Scalable and Dynamic Cooperative Perception: A Data/Model Co-Driven Framework

Kaige Qu, *Member, IEEE*, and Weihua Zhuang, *Fellow, IEEE*

***Abstract*—Cooperative perception (CP) is a key approach to ensuring reliable situation awareness of connected and autonomous vehicles (CAVs). In this article, we discuss the key challenges in terms of scalability, dynamics, and performance uncertainty for supporting CP in a practical network environment. Then, we present a data/model co-driven framework for scalable and dynamic CP with performance awareness, as an engineering solution to address the challenges. Specifically, we propose a performance-aware scalable CP scheme based on a learning-assisted optimization approach and a dynamic CP scheme based on an optimization-assisted learning approach for different scenarios, both exploiting data-driven and model-based methods to enhance each other. Finally, a case study is presented to show the effectiveness of our scheme in handling the network dynamics with resource efficiency.**

***Index Terms*—Connected and autonomous vehicles (CAVs), cooperative perception, data fusion, performance estimation, machine learning, data/model co-driven methods.**

I. Introduction

Connected and autonomous vehicles (CAVs) will play a crucial role in the sixth-generation (6G) communication network for use cases such as smart city, smart factory, and smart port, for enhancing the vehicle traffic safety and efficiency [1]–[3]. They rely on real-time environment perception to maintain situation awareness about the surroundings, which includes continually detecting and tracking nearby objects in a time-varying region of interest (RoI). Object detection involves detecting the presence, locations, and classes of surrounding objects, based on computation-intensive deep learning techniques [4], [5]. Object tracking involves maintaining the identity and trajectory of detected objects over time, which is usually lightweight, e.g., based on optical flow. In *tracking-by-detection*, object detection is periodically triggered over multiple time frames, while object tracking associates same objects across frames between consecutive object detections to form object trajectories, during which tracking errors accumulate over time [6], [7]. Hence, object detection should have a low latency and a high accuracy, to promptly compensate for the tracking errors and reset the tracking accuracy to a high level.

Individual CAVs will have the *stand-alone perception (SP)* capability for object detection and tracking by using onboard sensors. Ideally, a CAV should operate without dependencies on external sources of perception information. However, in practice, it may lack a comprehensive understanding of the environment beyond the visible immediate surroundings. Each CAV-object pair is subject to constantly-changing viewing distances/angles and occasional occlusions due to mobility, which hinders consistent and reliable object detection and tracking with SP. In such situations, adequate situational awareness of a CAV requires obtaining and fusing complementary or enhancing sensory information from other sources.

The emerging vehicle-to-everything (V2X) and vehicular edge computing (VEC) technologies provide an opportunity for CAVs to share sensory information and to support the computation-intensive object detection, allowing them to collaboratively perceive a more accurate view of the environment [4], [8]. Fig. 1 shows a V2X and VEC enabled network architecture for supporting the *cooperative perception (CP)*. Benefiting from the expanded sensing range and improved perception accuracy, CP enhances the accuracy, robustness, and reliability of object detection and tracking. When CP is unavailable due to the lack of cooperating CAVs or communication resources, SP serves as a fallback mechanism.

CP is a hot research topic in computer vision field, which focuses on the design of advanced deep neural network (DNN) models for object detection and data fusion. This work has a different focus to address challenges of CP from the networking perspective. To support highly-efficient CP in a practical network environment, there are several challenges in terms of scalability, dynamics, and uncertainty. *First*, as the CAV number increases, the collective environmental knowledge grows, potentially enhancing the perception performance, at the cost of almost linearly increasing amount of exchanged sensor data. It poses a substantial burden on the limited network resources for communication and computation, potentially leading to high object detection latency. Therefore, scalable engineering solutions and efficient resource management strategies are required for the CP system to effectively scale. *Second*, the CP system is highly dynamic, due to vehicle mobility, dynamic perception workload, changing sensor data quality, and time-varying resource availability. Addressing these dynamics requires adaptive algorithms or protocols that can handle the changes while maintaining timely and reliable information sharing. *Third*, there are several factors that bring performance uncertainty in CP, including the black-box nature of DNNs and the unpredictable sensor data quality variations. These uncertainties hinder performance-aware decisions such as how to select the CAVs to cooperate with and which part of sensor data to share while ensuring the perception performance. The main contributions of this work are summarized as follows.

- We present a data/model co-driven scalable and dynamic CP framework and design two modules for different sce-

This work was supported by the Natural Sciences and Engineering Research Council (NSERC) of Canada.

Kaige Qu and Weihua Zhuang are with the Department of Electrical and Computer Engineering, University of Waterloo, Waterloo, ON, Canada, N2L 3G1 (emails: {k2qu, wzhuang}@uwaterloo.ca).

