Proceedings of the 4th International Conference on Machine Learning and Automation
DOI: 10.54254/2755-2721/2026.38296

Intermediate Feature Fusion Method for Collaborative Perception of Vehicles and Infrastructure

Yunpeng Chen

Information Engineering College, Chang'an University, Xi'an, China
2023901132@chd.edu.cn

Abstract. The single vehicle intelligent system has inherent limitations in the perception level, such as insufficient over the horizon capability, lack of global perspective and limited stability of extreme scenes. The vehicle-to-infrastructure group intelligent collaborative perception technology provides an effective solution to the above problems through information sharing and fusion among multi-agent. Between data level fusion and decision level fusion, intermediate feature fusion achieves the balance between communication efficiency and perception accuracy by realizing the interaction and integration of multi-source information at the feature level, which has become a research hotspot in the field of vehicle-to-infrastructure collaborative perception. This paper reviews the development process of intermediate feature fusion, and reveals the evolution logic of collaborative perception from "result sharing" to "feature collaboration" from the paradigm shift from late collaboration to medium-term collaboration; The three core technologies of feature extraction and representation, spatial feature alignment, feature compression and communication optimization are analyzed; Four typical paradigms of global fusion, sparse fusion, hierarchical fusion and attention fusion and their application scenarios are summarized; The research hotspots of transformer architecture, task oriented communication optimization and the key challenges faced by system deployment are discussed. On this basis, the future directions of generative AI empowerment, evolution from collaborative reasoning to collaborative generation, and standardized deployment are prospected, in order to provide reference for the in-depth research and engineering implementation of intermediate feature fusion technology.

Keywords: Collaborative perception, Intermediate feature fusion, Internet of vehicles, Deep learning, Distributed reasoning

1. Introduction

With the large-scale development of high-level automatic driving, the coverage, accuracy and robustness of environmental perception are the core premise to ensure driving safety and traffic efficiency. Restricted by physical sight distance, building shelter, extreme weather and other factors, single vehicle intelligent perception has inherent bottlenecks, which makes it difficult to independently cope with the full-time perception needs of complex urban scenes. Relying on the Internet of vehicles, vehicle-to-infrastructure collaborative perception can achieve multi-agent

© 2026 The Authors. This is an open access article distributed under the terms of the Creative Commons Attribution License 4.0 (https://creativecommons.org/licenses/by/4.0/).

DOI: 10.54254/2755-2721/2026.38296

information interaction and deep integration, which can break through the boundary of single vehicle perception, fill the blind spot, and improve the environmental adaptability. It has become the core technology direction in the field of intelligent transportation.

According to the integration stage, vehicle-to-infrastructure collaborative perception includes three types of technical paths: data level, feature level and decision level [1]. Data level fusion shares the original sensor data, and the information is complete, but the bandwidth and synchronization requirements are extremely high, so it is difficult to land. Decision level fusion only transmits the detection results, with low communication overhead but obvious information loss, which is difficult to correct the single vehicle perception error, and the performance improvement is limited. Intermediate feature fusion realizes information integration in the feature layer of neural network, takes into account semantic information retention and transmission efficiency, and achieves a better balance between perception accuracy and communication cost.

At present, the academic circle has formed a multi-dimensional research accumulation around the vehicle-to-infrastructure collaborative perception. Some studies have combed the overall technical framework from the perspective of swarm intelligence collaboration, and clarified the core advantages of multi-agent collaboration [1]. Some works also reviewed the evolution of end-to-end architecture from the perspective of multimodal integration and multi node collaboration, and pointed out that collaboration is extending to the decision-making level [2]. The overview from the perspective of vehicle-infrastructure-cloud integration summarizes the three-tier collaborative architecture and proposes the technical route of AI empowerment [3]. From the group intelligence collaboration framework to multi node collaboration, and then to the vehicle-infrastructure-cloud integration architecture, the existing reviews show a gradual trend, but none of them specifically combs the core paradigm of intermediate feature fusion, which is the work of this paper.

At the level of specific integration methods, relevant research has also formed a series of breakthroughs. Aiming at the problem of point cloud feature aggregation accuracy, some researches have proposed an adaptive weight fusion mechanism, which improves the effect of multi node aggregation [4]. For real communication packet loss scenarios, relevant work has built a feature repair and attention fusion framework to alleviate the performance degradation caused by data loss [5]. Aiming at the problem of heterogeneous node fusion between vehicle and road, V2X-ViT series methods enhance the robustness of the system to pose error and time delay through heterogeneous attention and multi-scale mechanism [6]. In the downstream task of collaborative sensing, the vehicle-to-infrastructure collaborative heavy recognition improves the robustness of cross node target matching by fusing multiple information of image and trajectory through lightweight twin network [7]. The latest review further clarifies that bandwidth constraints, heterogeneous alignment and spatio-temporal synchronization are the core bottlenecks for large-scale implementation [8].

