP has recently emerged as a promising solution to enhance AV safety by leveraging collaborative information sharing among vehicles.

A typical CP approach employs Vehicle-to-Vehicle (V2V) communication technology to share sensory data among vehicles for collective processing [3]–[5]. To realize the collective processing of shared visual information, there exist three mainstream fusion methods, as illustrated in Figure 1. The first approach is *early fusion*, where the raw captured images by multiple vehicles are translated to the same field of view through projection transformation, and then mixed to obtain a single high-quality image to be used by the downstream learning-based tasks such as object detection (OD), depth estimation, reinforcement learning (RL), etc. (Fig. 1 (a)). The main issue of this approach is the sensitivity of the fused image to the parameters of projection transformation. The second approach is *late fusion*, where each image is first processed by the learning-based task, then the obtained results are mixed (Fig. 1 (b)). For instance, one may execute OD on individual images to obtain object classes and bounding boxes, to be fused to get more precise results. The advantage of these two methods is their easy integration with existing deep learning (DL)-based visual processing methods [6]. The third approach is more intricate and involves developing new DL architectures that jointly process multiple image/video channels (Fig. 1 (c)). This method can yield better results for specific tasks but requires custom-designed DL architectures that can limit their applicability. In this work, we use the *late fusion* method for its simplicity and practicality.

While existing CP methods have demonstrated success in addressing these challenges, most of them have taken the unrealistic assumptions of perfect communication conditions and unlimited transmission resources [3], [4], which may not align with the real-world scenarios. To the best of our knowledge, there exist only two works that consider imperfect communications in CP. First is [7], where the authors present a vision transformer, V2X-ViT, along with a CP framework that utilizes Vehicle-to-Everything (V2X) communication to improve AV perception in terms of 3D object detection. While the authors try to address V2X challenges, their primary

This material is based upon work supported by the National Science Foundation under Grant Number CNS2204721 and MIT Lincoln Lab under Grant Number 7000565788.

focus is on communication delays, overlooking other essential communication aspects (packet drop rate, throughput, ...). Their proposed architecture aimed to enhance robustness to delay in the network.

Likewise, [8] intends to address the impact of imperfect V2X communication in cooperative perception for 3D object detection. They proposed an innovative approach with an LC-aware Repair Network (LCRN) and a V2V Attention Module (V2VAM) to effectively mitigate the adverse effects of imperfect communication. They significantly improved the detection performance demonstrated on the OPV2V dataset. However, they only tried to adjust their method for varying channel quality without incorporating channel quality into the vehicle selection process.

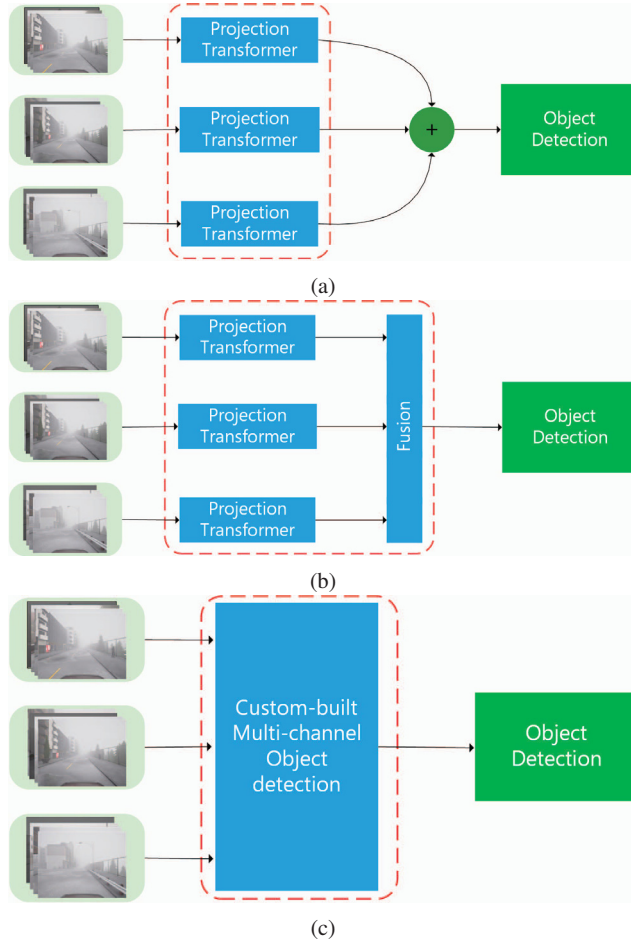


Fig. 1: Three fusion approaches used in cooperative perception, including (a) early fusion, (b) late fusion, and (c) integrative analysis.

