Beam Design and Cooperative Positioning Algorithm for Vehicle-mounted Communication and Sensing Integrated System

Zhuo Chen
School of Electronic and Information Engineering, University of Science and Technology Liaoning, Anshan, 114051, Liaoning, China
Corresponding Author: Zhuo Chen

Abstract: With the rapid development of intelligent connected vehicles (ICVs) and 6G vehicular networks, integrated sensing and communication (ISAC) has become the core enabling technology for high-level autonomous driving. However, the practical deployment of vehicular ISAC systems is severely restricted by three key challenges: insufficient dynamic beam adaptation in high-mobility scenarios, degraded cooperative localization accuracy in non-line-of-sight (NLOS) environments, and the difficulty of collaborative optimization for communication and sensing performance. To address these issues, this paper proposes a deep learning-based joint beam design and cooperative localization algorithm framework for vehicular ISAC systems. Specifically, we first develop a Transformer-based time-varying channel prediction and multi-objective beam optimization network, which realizes the joint optimization of communication spectral efficiency and sensing measurement accuracy. Then, a graph neural network (GNN)-driven multi-node cooperative localization model is designed to suppress NLOS errors using high-precision sensing parameters from optimized beams, with a closed-loop optimization mechanism between beam design and localization established. Simulation results show that compared with traditional separate benchmark algorithms, the proposed scheme improves communication spectral efficiency by 22.3%, sensing angle measurement accuracy by 31.5%, and reduces the root mean square error of vehicular cooperative localization by 46.8%, with end-to-end inference latency within 1ms. The proposed method meets the real-time and reliability requirements of high-dynamic vehicular scenarios, providing an effective technical solution for the engineering application of vehicular ISAC systems.

Keywords: Communication and sensing integration; Vehicle-mounted beam design; Cooperative positioning; Deep learning; Intelligent Connected Vehicles

1. Introduction

With the accelerated iteration of intelligent connected vehicles and 6G vehicle networking technologies, L4 and above-level high-level autonomous driving has imposed rigid demands on the vehicle system for low-latency high-reliability communication, high-precision environmental perception, and centimeter-level positioning capabilities [1]. The traditional discrete architecture of vehicle communication and perception systems has inherent defects such as spectrum resource competition, high hardware redundancy, and low information coordination efficiency [2], which have

become difficult to meet the strict performance requirements of high-level autonomous driving [3]. The integrated communication and perception (ISAC) technology, through sharing spectrum resources and radio frequency hardware platforms, realizes the deep integration and performance coordination of communication and perception functions, has become the core supporting technology of 6G vehicle networks [4]. Beamforming design and multi-node collaborative positioning are the two core modules that ensure the dual functions of communication and perception and achieve core applications of autonomous driving in the ISAC system. The performance of these two modules directly determines the practical level of the ISAC system [5]. Conducting joint optimization research on beam design and collaborative positioning for ISAC in high-dynamic vehicle scenarios has important theoretical value and engineering application significance for promoting the large-scale implementation of high-level autonomous driving [6].

Currently, scholars at home and abroad have conducted extensive research on beam design and collaborative positioning algorithms for vehicle communication and perception integration systems [7]. In the beamforming field, traditional integrated beam design methods based on convex optimization and game theory can achieve trade-off optimization of communication and perception performance in static or quasi-static scenarios. However, in vehicle high-speed mobile scenarios, they face problems such as high complexity of iterative optimization for time-varying channels, large adaptation delay for beam tracking, and insufficient anti-interference ability in dense multi-user scenarios [8], which are difficult to meet the millisecond-level real-time requirements of vehicle scenarios. In the collaborative positioning field, traditional positioning algorithms based on measurement parameters such as angle of arrival and flight time can achieve high positioning accuracy in line-of-sight scenarios, but in typical non-line-of-sight scenarios such as urban canyons and underground tunnels, multipath effects lead to significant degradation in measurement accuracy [9]. The robustness and generalization ability of multi-node collaboration are insufficient. Although existing studies have introduced machine learning techniques such as deep learning into the beam design or positioning optimization of vehicle ISAC, most of them only perform independent optimization for a single module and lack a closed-loop joint optimization mechanism for beam design and collaborative positioning, unable to use positioning results to iteratively optimize beam configuration, and unable to simultaneously consider the coordinated improvement of communication and perception performance and positioning accuracy. There are still significant shortcomings in the real-time performance and robustness of algorithms in high-dynamic vehicle scenarios, which have become the core challenges restricting the large-scale implementation of vehicle ISAC systems. In response to the core challenges faced by the ISAC system, this paper takes high-dynamic vehicle scenarios as the core research object and conducts research on the joint optimization algorithm for beam design and collaborative positioning of the integrated communication and perception system.