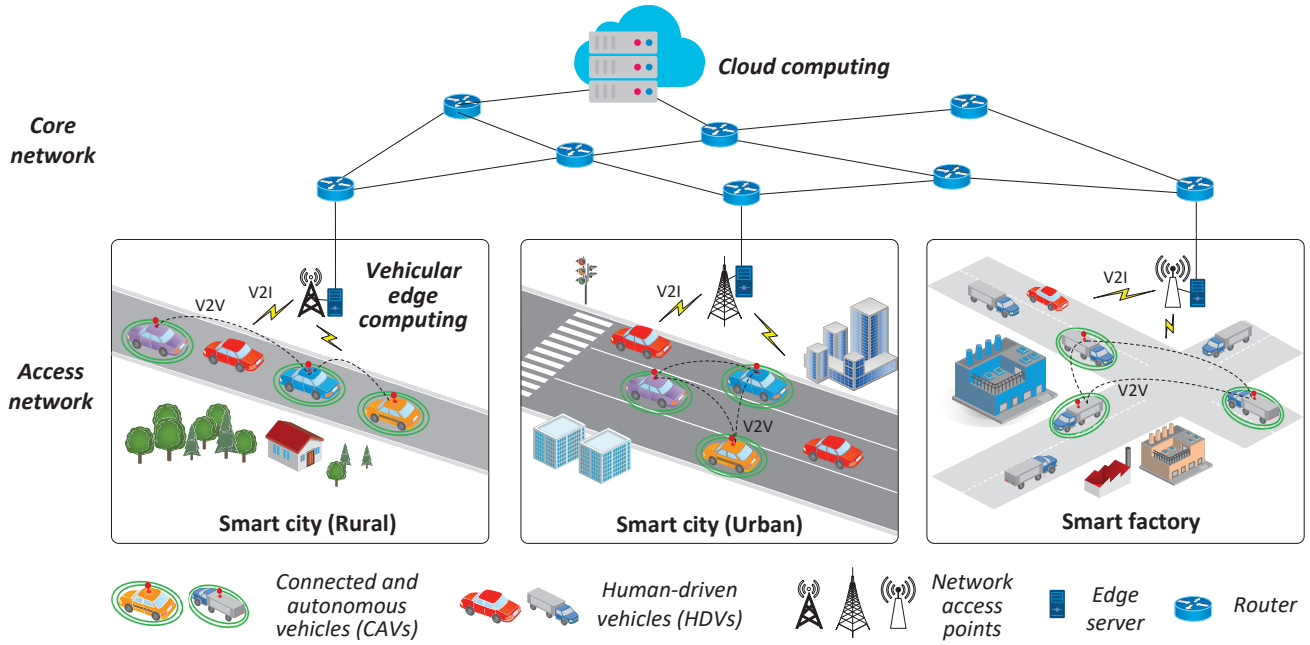


Fig. 1: A V2X and VEC enabled network architecture for supporting the cooperative perception among CAVs.

narios. One is a scalable CP scheme based on a learning-assisted optimization method, addressing the scalability and uncertainty issues. The other is a dynamic CP scheme that focuses on accommodating dynamics in a CP system based on an optimization-assisted learning approach;
- A case study is presented to demonstrate the effectiveness of the proposed CP scheme.

The remainder of this article is organized as follows. The challenges and potential approaches are discussed in Section II. Section III presents the data/model co-driven scalable and dynamic CP framework, with a case study given in Section IV. Conclusions are drawn in Section V.

II. Cooperative Perception: Challenges and Potential Approaches

According to the format of shared sensory information, CP can be categorized into raw level, feature level, and decision level [4], [9]. Fig. 2 illustrates three CP levels between two CAVs. For different CP levels, a trade-off is observed between performance gain and resource efficiency. In the *raw-level CP*, CAVs share and fuse a large volume of raw sensor data, such as camera images, light detection and ranging (LiDAR) point clouds, or radar measurements, which preserve the most fine-grained environmental information and contribute to the highest perception performance gain, while incurring the highest communication overhead. In the *feature-level CP*, CAVs extract, share, and fuse higher-level features based on the raw data. As the features usually capture relevant perception information, feature-level CP reduces the communication resource requirements while still delivering sufficient perception performance. The *decision-level CP* integrates lightweight sensing decisions (e.g., object detection results) of individual CAVs, which is the most communication-efficient but achieves limited

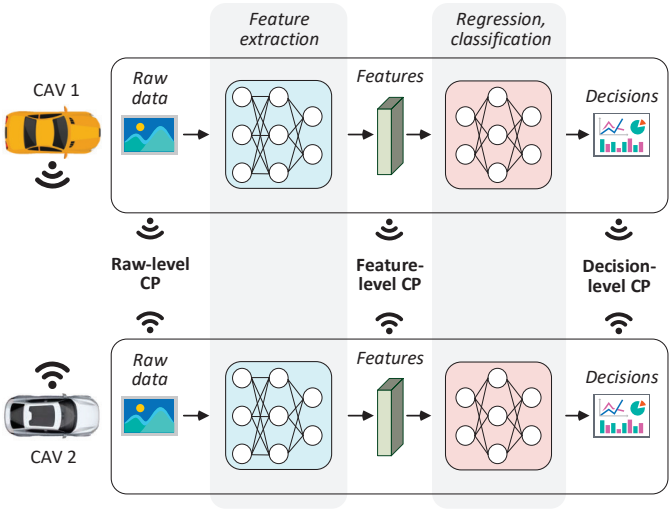


Fig. 2: An illustration of three cooperative perception levels.

perception performance gain. Here, we focus on challenges of raw-level and feature-level CP schemes, where the exchanged sensory data have non-negligible data sizes.

A. Scalability

In a practical network with limited resources, it is challenging for raw-level and feature-level CP schemes to effectively scale when the number of CAVs or objects increases. Typically, the decision-level CP assumes data sharing via radio broadcast, which incurs acceptable delay due to the small data size [10]. However, if broadcasting is directly used in raw-level/feature-level CP, the object detection latency is high, which hinders prompt compensation for the accumulated tracking errors. We consider two potential approaches to

addressing this challenge, through reducing the exchanged data based on *data relevance* or through reducing the object detection delay by exploiting *data parallelism*.