We propose a novel approach to CP to overcome the limitations of the previous methods by enabling selective communication, where the helper vehicles are selected based on their visual information as well as their channel quality. The networking parameters are adopted from the LTE Release 14 Mode 4 side-link communication used for V2V communica-

tion. By doing so, we achieve enhanced situational awareness through higher CV-based detection accuracy under imperfect and constrained communication scenarios, compared to CV methods that blindly or randomly selected helper vehicles. In this work, we select helper vehicles based on their contribution to extending the visual range of the ego vehicle, considering their geo-locations as well as their imperfect channel conditions. Additionally, we take into account the vehicles' relative velocities, because the motion blur impacts their imaging quality [9]–[11] and consequently the ultimate accuracy of the developed CP-based object detection accuracy.

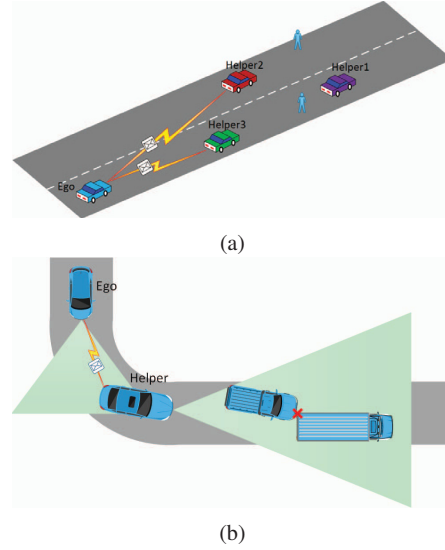


Fig. 2: Two scenarios that may occur on the road are: (a) in foggy weather, helper vehicles closer to pedestrians (Green and Purple) can share their cameras with the ego vehicle (Red vehicle), and (b) on a curved road, a helper can detect an accident and share its camera with the ego vehicle.+

To adopt more realistic assumptions, we note that transmission errors in C-V2X mode 4 arise from different factors, including (1) using half-duplex transmission, (2) receiving signal strengths below the sensing threshold, (3) propagation effects, and (4) packet collisions [12]. In addition to considering channel quality in selecting the best helper vehicles, we also enhance the resulting communication efficiency through radio block optimization among the selected vehicles [13].

To investigate the effectiveness of the proposed approach, we concentrate on pedestrian detection as an exemplary scenario. However, the proposed framework is generic and can be used for other applications, such as speed estimation, lane detection, vehicle classification, traffic sign detection, traffic light interpretation, etc. Through the generation of frames using the CARLA simulator [14], annotation of data, and compilation of a comprehensive dataset, we conduct experiments to validate the performance of our proposed method.

The results obtained from our two-step optimization process, present notable improvements in CP by selecting the best helper to increase the ego vehicle's visual range.

The rest of the paper is organized as follows. In section 2, a more inclusive review of related work is presented. In section 3, the system model is presented. Finally, the proposed method is evaluated through intensive simulations in section 4, followed by concluding remarks in section 5.

II. RELATED WORKS

A. CAVs with Imperfect Communications

V2V LTE Release 14 presents advancements in V2V communication, enhancing connectivity and communication protocols for improved safety and efficiency in vehicular networks. However, there are limitations in V2V communications, that that should be considered when developing networked service provisioning systems. Authors in [12] evaluate the communication performance of C-V2X or LTE-V Mode 4, focusing on average Packet Delivery Ratio (PDR) and four types of transmission errors, with validating various transmission parameters and traffic densities. Authors in [15], [16] highlight the importance of wireless factors, such as packet errors and limited bandwidth in Federated Learning (FL) implementations. Li et al. [17] presents the importance of efficient resource allocation techniques in V2V communication networks in FL, particularly in the vehicular safety services area, where delay and reliability requirements need optimal utilization of limited spectrum resources. The authors state that such resource allocation strategies are essential to mitigate interference, support various quality-of-service requirements, and ensure the success of emerging vehicle-related services in dynamic and fast-changing vehicular environments. These researches show that imperfect communication can play a key role in CP; therefore, it should not be overlooked when forming a network of vehicles for collective perception.

B. 3D Object Detection

Our target application, in this work, is 3D OD, which involves identifying and locating objects in three-dimensional space. It is an integral part of autonomous driving and crash avoidance [8]. 3D OD can be applied to different perception domains, including visions-based (e.g., regular and IR cameras), sensor-based (e.g., RADAR/LiDAR), and hybrid methods, with the following notable implementations.

Reading et al. [18] proposed Categorical Depth Distribution Network (CaDDN), as a monocular vision-based 3D OD method. This method utilizes predicted categorical depth distributions for each pixel to enhance depth estimation accuracy and achieve a top-ranking performance on the KITTI dataset, showcasing its effectiveness in addressing the challenges of monocular 3D detection. Authors of [19] introduced a multi-camera 3D OD framework that operates directly in 3D space, leveraging sparse 3D object queries to index 2D features from multiple camera images. [20] proposes PointRCNN which is a 3D OD framework, where using point cloud data directly to generate high-quality 3D proposals instead of generating proposals from RGB image or projecting point cloud to bird's view or voxels.