Although the existing achievements are rich, the systematic carding of the core paradigm of intermediate feature fusion is still insufficient. Most reviews only briefly mention it as a branch of collaborative perception, but have not yet clarified its development context, technical system, integration paradigm and landing challenges. Relevant achievements are scattered in each branch field, which is not conducive to researchers' quick grasp of the full picture of the field. In view of the above situation, this paper systematically summarizes the intermediate feature fusion methods for the vehicle-to-infrastructure collaborative perception scene. Firstly, this paper combs the technological development process of the integration of intermediate features, and reveals the paradigm shift logic of the evolution from late collaboration to medium-term collaboration; Secondly, the research progress of three core technology modules, feature extraction and representation, spatial feature alignment, feature compression and communication optimization, is

DOI: 10.54254/2755-2721/2026.38296

deeply analyzed; Then, four typical fusion paradigms and their applicable scenarios are summarized: global fusion, sparse fusion, hierarchical fusion and attention fusion; On this basis, the research hotspots in the current field and the key challenges faced by the system deployment are analyzed; Finally, the future development directions of generative AI empowerment, collaborative paradigm evolution and standardized deployment are prospected, in order to provide reference for the in-depth research and engineering implementation of related technologies.

2. The development of intermediate feature fusion

2.1. Early exploration stage

The concept of intermediate feature fusion originates from the development of multi-sensor fusion technology. In the early research of vehicle-to-infrastructure collaborative sensing, researchers mainly focused on the late collaborative scheme, and each agent only shared the detection results after completing the sensing task independently. Although this method has low communication overhead and flexible deployment, it is difficult to correct the perception errors at the single end, and the improvement of perception performance is limited [1]. With the extensive application of deep learning in automatic driving, researchers began to explore information interaction at the feature level. V2vnet is an early representative work in this direction, which realizes joint perception and prediction by sharing intermediate feature representation between vehicles. The f-cooper framework proposes two strategies, voxel feature fusion and spatial feature fusion, which effectively integrate the lidar features of multiple vehicles [1].

2.2. Paradigm shift stage

The transformation from late collaboration to mid-term collaboration marks the paradigm transition of collaborative perception from "result interaction" to "feature fusion". The core driving force of this transformation is that the intermediate features extracted by deep neural network not only retain rich environmental semantic information, but also significantly reduce the amount of data transmission through feature compression [1]. Coopernaut realizes end-to-end cooperative driving based on V2V communication, encodes sensor data into compact messages and aggregates features in the self vehicle coordinate system [1]. Where2comm introduces spatial confidence map to achieve selective transmission of key area features, which significantly improves the communication efficiency [2]. In China, V2X-ViT designed a heterogeneous self attention module and a location delay sensing mechanism to solve the problem of feature alignment of heterogeneous sensing nodes in the vehicle-to-infrastructure collaboration scene, effectively alleviating the problem of feature dislocation caused by type differences and synchronization errors [1].

3. Core technology of intermediate feature fusion

3.1. Feature extraction and representation

Feature extraction is the basic link of intermediate feature fusion, and its goal is to extract compact and semantic feature representation from the original sensor data [9]. In terms of point cloud feature extraction, the mainstream methods include voxel based methods such as VoxelNet, point set based methods such as PointNet and graph based methods. Pointpillars greatly improves the computational efficiency by encoding point clouds as pseudo image features. In terms of image feature extraction, convolutional neural networks such as RESNET effectively extract semantic features through jump

DOI: 10.54254/2755-2721/2026.38296

connection, and CoBEVT cooperatively generates BEV features through sparse transformer, realizing the unified representation of multi vehicle and multi camera perception. In terms of multimodal feature fusion, HM vit is the first unified multi-agent heterogeneous modal collaborative sensing framework, which supports the collaborative prediction of 3D objects by different numbers and types of agents [1]. In the multi information fusion at the task level, the combination of image and trajectory information is also concerned. The vehicle-to-infrastructure collaborative heavy recognition algorithm based on lightweight twin network and weight distribution fusion function effectively improves the problem of insufficient robustness of a single information element [7].