Data Relevance: In raw-level CP, partial raw sensor data can be selected for transmission and processing based on data relevance. For example, when dealing with a union RoI of multiple nearby CAVs, one approach is to divide the union RoI into non-overlapping spatial zones. Each individual CAV then shares the high-resolution partial raw sensor data exclusively for its nearest zone. This strategy, when compared with a basic scheme where all CAVs transmit the full raw sensor data, potentially brings a substantial reduction in the data volume without a remarkable performance degradation. To further reduce irrelevant raw sensor data, a finer-grained raw-level CP scheme enables the selective sharing of partial raw sensor data pertaining to objects within the scene by eliminating background information [5], [6]. However, when the background-foreground ratio in the scene is low, the background removal is insufficient to significantly reduce the data volume. We observe that, in tracking-by-detection, the tracking accuracy varies among objects due to their distinct motion patterns. Based on such property, the data can be further reduced to those pertaining to objects suffering from low tracking accuracy and newly appearing objects [7].

Data Parallelism: As object detection can be parallelized in per-object granularity, parallel computing subtasks can be created, each for the detection of at least one object [11]. By leveraging the computing resources of CAVs and nearby edge servers, the subtasks can be distributed, allowing for parallel processing and reducing the burden on individual CAVs. The delay performance can be improved, enhancing the scalability.

B. Dynamics

A CP system is highly dynamic in multiple dimensions. *First*, the environment is constantly evolving, leading to perception workload and computing demand variations for the CAVs. *Second*, CAVs may join or leave a CP system and change their positions, which affects the overall sensor data availability and quality. *Third*, due to CAVs' mobility and resource competition with other vehicles, the availability, quality, and resources for V2X communications change over time. Such dynamics pose a significant challenge on consistently accurate and real-time perception. There are several potential approaches in handling the dynamics. The CAVs can choose their perception mode in terms of whether or not to cooperate, which CAVs to cooperate with, and what sensor data to share. Moreover, by intelligently allocating network resources among vehicles, the system can adapt to varying network conditions. To deal with the dynamic computing demand while ensuring delay satisfaction, techniques such as dynamic voltage and frequency scaling can be employed at the CAVs to scale up/down the CPU frequency on demand.

C. Uncertainty

There are several factors that lead to performance uncertainty for CP. *First*, due to the inherent black-box nature of DNNs, it is intractable to interpret the feature learning and decision-making processes. It is unclear how the extracted features from different views interact to yield a CP result. This ambiguity introduces performance uncertainty. *Second*, due to unpredictable photometric factors (e.g., illumination, blur) that affect the sensor data quality, the perception performance for a CAV-object pair varies in an unknown manner even if there are no relative displacement or blockage between them [12]. This variability becomes more pronounced when considering the data fusion among multiple CAVs. *Third*, the mobility of CAVs and objects results in constantly-changing viewpoints, further exaggerating the performance unpredictability. A potential approach to addressing the uncertainty is performance estimation based on data-driven techniques.

III. Data/Model Co-Driven Scalable and Dynamic Cooperative Perception

To address the challenges, we present a data/model co-driven framework for performance-aware scalable and dynamic cooperative perception. Model-based methods use mathematical models for decision making, based on prior knowledge or assumptions. They have the generalities for different networking scenarios, but are usually not applicable to complex networks. Data-driven methods are based on machine learning and big data, where data are exploited to derive insights and make decisions, even if the underlying mechanisms are not well understood. However, it usually takes a long time to learn an optimal solution from the data, especially for a complex problem with many decisions. Here, we present two novel CP schemes, to demonstrate how model-based and data-driven methods enhance each other in addressing the scalability, dynamics, and uncertainty issues.

A. Performance-Aware Scalable Cooperative Perception

Consider a group of CAVs under the service coverage of a road-side unit (RSU). Both CAVs and RSU have sensing, computing, and communication capabilities, which cooperatively perceive a union RoI of the CAVs. Fig. 3 show both ego-centric and birds' eye views of two CAVs, each with a 360° LiDAR sensor mounted on the roof, in a simulated scenario. Software-defined networking (SDN) is employed to separate the control and data planes of network devices, allowing for flexible network management. Fig. 4 shows the data and control plane operations, which are elaborated in the following.

1) Data Plane: To enhance the scalability, we consider fine-grained partial raw sensor data selection, transmission, fusion, and processing on a per-object basis, and leverage distributed computation by exploiting the parallelism among computing subtasks associated with each object. A network device, either a CAV or an RSU, may serve as a sensing device, a computing device, or both. To support scalable CP, several key operations are performed among the network devices.

Data Acquisition and Abstraction. Each sensing device collects raw sensor data for objects within its sensing range. A LiDAR sensor generates a 3D point cloud for each environmental scan, which contains a set of 3D location coordinates for observation data points. A data point pertains to a point in space, e.g., due to a LiDAR reflection off a surface at the point.