An exemplary implementation of LiDAR-based OD methods is PointPillars, a highly effective point cloud encoder, that employs PointNets to organize point clouds into vertical columns (pillars), demonstrating superior speed and accuracy [21]. They outperform existing methods on the KITTI benchmarks while achieving a higher accuracy in OD in comparison to [22]–[24]. Finally, a hybrid camera-LiDAR fusion detection method, called RCBEV, is proposed in [25]. RCBEV implements a radar-camera feature fusion method for 3D OD in autonomous driving using nuScenes and VOD dataset through an efficient top-down feature representation and a two-stage fusion model which bridges the view disparity between radar and camera features.

C. Fusion Methods

The information of multiple perception systems can be collectively processed in three different modes, namely *early*, *integrative*, and *late fusion*. The authors of [26] introduce an *early fusion* approach enabling cooperative perception through the fusion of LiDAR 3D point clouds from diverse positions, improving detection accuracy and expanding the effective sensing area. They validated their method using KITTI and T&J dataset, demonstrating the feasibility of transmitting point cloud data via existing vehicular network technologies. Qiu et al. present an infrastructure-less CP system, leveraging direct vehicle-to-vehicle communication to efficiently share sensor readings while optimizing transmission schedules to enhance safety in dense traffic scenarios [27].

A *late fusion* machine approach is adopted in [28] to enhance the accuracy of shared information for collaborative automated driving. First, they employed Convolutional Neural Networks (CNN) for OD and classification to extract positional and dimensional information. Next, they combined extracted information to enhance driving safety in V2V collaborative systems. Article [29] explores CP for 3D OD using two fusion schemes, *early* and *late fusion*, showing that *early fusion* outperforms *late fusion* in challenging scenarios, while authors in [30] introduce the DAIR-V2X dataset for Vehicle-Infrastructure Cooperative Autonomous Driving (VICAD) and formulates the Vehicle-Infrastructure Cooperative 3D OD (VIC3D) problem, proposing the Time Compensation *late fusion* framework as a benchmark, highlighting the importance of collaborative solutions and addressing challenges in autonomous driving.

Both *early* and *late fusion* methods try to combine information either before or after applying the target DL-based application, like 3D OD. This brings the convenience of deploying existing methods. Recently, some attempts have been made to develop integrative processing of multi-channel inputs to achieve elevated performance. Xu et al. [7] introduced a fusion model V2X-ViT, based on the popular vision transformer architecture equipped with heterogeneous multi-agent and multi-scale window self-attention modules for robust CP in autonomous vehicles, achieving state-of-the-art 3D OD performance in challenging and noisy environments using a large-scale V2X perception dataset. Another *integrative*

end-to-end DL-based architecture is presented in [31] that utilizes cross-vehicle perception with LiDAR data encoded into compact point-based representations, demonstrating an improvement in average success rate over egocentric driving models in challenging scenarios, with a 5× smaller bandwidth requirement compared to prior work, as evaluated in the AUTOCASIM simulation framework.

III. SYSTEM MODEL

In our approach, the ego vehicle sends request-for-help messages to front vehicles when individual perception quality is not satisfactory. We assume that N vehicles, $V_1, V_2, \dots, V_N$, notify their intention to provide CP assistance through Ack messages along with their position (x_i, y_i) , speed v_i , channel conditions, and sample images. Additional details, such as the distance to the ego vehicle, approximate vision range, motion blur, field of view, required transmission energy budget E , and communication errors in the C-V2X mode 4, are estimated based on the gathered information (position, channel conditions) and refined through an examination of the sample images. Then, various error factors, including half-duplex transmission, received signal power below the sensing power threshold, propagation effects, and packet collisions, are combined into β , denoting the effective packet error probability. Once the information is obtained, the ego vehicle selects M out of N volunteered vehicles that collectively achieve the highest performance under given conditions.

The main objective of selecting helpers is to maximize contextual information diversity through an effective and informed choice of helpers. In our case, the key contributing factors include the vehicle's contribution to extending the visual range of the ego vehicle (based on their positions), the quality of their image based on motion blur (based on their velocities), and the channel conditions. We develop a set of objective functions, as elaborated below, that quantify these factors and facilitate selecting the best helper.

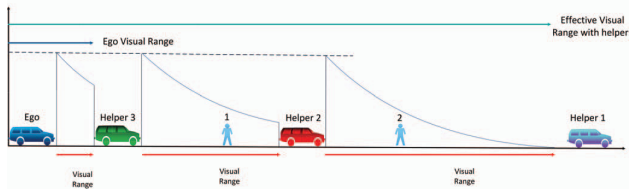


Fig. 3: The individual visual range for vehicles. Recruiting helpers extend the visual range of the ego vehicle through cooperative perception.