3.2. Spatial feature alignment

The core challenge of spatial feature alignment is to establish accurate pixel level or voxel level correspondence between the feature maps of different agents. The early solution relied on high-precision gps/imu for coordinate transformation, but it was not reliable in urban canyon, tunnel and other scenarios with unstable GNSS signals. To enhance robustness, researchers have proposed a variety of methods: learning error correction through the pose regression module, using the semantic labels of 3D points to find the point-to-point correspondence, designing the pose error correction module to achieve accurate alignment of the misaligned parts of the feature level, and performing accurate spatial position calibration through sparse uniform features [1]. Aiming at the misalignment in the time dimension, the combination of heterogeneous multi-agent self attention and multi-scale window self attention can simultaneously capture the spatial feature relationship and temporal dependence, and effectively alleviate the asymmetry problem caused by asynchronous information [1].

3.3. Feature compression and communication optimization

Communication bandwidth limitation is one of the core engineering constraints faced by the vehicle-to-infrastructure cooperative sensing system [3]. Although the integration of intermediate features significantly reduces the communication load, the transmission of high-dimensional BEVfeatures and point cloud features may still cause network congestion in multi vehicle intensive scenarios [8]. Compressing the necessary information at each stage through a learnable mechanism is the direction of early exploration. Learning based communication group construction and communication timing decision significantly reduce bandwidth consumption while maintaining accuracy [1]. The introduction of the spatial confidence map method provides a new idea for communication optimization- by evaluating the importance of different spatial regions to perception tasks, only selectively transmitting the characteristics of high confidence regions, achieving a better efficiency accuracy balance than uniform compression [2].

4. Comparison of fusion paradigms and application case analysis

4.1. Comparison of fusion paradigms

According to the fusion granularity and interaction mode, the intermediate feature fusion can be summarized into four typical paradigms. The global feature fusion sends the complete feature map to the central node for splicing, which is simple to implement but has high communication overhead; Sparse feature fusion is based on key points or region selective interaction, which has low communication overhead and good scene adaptability; Hierarchical feature fusion constructs the

DOI: 10.54254/2755-2721/2026.38296

perception results from coarse to fine through multiple rounds of communication, which has high accuracy but high computational complexity; Attention feature fusion uses attention mechanism to dynamically calculate feature weights, which has the best flexibility and comprehensive performance, but the computational complexity is high. The four paradigms show different trade-offs between communication overhead, computing load and fusion accuracy. The actual deployment needs to be flexibly selected according to the scenario requirements and resource constraints.

4.2. Typical application cases

Vehicle-to-vehicle collaborative perception realizes continuous perception and motion prediction in occluded scenes through the spatio-temporal fusion of multi vehicle BEV features. The core value is to supplement the blind area of self vehicle perception by using the observation of different perspectives of adjacent vehicles. In typical scenes such as car following, meeting, and intersection traffic, adjacent vehicles Exchange semantic features in the middle layer, which can effectively eliminate the visual field occlusion caused by vehicles in front and roadside buildings, and perceive crosswise pedestrians and non motor vehicles in advance. Aiming at the problem of communication packet loss in actual driving, the mechanism of lossy perception repair and uncertain attention can further enhance the robustness of feature fusion and ensure the perceptual stability in non ideal communication environment [5]. This kind of method exchanges small communication overhead for blind area completion, but the coupling of pose error and time delay in high-speed dynamic topology is its main challenge.

Vehicle-to-infrastructure collaborative perception uses the global perspective of roadside infrastructure to enable vehicles. As the first large-scale real scene vehicle-to-infrastructure collaborative data set, DAIR-V2X provides VIC3D benchmark. Roadside perception has a wide field of vision, is not affected by vehicle movement, and can provide stable global context information. Roadside units are usually deployed at key points with limited sight distance, such as intersections and sharp turns, to cover areas beyond the vehicle's physical sight distance from a high-level top-down perspective, providing over the horizon target information and risk warning for vehicle perception. Aiming at the problems of hardware heterogeneity and pose deviation between roadside and on-board equipment, heterogeneous multi-agent attention, delayed sensing coding and space-time correction mechanisms work together to realize the feature alignment and fusion of two types of nodes [6].At the target correlation level, the re recognition algorithm based on the fusion of image and trajectory information provides a feasible scheme for the reliable matching of the same target across nodes [7]. The core technical contribution of this scenario lies in the joint correction of view difference and asynchronous delay by heterogeneous attention and delay aware coding [6].