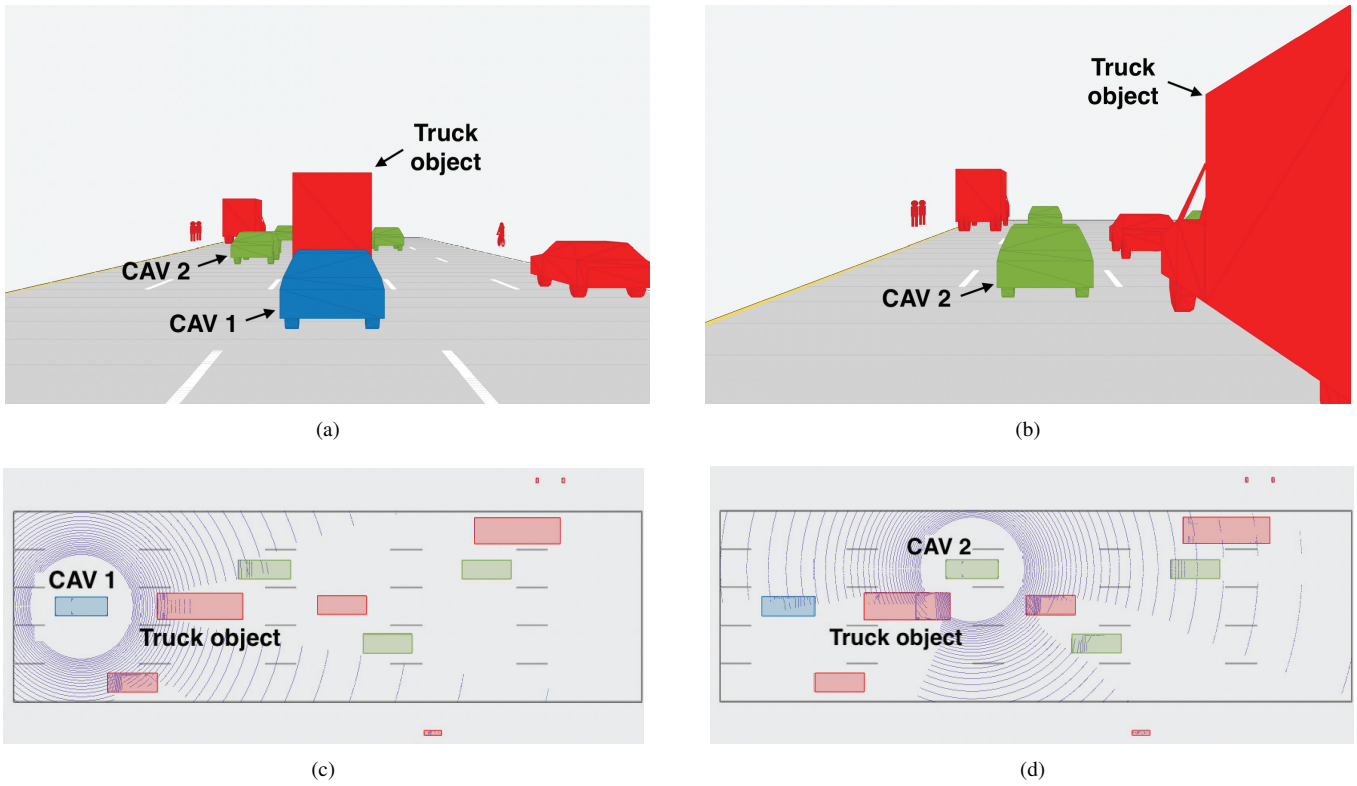


Fig. 3: An illustration of LiDAR-based sensing for two CAVs in a simulated driving scenario. (a) Ego-centric view of CAV 1. (b) Ego-centric view of CAV 2. (c) Bird's eye view of CAV 1. (d) Bird's eye view of CAV 2.

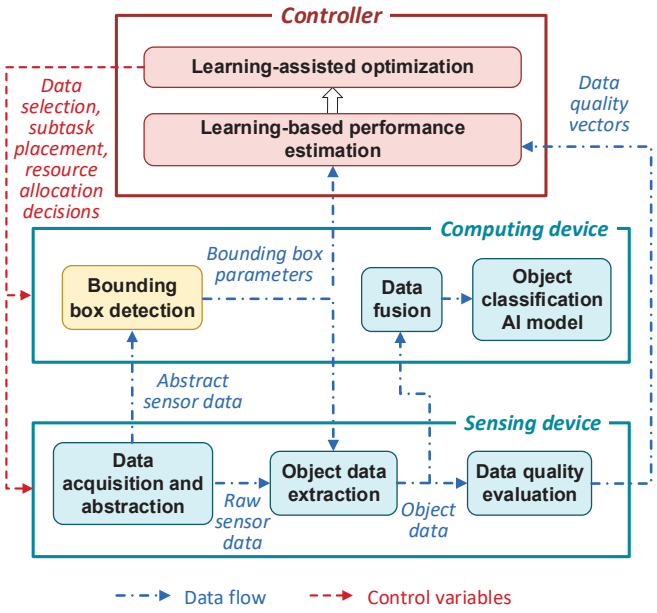


Fig. 4: Data and control plane operations in a performance-aware scalable cooperative perception scheme.

All the raw sensor data acquired at different sensing devices can be aligned in a global coordinate system via coordinate transformation. Each sensing device also generates abstract sensor data, which include the same general information as the raw sensor data but at a lower resolution.

Bounding Box Detection. A 3D bounding box containing an object can be characterized by a 9-tuple, specifying the center coordinates and the region lengths and rotations (i.e., roll, yaw, pitch) along the x, y, and z axes, all in the global coordinate system. A bounding box detection module is placed at one of the computing devices, which receives and fuses the abstract sensor data from all sensing devices. For example, by clustering data points in the fused data, the bounding box parameters of each object can be estimated, which are then distributed to all sensing devices and the controller.

Object Data Extraction and Quality Evaluation. For each object, a sensing device extracts the object data from its raw sensor data based on the estimated bounding box parameters. Depending on the sensing distance, viewpoint, and obstruction, the object data held by different sensing devices differ in their data quality. For LiDAR-based object data, a higher intensity and a more even spatial distribution indicate better data quality. To capture the object data quality, the bounding box is partitioned into M disjoint sub-regions. The numbers of data points located inside all the sub-regions composite a M-dimension data quality vector. Each sensing device evaluates the data quality vectors for different objects, which are then sent to the controller. Fig. 5(a) and Fig. 5(b) show the object data of both CAVs in Fig. 3 for a truck object, with the red solid lines showing the boundaries of $M = 8$ sub-regions. The data points of CAV 1 are concentrated at the back side with low diversity, while the data points of CAV 2 spread more evenly over the front and left sides, providing higher diversity.