A. Approximate Vision Range

A key factor in selecting helpers is the collective visual range of the ego vehicle that enables accurate object detection. The visual range of each vehicle varies with weather conditions, light conditions, and the presence of obstacles. The effective individual visual range of vehicles is demonstrated in Fig. 3.

By optimizing the selection process, we ensure that the chosen helpers effectively contribute to expanding the collective visual range of the ego vehicle.

We assume that all vehicles know their location (x_i, y_i) and their distance to the front obstacle, and share this information with the ego vehicle. When calculating the total visual range, we use function $h(x)$ to determine individual visual ranges

$$h(x) = u[x - (x_i + L_i)] - u[x - x_i], \quad (1)$$

$$L_i = \min(T, x_{i+1} - x_i),$$

where x_i represents the Longitudinal motion along the linear road on x-axis. Otherwise, we substitute $x \leftarrow x \cos\theta$ or $x \leftarrow y \sin\theta$ from positions if the road makes θ with x-axis. Here, $h_i(x)$ is a rectangular pulse of length L_i , representing the i^{th} vehicle's visual range until the threshold T defined by the weather and light conditions or the distance to the front vehicle $x_{i+1} - x_i$, whichever is shorter. The sum of these pulses from selected vehicles represents the augmented visual range.

In addition to expanding the visual range, we desire to prioritize selecting closer vehicles. This is crucial because having a clear vision for detecting closer objects is imperative for safe driving. For instance, in Fig. 3 Helper 3 can help with detecting both pedestrians at least partially, Helper 2 can help only with detecting one pedestrian, and Helper 1's contribution is irrelevant (not useful) in detecting either of pedestrians.

To capture this effect, we use an exponential decay function $g(x) = e^{-ax}$ with a tuning parameter a to weigh the augmented visual range to promote selecting closer helpers. If the binary vector $\vec{\alpha} = (\alpha_1, \alpha_2, \dots, \alpha_N)$ represents the volunteer vehicles ($\alpha_i = 1$: selected, $\alpha_i = 0$: not selected), the weighted augmented visual range of selected vehicles can be expressed as

$$f_1(\vec{\alpha}) = \int_{x=0}^{\infty} g(x) \cdot \left[\sum_{i=1}^N \alpha_i h_i(x) \right]. \quad (2)$$

B. Motion Blur Level

We assume that all vehicles are equipped with similar cameras with an equal resolution and field of view. In this paper, the only important parameter affecting motion blur is the speed of the vehicles. Motion blur represents the concept of the decline in image quality for higher speeds. We follow the characterization of motion blur in [9] to obtain

$$f_2(\vec{\alpha}) = \sum_{i=1}^N \alpha_i \frac{v'_i T [f \cos(\phi) - QS \sin(\phi)]}{v'_i T Q \sin(\phi) + zp}, \quad (3)$$

for the vehicles joining CP ($\alpha_i = 1$). Here, T is the exposure time interval, f is the camera focal length, Q is the charge-coupled device (CCD) pixel size in the horizontal direction, S is the starting position of the object in the image (in pixels), ϕ is the angle between the motion direction and the image plane, d is the perpendicular distance from the starting point of the moving object to the pinhole [10]. The effect is maximized and becomes linear in velocity, when the image plane is parallel to the motion direction ($\phi = 0$), as we get

$$f_2(\vec{\alpha}) = \sum_{i=1}^N \alpha_i \frac{v'_i T f}{zp}. \quad (4)$$

C. Communications Parameters

The transmission energy budget and packet drop rate are essential metrics in C-V2X communications, as they directly impact the reliability and effectiveness of V2X communication and the overall safety and efficiency of AVs by determining the successful exchange of critical information among vehicles and infrastructure.

For the definition of effective throughput, we absorb the impact of all error terms in LTE C-V2X into packet error rate β . As a result, the effective throughput ζ_i of vehicle V_i is

$$\zeta_i = R_{ch}/E[R_i] = R_{ch}(1 - \beta_i)w_i, \quad (5)$$

where R_{ch} is the rate of channel, and w_i is the bandwidth of the resource block assigned to vehicle i . Here, R_i is the number of re-transmissions with the following expected value

$$E[R_i] = 1/(1 - \beta_i). \quad (6)$$

Therefore, the average throughput under selection vector α is

$$f_3(\mathbf{w}) = \sum_{i=1}^N w_i \zeta_i = \sum_{i=1}^N R_{ch}(1 - \beta_i)w_i. \quad (7)$$

$$f_3(\vec{\alpha}, \vec{w}) = \sum_{i=1}^N \alpha_i \zeta_i = \sum_{i=1}^N R_{ch}(1 - \beta_i)w_i \alpha_i. \quad (8)$$