Vehicle-infrastructure-cloud collaborative sensing realizes effective division of labor between time delay sensitive and computing intensive tasks through a three-tier Cloud Architecture, namely edge cloud, regional cloud and central cloud. Holovic's large scale multi view multi-sensor intersection data set provides an important data base for it [1]. Among them, the edge cloud undertakes the real-time perceptual data processing of roadside and vehicle ends nearby, and supports low delay security applications such as collision warning and traffic guidance; At the specific task level, the existing work is based on the vehicle-infrastructure-cloud multi-source information fusion framework, combined with Bi-LSTM network and XGBoost algorithm to achieve accurate recognition of vehicle driving intention [10]. The regional cloud and the central cloud are responsible for computing intensive tasks such as wide area traffic situation awareness and awareness model iterative training to achieve the overall optimization of global awareness. This layered collaboration mode takes into account the real-time performance of end-to-end services and

DOI: 10.54254/2755-2721/2026.38296

the computing power advantage of the cloud, and can support the implementation of a wider range and higher-level vehicle-to-infrastructure collaborative automatic driving applications [3]. Multi source features are uploaded level by level after being compressed by edge cloud. Cross layer delay accumulation and compression semantic loss are the unique challenges of this scenario.

5. Research hotspots and technical challenges

5.1. Research hotspots

At present, the research hotspots in the field of intermediate feature fusion mainly focus on the following three directions.

First, the heterogeneous fusion architecture based on transformer has become the mainstream technology route. With its powerful global modeling and dynamic interaction capabilities, transformer has become the core architecture of multi-agent feature fusion. V2X-ViT captures the dynamic interaction between vehicle and road nodes through heterogeneous multi-agent attention modules [6]; heterogeneous graph attention networks learn the differentiated associations between different types of agents [1]; V2VFormer introduces the spatial-channel dual attention mechanism, marking the evolution of collaborative perception from simple feature superposition to the direction of fusing context information [2]. Aiming at the problem that the computational complexity of transformer increases with the square of the number of agents, knowledge distillation, model quantization and sparse attention optimization are the main technical means to achieve lightweight deployment.

Second, the task oriented communication optimization strategy continues to deepen. This direction aims to establish a direct mapping between communication behavior and perceived task performance, breaking the traditional mode of "uniform transmission". Where2comm evaluates the contribution of different regions to perception tasks through spatial confidence map, and only selectively transmits the characteristics of high-value regions; How2comm further introduces the flow guidance delay compensation mechanism to achieve real-time dynamic matching between communication content and task requirements [1]. Existing studies have shown that such strategies can reduce the communication load by more than 50% without sacrificing the perception accuracy, and significantly improve the engineering feasibility of the system. This feasibility has been confirmed in an engineering application case: the vehicle cooperative control based on 5Gimproves the average traffic efficiency of intersections by 35%, and the response time of accident identification is shortened from 30S to less than 6s [11].

Thirdly, the robustness enhancement in non ideal environment has been widely concerned. For the non ideal factors such as communication packet loss and pose deviation in the real deployment, the academic community has carried out a series of targeted research. For lossy communication scenarios, the lossy perception repair network and uncertain attention mechanism can repair and weight the damaged features to alleviate the performance degradation caused by data loss [5]; Aiming at the pose error and time delay of heterogeneous nodes on the road, the heterogeneous multi-agent attention, delay-aware encoding and spatio-temporal correction mechanisms work together to achieve feature alignment and fusion between vehicle and infrastructure nodes [6].

5.2. Technical challenges

Although the intermediate feature fusion technology has made significant progress, it still faces four key challenges from academic research to large-scale engineering.

DOI: 10.54254/2755-2721/2026.38296

First, the communication bandwidth constraint in high-density traffic scenarios is prominent. When the vehicles are densely distributed in the vehicle-to-infrastructure coordination scene, the parallel transmission of a large number of high-dimensional BEVfeatures and point cloud features is very easy to cause network congestion, and the existing compression and selective transmission methods are still difficult to take into account the accuracy and real-time in the extreme high-density scene. Even with 5G support, where end-to-end latency can be controlled within 10–20 ms, the per-vehicle data volume of existing lightweight schemes still reaches 128 kbps [11]; sustained transmission of uncompressed high-dimensional features is clearly unaffordable in terms of bandwidth.