Object Data Selection, Transmission, Fusion and Pro-

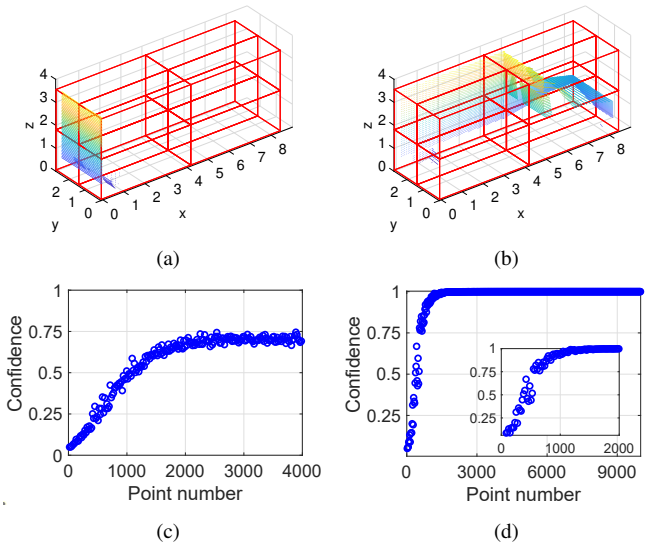


Fig. 5: Impact of object data intensity and distribution on object classification confidence. (a) Object data of CAV 1. (b) Object data of CAV 2. (c) Confidence obtained by using object data of CAV 1. (d) Confidence obtained by using object data of CAV 2.

cessing. After bounding box detection, the object classes are identified by using an AI model pre-configured at each computing device. Our scheme does not assume a specific AI model. Here, we use *Voxelnet* as an example due to its generalization ability and performance [9]. We use an automated driving toolbox in Matlab to simulate diverse scenarios with randomly positioned CAVs and objects. Point clouds are simulated for each scenario, based on which a dataset is created for training the *Voxelnet*. By exploiting data parallelism, each object is associated with one classification subtask. For each subtask, multiple sensing devices can be selected to provide object data to a selected computing device for data fusion and AI model processing, which incurs communication and computing delays.

2) Control Plane: For data selection, we aim to ensure the perception performance with minimum network resources for the subsequent data transmission, fusion and processing stages. In the *control plane*, a supervised learning based performance estimation method is employed to address the performance uncertainty issue. A learning-assisted optimization model is developed to facilitate performance-aware and resource-efficient data selection, subtask placement, and resource allocation.

Learning-Based Performance Estimation. For a trained object classification AI model, a performance metric is required to evaluate the impact of object data quality of each data sample, to facilitate performance-aware data selection [12]. Typical performance metrics, such as accuracy, recall, and precision, evaluate the average model performance for a dataset [9]. Existing studies have demonstrated that, for a well-calibrated classification model, the estimated probability associated with the true class label, i.e., the confidence, reflects its ground-truth correctness likelihood [13]. The positive correlation between confidence and accuracy has been validated in our previous work [14]. Hence, we use the confidence, which differs among different data samples, as the performance metric. Fig. 5(c) and Fig. 5(d) show the confidence of the trained *Voxelnet* for object data of CAVs 1 and 2, respectively, with a data down-sampling ratio from 0.01 to 1[1]. We see in Fig. 5(c) that the confidence increases slowly to below 75% as the point number increases, inferring that increasing the data intensity without improving the data diversity brings limited confidence gain. As the data points of CAV 2 are more evenly distributed, we see in Fig. 5(d) that the confidence increases more rapidly and finally approaches 100%. The object size also affects the confidence, as a smaller object tends to require less data points for accurate classification. Therefore, we profile the confidence as an unknown nonlinear function of the data quality vector and the object dimensions, and train a DNN model to learn the function.

Learning-Assisted Optimization. At a computing device with at least one subtask, a fraction of computing resources should be allocated. For each sensing-computing device pair with data transfer, a fraction of the available radio spectrum should be allocated. For a subtask, data selection from more sensing devices potentially improves the performance with a diminishing marginal gain, at a higher network resource cost for computing and communication. Hence, it is necessary to select the best sensing device group and computing device for each subtask, which satisfies the confidence and delay requirements with minimum resources. We develop a learning-assisted optimization model for the joint data selection, subtask placement, and resource allocation decision. By using the performance estimation DNN model, the confidence of each subtask can be estimated, which should not be below a threshold. The delay of each subtask, which depends on the data size, computing demand, and resource allocation, should not exceed an upper limit. We propose an iterative algorithm, where an outer module and an inner module interact to yield a suboptimal solution. The outer module uses a genetic algorithm to update the binary data selection and subtask placement decisions with confidence satisfaction, while the inner module optimizes the resource allocation decisions with delay satisfaction given the binary decisions in each iteration.

B. Dynamic Cooperative Perception

In the scalable CP scheme, we consider a static snapshot of an autonomous driving scenario for simplicity. The resource availability is not considered as a bottleneck. In practice, the surrounding objects change over time due to vehicle mobility, and the resource availability fluctuates and occasionally becomes the bottleneck. Hence, we also propose a dynamic CP scheme, which focuses on addressing the dynamics issue. A practical mixed-traffic driving scenario is considered, where a cluster of CAVs and human-driven vehicles (HDVs) traverse through an RSU's coverage area and share the radio resources [15]. For simplicity, we assume predetermined CAV pairing and object list for cooperative classification in ideal cases with sufficient resources. Under the network dynamics, we select a subset of CAV pairs for cooperation. The two CP schemes focus on different aspects in different scenarios. They

[1]The code is available at *https://github.com/kaigequ*.

complement each other, providing insights for addressing the challenges in a CP framework.