The transmission energy of vehicle V_i is

$$e_i = \frac{P_i^{tr}}{R_i \cdot (1 - \beta_i)} \alpha_i \quad (9)$$

where P_i^{tr} denotes the average transmission power, considering the number of re-transmissions represented by $1/(1 - \beta_i)$. Therefore, the average energy consumption for selected vehicles is

$$f_4(\vec{\alpha}, \vec{w}) = \sum_{i=1}^N \alpha_i \frac{P_i^{tr}}{R_{ch}(1 - \beta_i)}, \quad (10)$$

One may also include then End-to-End (E2E) delay in the optimization. In general, E2E delay accounts for sampling and perception delays in the sender, queuing delays, channel setup delays, congestion and re-transmissions, actual transmission delays, as well as processing delays in the receiver [32]. We take a simplistic assumption and model $D_i^{(1)}$, the E2E delay for one packet for vehicle V_i as an exponentially distributed continuous-valued Random Variable (RV) with vehicle-specific mean λ_i following some prior work [33], and then specifically, we define

$$D_i^{(1)} \sim f_{\lambda_i}(d_i), \quad f_{\lambda_i}(d_i) = \begin{cases} \lambda_i e^{-\lambda_i d_i}, & d_i \geq 0, \\ 0, & d_i < 0. \end{cases} \quad (11)$$

Here, we use capital letters for RVs and lowercase letters for their realizations. We consider that an average delay $E[D_i^{(1)}] = 1/\lambda_i$ for each vehicle V_i remains constant in one

transmission slot. Here, we consider $D_i^{(1)}$ captures all delay terms. However, if it accounts only for queuing delay, we can add the constant term $D_{tr} = l/R_{ch}$ accounting for the actual transmission delay, where l is the packet length (in bits). For convenience, we exclude delay in the optimization, since CP requires near-realtime synchronization between packets from helper vehicles.

D. Optimization Problem

With both the visual range and the speed of each vehicle, a selection problem arises. This process aims to select the vehicle that not only extends the overall visual coverage for the ego vehicle but also contributes to capturing higher-quality images with reduced motion blur.

We formulated our problem as two optimizations, in which first, we tried to select the best M vehicles out of N and then allocate them among the selected vehicles. Our optimization is based on a set of transmission Key Performance Indicators (KPIs) calculated for selected vehicles. Specifically if the binary vector $\vec{\alpha} = (\alpha_1, \alpha_2, \dots, \alpha_N)$ represents the selected vehicles.

The optimization problem is

$$\begin{aligned} \arg \max_{\alpha, \vec{w}} f(\vec{\alpha}, \vec{w}) &= \sum_{i=1}^4 k_i f_i(\vec{\alpha}, \vec{w}) \\ \text{s.t.} \quad &\alpha_i \in \{0, 1\}, \\ &\sum_{i=1}^N \alpha_i \leq M, \\ &\sum_{i=1}^M \alpha_i w_i = B, \end{aligned} \quad (12)$$

where k_i is a tuning factor to weigh the importance of different objectives (and/or standardize them).

The constraints mean that the number of selected vehicles should not exceed M and the utilized bandwidth by selected vehicles can not exceed the total bandwidth. To simplify this problem, we take a two-step process. We first select the vehicles solely based on their contribution to CP visual quality by solving

$$\begin{aligned} \arg \max_{\vec{\alpha}} f(\vec{\alpha}) &= \sum_{i=1}^2 2k_i f_i(\vec{\alpha}) \\ \text{s.t.} \quad &\alpha_i \in \{0, 1\}, \\ &\sum_{i=1}^N \alpha_i \leq M, \end{aligned} \quad (13)$$

and then we assign the networking resources among selected

vehicles, by solving

$$\begin{aligned} \arg \max_{\mathbf{w}} f(\mathbf{w}) &= \sum_{i=3}^4 k_i f_i(w_1, w_2, \dots, w_M) \\ \text{s.t.} \quad &\sum_{i=1}^M p_i^{tr} = P_T, \\ &\sum_{i=1}^M w_i = B, \\ &E[D_i] = \frac{1}{\lambda_i} \leq T_{tr} \end{aligned} \quad (14)$$

only for selected vehicles.

In characterizing bandwidth in (14), we consider LTE-V2X, a communication technology designed for V2V communication, which supports 10 and 20 MHz channel bandwidths. This bandwidth is divided into sub-channels in the frequency domain and sub-frames in the time domain. Resource Blocks (RBs) are crucial components in the frequency domain, each consisting of 12 sub-carriers spaced by 15 kHz. The signal is formed by resource block pairs (RBPs), defined by 12 sub-carriers and carrying 14 OFDM symbols. Subchannels, defined by groups of RBPs within the same subframe, serve as the minor units of resources allocated to vehicles for transmitting Cooperative Awareness Messages (CAM). The RB, with a width of 180 kHz, becomes the smallest unit of frequency resources allocated to LTE users. Subchannels transmit data and control information, with Transport Blocks (TBs) containing complete packets, beacons, or other messages. Resource block allocation is vital in LTE-V2X due to the need for efficient utilization of the available spectrum. The number of RBs per sub-channel can vary, and vehicles autonomously select resources without cellular infrastructure assistance.