Second, it is difficult to guarantee the position and attitude error and space-time synchronization in complex environment. Existing spatial alignment methods are highly dependent on GNSS positioning accuracy. In urban canyon, tunnel, underground parking lot and other signal occlusion or failure scenarios, positioning deviation will lead to feature dislocation, while the accuracy of unsupervised autonomous alignment method is still not enough to support the high-level perception needs. The downstream tasks are also affected by this. The real road acquisition data are inevitably disturbed by sensor noise and positioning error. The generalization ability of intention recognition model trained on conventional driving data for sudden scenes still has significant room for improvement [10].

Third, the ability of asynchronous information compensation under large delay jitter is insufficient. There are significant dynamic fluctuations in the communication delay in the actual Internet of vehicles environment. The existing delay compensation methods are mostly designed for fixed or small-scale delay, and have limited adaptability to large jitter. Asynchronous features are easy to lead to position deviation and missed detection of target detection.

Fourth, the generalization and scalability of the model are still insufficient. At present, most fusion models are trained under the fixed number of agents and scene types, and their performance will be significantly degraded in the face of different vehicle types, different roadside equipment configurations and the dynamic number of vehicles; At the same time, the computing and communication overhead increases rapidly with the number of agents, which is difficult to support the global coordination of large-scale road networks. Such constraints have appeared in specific systems: when the driving intention recognition method based on vehicle-infrastructure-cloud collaborative perception processes more than 20 target vehicles at the same time, the single recognition takes more than 100ms, and the real-time performance is difficult to guarantee [10].

6. Future trends

The integration of generative AI and collaborative perception is promoting the evolution of vehicle-infrastructure-cloud integration system from "data driven" to "generative intelligence" [3]. The diffusion model and world model can generate long tail scene data to make up for the lack of real data, and the closed-loop simulation platform can realize the complete closed-loop test from scene generation to decision control [3]. From collaborative reasoning to collaborative generation is another important direction: at the perception level, vehicles can actively generate blind area targets based on context; At the prediction level, multi-agent generates consistent future prediction by sharing motion features; At the decision level, the end-to-end collaborative framework has demonstrated the feasibility of directly generating planning trajectories. In terms of standard system and large-scale deployment, China has built six standard frameworks for vehicle-infrastructure-cloud integration system [3], and it is necessary to further establish a unified multi-source feature representation specification, cross platform interface standards and evaluation criteria. In addition,

DOI: 10.54254/2755-2721/2026.38296

the credibility verification of the fusion model after it goes online also lacks mature means: the real operation data is lack of annotation, and the deployed model is difficult to access. How to evaluate the credibility of the reasoning results in real time without annotation is directly related to the system operation safety. The credibility quantification method based on complex integrity constraints provides a feasible idea for this problem, which can find untrusted models in time without labeling, and provide the basis for model retraining and redeployment [12]. At the engineering verification level, in the application case of vehicle cooperative control based on 5g, the average traffic efficiency at intersections was increased by 35%, the congestion duration in morning rush hours was reduced by more than half, and the response time of accident identification was shortened from 30s to less than 6S [11], indicating that the low delay cooperative technology has real traffic benefits. The large scale deployment of 5g-v2x and the pilot implementation of " vehicle-infrastructure-cloud integration" are promoting the intermediate feature fusion technology from research to application.

7. Conclusion

Taking the fusion of intermediate features as the main line, this paper reviews the evolution process of this technology from late collaboration to medium-term collaboration, and explains the internal motivation of paradigm change; The representative schemes and their limitations are summarized around three modules: feature extraction and representation, spatial feature alignment, feature compression and communication optimization; The four paradigms of global, sparse, hierarchical and attention fusion are summarized, and the applicable conditions are discussed in comparison with the three scenarios of vehicle-to-vehicle, vehicle-to-infrastructure and vehicle-infrastructure-cloud; On this basis, the research hotspots and four landing challenges are analyzed, and the directions of generative AI, collaborative generation and standardized deployment are prospected.

As the core technology paradigm of vehicle-to-infrastructure collaborative perception, intermediate feature fusion achieves a good balance between communication efficiency and perception accuracy, and has become one of the key paths to support the large-scale implementation of high-level autonomous driving.

In terms of the development process, the intermediate feature fusion has experienced a paradigm transition from late collaboration to mid-term collaboration. The core breakthrough is to improve the collaborative perception from "result sharing" to "feature collaboration", so that multi-agent can realize the deep complementarity and fusion of information in the intermediate semantic layer. The driving force of this evolution comes from the improvement of deep learning feature expression ability and the dual pursuit of communication efficiency and perception accuracy.