For object classification, a CAV pair works in either an SP mode by using a default DNN model or a CP mode by using a feature-fusion DNN model. Here, we consider *Voxelnet* as the backbone for both models. In the CP mode, both CAVs extract features from their object data. One CAV transmits the features via vehicle-to-vehicle (V2V) communication to the other CAV, where the features from both CAVs are fused and processed. The results are returned to the sender, with negligible transmission cost. The total computing demand is reduced via CP, potentially enhancing the computing efficiency. For each CAV pair, a delay bound should not be violated in each time slot. Thus, each CAV pair should dynamically switch between the SP and CP modes under the network dynamics, and the radio resources and CPU frequencies for all cooperative CAV pairs should be jointly allocated for delay satisfaction. Define the computing efficiency gain of a CAV pair as the reduced computing energy consumption in comparison with that in the SP mode, which decreases proportionally with the computing demand (in CPU cycle) and the CPU frequency (in cycle/s) squared. Such a gain is equal to zero in the SP mode. For the dynamic perception mode selection and resource allocation, we aim to increase the total computing efficiency gain, while reducing the total switching cost between scheduling different DNN models and satisfying the delay requirement.

An optimization-assisted multi-agent reinforcement learning (MARL) framework is proposed, where each CAV pair is a learning agent. All agents cooperatively learn the perception mode selection *actions* based on dynamic *states* including radio resource availability, perception workload, and channel condition, to maximize an expected discounted total reward in the long run. The *reward* function integrates the total computing efficiency gain and the total switching cost. At each time slot, given the states and actions of all agents, the reward is obtained by solving a resource allocation optimization problem, which determines the CPU frequencies and radio resources among all cooperative CAV pairs, for a maximal total computing efficiency gain with delay satisfaction. If the problem is infeasible, a large negative reward is applied. This optimization-assisted learning approach reduces the action space and problem complexity in comparison with a pure learning approach which directly learns all the decisions.

C. Practical Implementation

Due to vehicle mobility, the communicating RSU for a CAV changes over time, and a CAV may move into an area without RSU coverage. For seamless and reliable control, a cluster head is selected among the vehicles to coordinate the communication, computing, and sensing in the CP system. In the scalable CP scheme, all the control plane functionalities are centralized at an SDN controller which can be placed at the cluster head. In the dynamic CP scheme, the cluster head is responsible for collecting the states and actions from the CAV pairs via a dedicated control channel, to calculate the reward by solving a centralized resource allocation optimization problem. With the coordination of the cluster head, the point clouds of different CAVs can be aligned in a global coordinate system, which is the basis for bounding box detection.

Another practical issue is the heterogeneity among DNN models for CP. With a pre-trained DNN, the input sensing data format should be consistent with the model's requirement. For example, images may need to be resized to fit the input resolution of a DNN. If a multi-modal DNN is used for data fusion between different sensor types, sensing data of required modality should be provided. The proposed CP schemes assume same DNN models among participating CAVs, but they are not limited to a single model. In practice, different DNNs can be pre-trained and maintained by a network controller. The controller deploys at least one DNN at each CAV. For a group of CAVs with CP requests, a protocol should be designed for the negotiation of data format and DNN model. Only the CAVs that reach a consensus can participate in a CP process. Moreover, it is possible to have multiple CP sub-systems using different data formats and DNN models.

IV. Case Study

We present a case study to evaluate the performance of the dynamic CP scheme. Consider a number of $K \in \{2, 3, 4, 5, 6\}$ CAV pairs, moving together with 10 HDVs. A pre-trained *Voxelnet* model is deployed at each CAV. The voxel representation is considered as the feature data. We consider a $1500m$ unidirectional highway segment for each learning episode, during which the vehicle cluster moves for a distance of $1000m$ through an RSU's coverage area. The delay requirement for object detection is $100\ ms$. The maximum CPU frequency for each CAV is $8\ GHz$. The amount of available radio resources for CAVs ranges from 2 to $7\ MHz$, the perception workload for each CAV ranges from 4 to 8, and the transmitter-receiver distance for each CAV pair ranges from 3 to $33\ m$, whose state transitions across consecutive time slots follow predefined Markov chains. To obtain the time-varying channel power gain for each CAV pair, we use the 3GPP NR-V2X 37.885 highway case for the V2V link path loss calculation.

We compare the performance between the MARL solution and three benchmark solutions for dynamic perception mode selection, with results shown in Fig. 6. The benchmarks include a random solution, a solution based on brute force search, and a solution that lets all CAV pairs cooperate if there exists a feasible resource allocation solution or work in the default SP mode otherwise, referred to as *random*, *BF*, and *all* respectively. For the benchmark solutions, given a candidate perception mode selection decision, a resource allocation optimization problem is solved to maximize the instantaneous reward in the current time slot. For the *BF* solution, we conduct a brute-force search among 2^K candidate decisions in each time slot, for a maximum instantaneous reward. As K increases, we see a turning point in the total computing efficiency gain in only the random solution, as the average number of selected cooperative CAV pairs increases linearly with K. This is because the radio resources are shared among more cooperative CAV pairs for feature data transmission, and each cooperative CAV pair should increase the CPU frequency to compensate for the lower average