We have used Genetic Algorithm to solve optimization problems in 13 and (14) for its power in solving binary-vector optimization problems.

E. Fusion

In our implementation, each vehicle utilized YOLOv8 for OD. After that, the ego vehicle captures the helpers to increase the visual range and detect objects more accurately. For the fusion part, we select the best IoU between the ego vehicle and the helper.

$$IoU_s = \text{Max}(IoU_e, IoU_h) \quad (15)$$

Consequently, we always select the best IoU, and the accuracy of OD will increase.

IV. SIMULATION

In this section, we investigate the performance of the proposed CP of the OD accuracy under foggy weather conditions while selecting the best vehicle under imperfect communication. In our experiments, we set the number of Vehicles $N = 10$, the number of selected vehicles as $M = 3$, channel conditions (β_i, λ_i) , relative velocity (v_i) , vehicle position (x_i, y_i) are selected randomly. We use Beta and exponential distributions

to generate β_i and $1/\lambda_i$, respectively. We trained our model using scenarios generated by the CARLA simulator. As there is no appropriate dataset for a visual CP, we developed over 2000 images with CARLA in which there is one ego vehicle and three helpers in foggy weather. We manually annotated the dataset and then used the YOLOv8 to detect pedestrians in foggy weather.



Fig. 4: Objects detection by YOLOv8 to locate the pedestrian. Left: ego vehicle's view, Right: helper vehicle's view. The green box is the ground truth and the red box is the predicted bounding box by YOLOv8.

Fig. 6 shows our method performs better than other selections regarding the vehicle's visual range. We ran the selection 100 times in random selection and got the average result. Fig. 7 shows that our method can select vehicles with lower motion blur, impacting image quality.

Fig. 8 and Fig. 9 demonstrated better performance in allocating resource blocks to vehicles in V2V communication. In random selection, resources are allocated randomly among vehicles, while in uniform allocation, resources are equally distributed among vehicles.

Table I shows that when an ego vehicle recruits a helper, it can enhance OD accuracy more than when each vehicle (ego vehicle and helpers) detects objects individually. In table I, we consider the scenario in a perfect communication. The result shows how much recruiting a helper can improve OD accuracy. However, we observe that Helper1 performs poorly in IoU, recall and F1 score as it fails to detect objects due to its distance from the ego vehicle, while the best results are achieved by recruiting Helper 2.

We introduce a packet drop rate to the images and subsequently apply the method for OD. Fig. 4 illustrates an example of the packet drop rate impact on the image. The experiment result is presented in table II, with the highest accuracy for our method. The result shows how imperfect communication can impact the OD when we compare it with table I.

V. CONCLUSION

Our novel approach to Cooperative Perception (CP) for Autonomous Vehicles (AVs) addresses imperfect communication. It optimizes helper selection, which leads to significant improvements in perception quality and driving safety under foggy weather. We had extensive experimentation and evaluated the validation of our method via generating an



Fig. 5: Data set generated with CARLA simulator

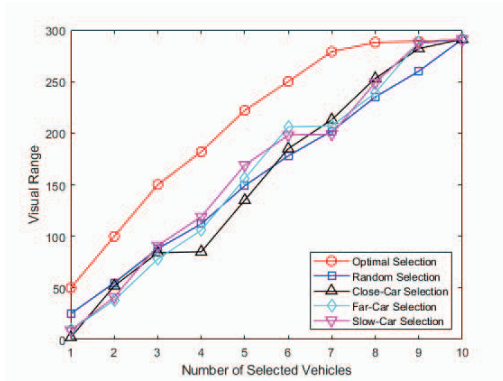


Fig. 6: Optimal vehicle selection has been compared across various scenarios, including random selection (averaged over 100 runs), close-car selection (choosing the closest cars), far-car selection (choosing the farthest cars), and slow-car selection (selecting the slowest cars), all based on visual range considerations.

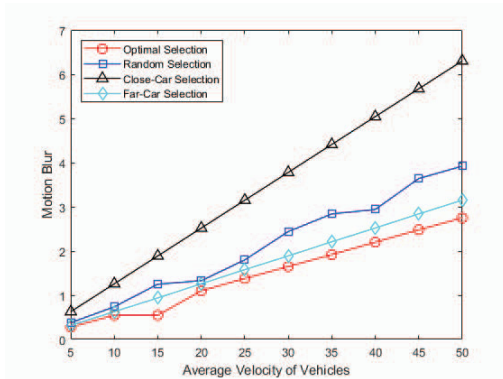


Fig. 7: Optimal selection vehicle compared to other selection method.