In terms of core technologies, feature extraction and representation, spatial feature alignment, feature compression and communication optimization constitute three major technical pillars. BEV feature representation has become the mainstream paradigm of unifying multi-source heterogeneous features. Selective communication based on spatial confidence graph achieves an effective balance between efficiency and accuracy. Transformer architecture shows unique advantages in dynamic feature interaction. However, the existing technology still has obvious shortcomings: the Pareto optimal boundary of feature compression is not clear, the generalization ability of pose compensation method for no error distribution is limited, and the adaptability of time series asynchronous modeling to large delay jitter needs to be improved.

In terms of application practice, the three scenarios of vehicle-to-vehicle collaboration, vehicle-to-infrastructure collaboration and vehicle-infrastructure-cloud collaboration have formed their own distinctive technical solutions, but there is still a significant gap from academic research to

DOI: 10.54254/2755-2721/2026.38296

engineering implementation. The existing methods are mostly verified under ideal communication conditions, and are not robust to non ideal factors such as bandwidth fluctuations and positioning errors in the real deployment environment.

Looking forward to the future, the introduction of generative AI will promote the transition of intermediate feature fusion from "collaborative reasoning" to "collaborative generation", and the integration of communication perception computing will break through the current bottleneck from the system level. The key research directions include: fusion of rule constraint and data-driven hybrid intelligent framework to enhance interpretability and security controllability; Developing weakly supervised or unsupervised pose calibration methods to reduce the dependence on labeled data; Establish a unified multi-source feature representation specification and cross platform interface standard to provide institutional guarantee for large-scale deployment. With the breakthrough in the above direction, the intermediate feature fusion technology will provide a more solid technical base for the large-scale application of high-level automatic driving.

References

[1] Hong, D. Y., Zhang, H., Zhang, Y. Y., et al. (2025). A review of collaborative intelligent perception methods for vehicle-road systems. *Chinese Journal of Intelligent Science and Technology, 7*(2), 144-164.

[2] Feng, H., Hou, W. T., Qu, K. G., et al. (2026). Research on multimodal and multi-node cooperation in autonomous driving. *AI-VIEW*, (1), 37-48.

[3] Ran, B., Deng, Y., Rui, Y. K., et al. (2026). Key technologies of vehicle-infrastructure-cloud integrated system enabled by artificial intelligence: Review and prospect. *China Journal of Highway and Transport, 39*(5), 158-189.

[4] Qiao, D., & Zulkernine, F. (2023). Adaptive feature fusion for cooperative perception using LiDAR point clouds. In *2023 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)* (pp. 1186-1195).

[5] Li, J., Xu, R., Liu, X., et al. (2023). Learning for vehicle-to-vehicle cooperative perception under lossy communication. *IEEE Transactions on Intelligent Vehicles, 8*(4), 2650-2660.

[6] Xu, R., Chen, C. J., Tu, Z., et al. (2025). V2X-ViTv2: Improved vision transformers for vehicle-to-everything cooperative perception. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 47*(1), 650-662.

[7] Qian, M., Geng, K. K., Yin, G. D., et al. (2024). Vehicle-to-infrastructure collaborative re-identification algorithm based on multiple information fusion. *Journal of Mechanical Engineering, 60*(16), 231-240.

[8] Huang, T., Liu, J., Zhou, X., et al. (2025). Vehicle-to-everything cooperative perception for autonomous driving. *Proceedings of the IEEE, 113*(5), 443-477.

[9] Yi, X. Y., Rui, Y. K., Ran, B., et al. (2024). Research progress and prospect of vehicle-to-infrastructure collaborative sensing technology. *Strategic Study of CAE, 26*(1), 178-189.

[10] Li, H. Q., Lei, Y. C., Dai, Y. Y., et al. (2025). Vehicle driving intention recognition method based on vehicle-infrastructure-cloud collaborative perception. *Journal of Electronic Measurement and Instrumentation, 39*(11), 98-107.

[11] Wang, W. L. (2026). Design of vehicle cooperative control algorithm in intelligent transportation system based on 5G. *China Broadband, 22*(4), 13-15.

[12] Hu, W. T., Jiang, D. W., Wu, S., et al. (2022). Complex integrity constraint discovery: Measuring trust in modern intelligent railroad systems. *Journal of Zhejiang University-SCIENCE A (Applied Physics & Engineering), 23*(10), 832-837.