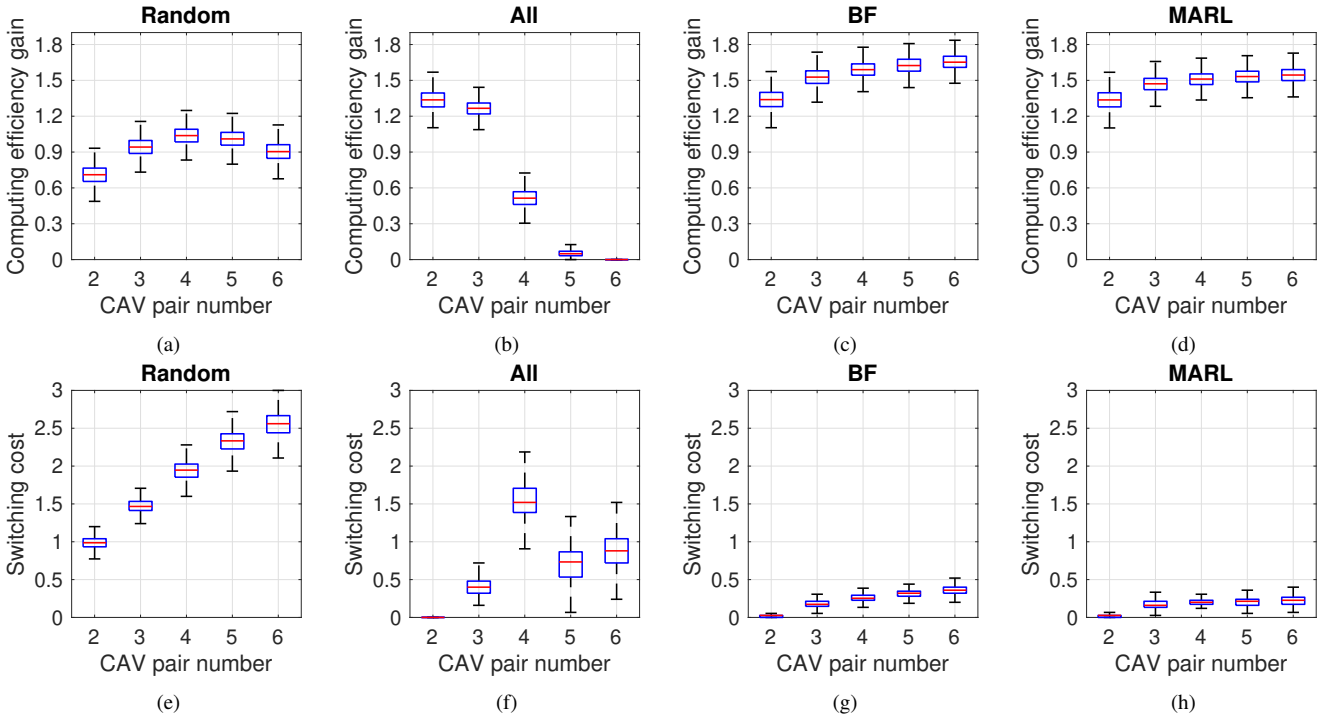


Fig. 6: Performance comparison between the proposed and benchmark solutions for dynamic cooperative perception.

transmission rate, for delay satisfaction. Thus, even though the total computing demand reduction increases in proportion to the number of cooperative CAV pairs, the total computing efficiency gain first increases and then gradually decreases after a turning point. The *all* benchmark achieves a gradually degraded computing efficiency gain when K increases, due to higher chance for infeasible resource allocation among all CAV pairs. However, both the MARL and BF solutions achieve an increasing total computing efficiency gain as K increases, with a small gap. The initial increasing trend is due to the selection of more CAV pairs for cooperation. As K further increases, adding more cooperative CAV pairs does not contribute to more computing efficiency gain, but both solutions are able to select the best group of cooperative CAV pairs among more candidates to further improve the gain. For the total switching cost, we observe an almost linear increasing trend for the random solution, while both the MARL and BF solutions achieve a significant reduction. The MARL solution incurs the lowest total switching cost, due to its ability to learn the long-term optimal perception mode switchings. For the *all* benchmark, as K increases, the dominant solution gradually changes from all CAV pairs working in the CP mode to all CAV pairs working in the SP mode, achieving the highest switching cost at a medium K value.

V. Conclusion

In this article, we focus on the networking perspective of cooperative perception, and present a data/model co-driven framework for performance-aware scalable and dynamic cooperative perception. By exploiting the complementary strengths of data-driven and model-based methods, such a framework facilitates the joint orchestration of sensing, computing, and communication resources, to enhance the network resource efficiency with performance satisfaction. To accelerate the pace of autonomous driving, extensive research efforts are required. For example, we will further explore the adaptive selection of cooperative perception levels, to gain more flexibility in handling the network dynamics and to find a better trade-off between performance gain and resource efficiency.

References

[1] V.-L. Nguyen, R.-H. Hwang, P.-C. Lin, A. Vyas, and V.-T. Nguyen, "Toward the age of intelligent vehicular networks for connected and autonomous vehicles in 6G," *IEEE Netw.*, vol. 37, no. 3, pp. 44–51, 2023.

[2] J. Zhang and K. B. Letaief, "Mobile edge intelligence and computing for the Internet of vehicles," *Proc. IEEE*, vol. 108, no. 2, pp. 246–261, 2019.

[3] X. Shen, J. Gao, W. Wu, M. Li, C. Zhou, and W. Zhuang, "Holistic network virtualization and pervasive network intelligence for 6G," *IEEE Commun. Surv. Tutor.*, vol. 24, no. 1, pp. 1–30, 2021.

[4] Q. Yang, S. Fu, H. Wang, and H. Fang, "Machine-learning-enabled cooperative perception for connected autonomous vehicles: Challenges and opportunities," *IEEE Netw.*, vol. 35, no. 3, pp. 96–101, 2021.