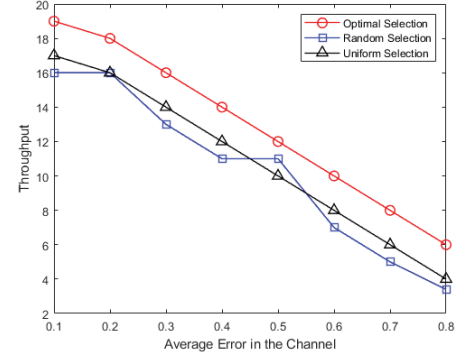


Fig. 8: Optimal selection vehicle compared to random selection and Uniform selection when increasing the error versus Throughput.

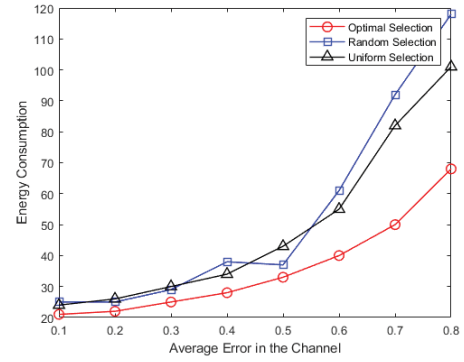


Fig. 9: Optimal selection vehicle compare to random selection and Uniform selection when increase the error versus Energy Consumption.

TABLE I: OD metrics under perfect communication.

	Ego	H_3	H_2	H_1	Ego + H_3	Ego + H_2	Ego + H_1
IoU	0.42	0.68	0.74	0.43	0.76	0.86	0.59
Recall	0.224	0.343	0.373	0.208	0.382	0.435	0.304
F1 Score	0.36	0.501	0.533	0.347	0.544	0.591	0.462

TABLE II: Comparison our method to other selections when we add error channel to the images

	Our Method	Randomly Selection	Close-Car Selection	Far-Car Selection
IoU	0.83	0.55	0.31	0.49
Recall	0.4	0.27	0.15	0.24
F1 Score	0.57	0.4	0.24	0.38

appropriate dataset in foggy weather in the CARLA simulator, demonstrating notable enhancements in CP in pedestrian detection scenarios. Our framework considered CP in imperfect communications, while most of the method works in perfect communication. Also, our method can be used for applications in AVs.

REFERENCES

- [1] L. Chen, Y. Li, C. Huang, B. Li, Y. Xing, D. Tian, L. Li, Z. Hu, X. Na, Z. Li, et al., "Milestones in autonomous driving and intelligent vehicles: Survey of surveys," *IEEE Transactions on Intelligent Vehicles*, vol. 8, no. 2, pp. 1046–1056, 2022.
- [2] S. P. H. Boroujeni, A. Razi, S. Khoshdel, F. Afghah, J. L. Coen, L. O'Neill, P. Fule, A. Watts, N.-M. T. Kokolakis, and K. G. Vamvoudakis, "A comprehensive survey of research towards ai-enabled unmanned aerial systems in pre-, active-, and post-wildfire management," *Information Fusion*, p. 102369, 2024.
- [3] 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), pp. 2583–2589, IEEE, 2022.
- [4] R. Xu, Z. Tu, H. Xiang, W. Shao, B. Zhou, and J. Ma, "Cobevt: Cooperative bird's eye view semantic segmentation with sparse transformers," *arXiv preprint arXiv:2207.02202*, 2022.
- [5] A. Sarlak, A. Razi, X. Chen, and R. Amin, "Diversity maximized scheduling in roadside units for traffic monitoring applications," in 2023 IEEE 48th Conference on Local Computer Networks (LCN), pp. 1–4, IEEE, 2023.
- [6] S. P. H. Boroujeni and A. Razi, "Ic-gan: An improved conditional generative adversarial network for rgb-to-ir image translation with applications to forest fire monitoring," *Expert Systems with Applications*, vol. 238, p. 121962, 2024.
- [7] 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, pp. 107–124, Springer, 2022.
- [8] 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.
- [9] H. Xiao, J. Zhao, Q. Pei, J. Feng, L. Liu, and W. Shi, "Vehicle selection and resource optimization for federated learning in vehicular edge computing," *IEEE Transactions on Intelligent Transportation Systems*, vol. 23, no. 8, pp. 11073–11087, 2021.
- [10] J. A. Cortés-Osorio, J. B. Gómez-Mendoza, and J. C. Riaño-Rojas, "Velocity estimation from a single linear motion blurred image using discrete cosine transform," *IEEE Transactions on Instrumentation and Measurement*, vol. 68, no. 10, pp. 4038–4050, 2018.
- [11] D. Ye, R. Yu, M. Pan, and Z. Han, "Federated learning in vehicular edge computing: A selective model aggregation approach," *IEEE Access*, vol. 8, pp. 23920–23935, 2020.
- [12] M. Gonzalez-Martín, M. Sepulcre, R. Molina-Masegosa, and J. Gozalvez, "Analytical models of the performance of c-v2x mode 4 vehicular communications," *IEEE Transactions on Vehicular Technology*, vol. 68, no. 2, pp. 1155–1166, 2018.
- [13] M. Chinipardaz, S. Amraee, and A. Sarlak, "Joint downlink user association and interference avoidance with a load balancing approach in backhaul-constrained hetnets," *Plos one*, vol. 19, no. 3, p. e0298352, 2024.
- [14] A. Dosovitskiy, G. Ros, F. Codevilla, A. Lopez, and V. Koltun, "Carla: An open urban driving simulator," in *Conference on robot learning*, pp. 1–16, PMLR, 2017.
- [15] M. Chen, Z. Yang, W. Saad, C. Yin, H. V. Poor, and S. Cui, "A joint learning and communications framework for federated learning over wireless networks," *IEEE Transactions on Wireless Communications*, vol. 20, no. 1, pp. 269–283, 2020.
- [16] T. Li, A. K. Sahu, A. Talwalkar, and V. Smith, "Federated learning: Challenges, methods, and future directions," *IEEE signal processing magazine*, vol. 37, no. 3, pp. 50–60, 2020.
- [17] X. Li, L. Lu, W. Ni, A. Jamalipour, D. Zhang, and H. Du, "Federated multi-agent deep reinforcement learning for resource allocation of vehicle-to-vehicle communications," *IEEE Transactions on Vehicular Technology*, vol. 71, no. 8, pp. 8810–8824, 2022.
- [18] C. Reading, A. Harakeh, J. Chae, and S. L. Waslander, "Categorical depth distribution network for monocular 3d object detection," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 8555–8564, 2021.
- [19] Y. Wang, V. C. Guizilini, T. Zhang, Y. Wang, H. Zhao, and J. Solomon, "Det3d: 3d object detection from multi-view images via 3d-to-2d queries," in *Conference on Robot Learning*, pp. 180–191, PMLR, 2022.
- [20] S. Shi, X. Wang, and H. Li, "Pointcnn: 3d object proposal generation and detection from point cloud," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 770–779, 2019.
- [21] 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*, pp. 12697–12705, 2019.
- [22] Y. Yan, Y. Mao, and B. Li, "Second: Sparsely embedded convolutional detection," *Sensors*, vol. 18, no. 10, p. 3337, 2018.
- [23] Y. Xiang, W. Choi, Y. Lin, and S. Savarese, "Subcategory-aware convolutional neural networks for object proposals and detection," in 2017 IEEE winter conference on applications of computer vision (WACV), pp. 924–933, IEEE, 2017.
- [24] J. Ku, M. Mozifian, J. Lee, A. Harakeh, and S. L. Waslander, "Joint 3d proposal generation and object detection from view aggregation," in 2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), pp. 1–8, IEEE, 2018.
- [25] T. Zhou, J. Chen, Y. Shi, K. Jiang, M. Yang, and D. Yang, "Bridging the view disparity between radar and camera features for multi-modal fusion 3d object detection," *IEEE Transactions on Intelligent Vehicles*, vol. 8, no. 2, pp. 1523–1535, 2023.
- [26] Q. Chen, S. Tang, Q. Yang, and S. Fu, "Cooper: Cooperative perception for connected autonomous vehicles based on 3d point clouds," in 2019 IEEE 39th International Conference on Distributed Computing Systems (ICDCS), pp. 514–524, IEEE, 2019.
- [27] H. Qiu, P. Huang, N. Asavisanu, X. Liu, K. Psounis, and R. Govindan, "Autocast: Scalable infrastructure-less cooperative perception for distributed collaborative driving," *arXiv preprint arXiv:2112.14947*, 2021.
- [28] Z. Y. Rawashdeh and Z. Wang, "Collaborative automated driving: A machine learning-based method to enhance the accuracy of shared information," in 2018 21st International Conference on Intelligent Transportation Systems (ITSC), pp. 3961–3966, IEEE, 2018.
- [29] E. Arnold, M. Dianati, R. de Temple, and S. Fallah, "Cooperative perception for 3d object detection in driving scenarios using infrastructure sensors," *IEEE Transactions on Intelligent Transportation Systems*, vol. 23, no. 3, pp. 1852–1864, 2020.
- [30] 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*, pp. 21361–21370, 2022.
- [31] J. Cui, H. Qiu, D. Chen, P. Stone, and Y. Zhu, "Coopernaut: End-to-end driving with cooperative perception for networked vehicles," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 17252–17262, 2022.
- [32] R. L. Cruz, "A calculus for network delay. i. network elements in isolation," *IEEE Transactions on information theory*, vol. 37, no. 1, pp. 114–131, 1991.
- [33] A. Razi, A. Valehi, and E. Bentley, "Delay minimization by adaptive framing policy in cognitive sensor networks," in 2017 IEEE Wireless Communications and Networking Conference (WCNC), pp. 1–6, IEEE, 2017.