[5] S.-W. Kim, K. Ko, H. Ko, and V. C. M. Leung, "Edge-network-assisted real-time object detection framework for autonomous driving," *IEEE Netw.*, vol. 35, no. 1, pp. 177–183, 2021.

[6] H. Wang, Q. Li, H. Sun, Z. Chen, Y. Hao, J. Peng, Z. Yuan, J. Fu, and Y. Jiang, "VaBUS: Edge-cloud real-time video analytics via background understanding and subtraction," *IEEE J. Select. Areas Commun.*, vol. 41, no. 1, pp. 90–106, 2023.

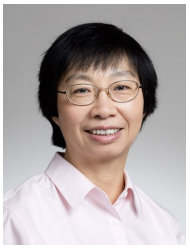
[7] K. Yang, J. Yi, K. Lee, and Y. Lee, "FlexPatch: Fast and accurate object detection for on-device high-resolution live video analytics," in *IEEE Proc. INFOCOM*, 2022, pp. 1898–1907.

[8] C. Li, Q. Luo, G. Mao, M. Sheng, and J. Li, "Vehicle-mounted base station for connected and autonomous vehicles: Opportunities and challenges," *IEEE Wirel. Commun.*, vol. 26, no. 4, pp. 30–36, 2019.

[9] E. Arnold, M. Dianati, R. de Temple, and S. Fallah, "Cooperative perception for 3D object detection in driving scenarios using infrastructure sensors," *IEEE Trans. Intell. Transp. Syst.*, vol. 23, no. 3, pp. 1852–1864, 2020.

[10] ETSI, "Intelligent Transport Systems (ITS); Vehicular Communications; Basic Set of Applications; Analysis of the Collective Perception Service," 3rd Generation Partnership Project (3GPP), ETSI TR 103 562, 2019, v2.1.1.
[11] W. Zhang, Z. He, L. Liu, Z. Jia, Y. Liu, M. Gruteser, D. Raychaudhuri, and Y. Zhang, "Elf: accelerate high-resolution mobile deep vision with content-aware parallel offloading," in *Proc. ACM MobiCom*, 2021, pp. 201–214.
[12] B. Yang, X. Cao, K. Xiong, C. Yuen, Y. L. Guan, S. Leng, L. Qian, and Z. Han, "Edge intelligence for autonomous driving in 6G wireless system: Design challenges and solutions," *IEEE Wirel. Commun.*, vol. 28, no. 2, pp. 40–47, 2021.
[13] C. Guo, G. Pleiss, Y. Sun, and K. Q. Weinberger, "On calibration of modern neural networks," in *Proc. ICML'17*, 2017, pp. 1321–1330.
[14] K. Qu, W. Zhuang, W. Wu, M. Li, X. Shen, X. Li, and W. Shi, "Stochastic cumulative DNN inference with RL-aided adaptive IoT device-edge collaboration," *IEEE Internet Things J.*, vol. 10, no. 20, pp. 18 000–18 015, 2023.
[15] B. Fan, Z. Su, Y. Chen, Y. Wu, C. Xu, and T. Q. S. Quek, "Ubiquitous control over heterogeneous vehicles: A digital twin empowered edge AI approach," *IEEE Wirel. Commun.*, vol. 30, no. 1, pp. 166–173, 2023.

Kaige Qu (S'19−M'21) received the Ph.D. degree in electrical and computer engineering from the University of Waterloo, Waterloo, ON, Canada, in 2021. She received the B.Sc. degree in communication engineering from Shandong University, Jinan, China, in 2013, and M.Sc. degrees in integrated circuits engineering and electrical engineering from Tsinghua University, Beijing, China, and KU Leuven, Leuven, Belgium, respectively, in 2016. From February 2021 to December 2023, she was a Post-doctoral Fellow and then a Research Associate with the Department of Electrical and Computer Engineering, University of Waterloo. Her research interests include connected and autonomous vehicles, network intelligence, network virtualization, and digital twin assisted network automation.



Weihua Zhuang (M'93−SM'01−F'08) received the B.Sc. and M.Sc. degrees from Dalian Maritime University, China, and the Ph.D. degree from the University of New Brunswick, Canada, all in electrical engineering. Since 1993, she has been a faculty member in the Department of Electrical and Computer Engineering, University of Waterloo, Canada, where she is a University Professor and a Tier I Canada Research Chair in wireless communication networks. Her current research focuses on network architecture, algorithms and protocols, and service provisioning in future communication systems. She is the recipient of Women's Distinguished Career Award in 2021 from IEEE Vehicular Technology Society, R.A. Fessenden Award in 2021 from IEEE Canada, Award of Merit in 2021 from the Federation of Chinese Canadian Professionals (Ontario), and Technical Recognition Award in Ad Hoc and Sensor Networks in 2017 from IEEE Communications Society. Dr. Zhuang is a Fellow of the IEEE, Royal Society of Canada (RSC), Canadian Academy of Engineering (CAE), and Engineering Institute of Canada (EIC). She is the President and an elected member of the Board of Governors (BoG) of the IEEE Vehicular Technology Society. She was the Editor-in-Chief of IEEE Transactions on Vehicular Technology (2007-2013), an editor of IEEE Transactions on Wireless Communications (2005-2009), General Co-Chair of IEEE/CIC International Conference on Communications in China (ICCC) 2021, Technical Program Committee (TPC) Chair/Co-Chair of IEEE Vehicular Technology Conference 2017 Fall and 2016 Fall, TPC Symposia Chair of the IEEE Globecom 2011, and an IEEE Communications Society Distinguished Lecturer (2008-2011).