2. Methods

2.1 Modeling of Vehicle-Integrated Sensing System and Evaluation Index System

This paper considers a single-cell multi-user vehicle-based ISAC system. The roadside base station (BS) is equipped with an M-element uniform linear array (ULA), and within the cell, there are K intelligent connected vehicles, each equipped with an N-element ULA [10]. The system supports cooperative transmission between vehicle-to-infrastructure (V2I) and vehicle-to-vehicle (V2V), operates in the millimeter wave frequency band, adopts the time division duplex (TDD) mode, and

achieves downlink communication and target echo perception simultaneously within the same time slot by sharing spectrum resources, radio frequency links, and transmission waveforms. The system topology and signal model are shown in Figure 1.

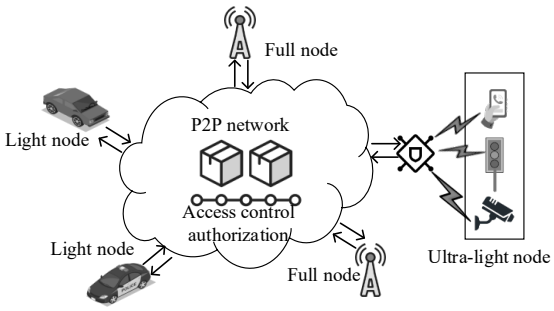

Figure 1: System topology and signal model.

The transmitted baseband signal of the base station can be expressed as:

$$s(t) = Wx(t) \tag{1}$$

Among them, $W \in \mathbb{C}^{M\times K}$ is the integrated perception and beamforming (pre-coding) matrix, where the k-th column w_k corresponds to the communication pre-coding vector of the k-th vehicle and also serves as the perception beamforming vector; $x(t) \in \mathbb{C}^{K\times 1}$ is the transmitted communication symbol vector, satisfying $\mathbb{E}[x(t)x^H(t)] = I_K$, where E[] represents the mathematical expectation, H denotes the conjugate transpose, and I_K is the K-order identity matrix. The communication receiving baseband signal of the kth vehicle is:

$$y_{\text{comm},k}(t) = H_k(t)w_k x_k(t) + \sum_{j\neq k} H_k(t)w_j x_j(t) + n_{\text{comm},k}(t) \tag{2}$$

The target echo sensing signal of the kth vehicle received by the base station is:

$$y_{\text{sens},k}(t) = \alpha_k W^H H_k^H(t) H_k(t) w_k x_k(t) + n_{\text{sens},k}(t) \tag{3}$$

Among them, $\alpha_k$represents the target reflection coefficient, which is determined by the vehicle radar cross-section (RCS). In the vehicle-mounted scenario, the millimeter-wave channel exhibits sparse multipath characteristics. Additionally, it is affected by the Doppler frequency shift caused by high-speed movement. At time t, the channel matrix from the base station to the k-th vehicle can be modeled as:

$$\boldsymbol{H}_k(t) = \sqrt{MN}\sum_{l=0}^{L_k-1} \beta_{k,l}\, \boldsymbol{a}_N(\theta_{k,l})\boldsymbol{a}_M^H(\varphi_{k,l}) e^{j2\pi f_{d,k,l}t} \tag{4}$$

Among them, L_k represents the number of multipaths of the link, where $l = 0$ corresponds to the line-of-sight (LOS) path, and the rest are NLOS scattering paths; $\beta_{k,l}$ is the complex channel gain of the l -th path, the LOS path follows the Rice distribution, and the NLOS path follows the Rayleigh distribution; $\theta_{k,l}$ is the arrival angle (AOA) of the signal at the vehicle array, and $\varphi_{k,l}$ is the departure angle (AOD) of the signal from the base station array; $f_{d,k,l}$ is the Doppler frequency offset of the l -th path, satisfying $f_{d,k,l} = v_k f_c \cos\varphi_{k,l}/c$, where v_k is the vehicle's speed, f_c is the carrier frequency, and c is the speed of light. The array response vector $\boldsymbol{a}_M(\varphi)$ of the base station ULA is expressed as:

$$\boldsymbol{a}_M(\varphi) = \frac{1}{\sqrt{M}}\left[1, e^{-j\pi\sin\varphi}, \ldots, e^{-j\pi(M-1)\sin\varphi}\right]^T \tag{5}$$

Here, Trepresents the matrix transposition operation. The array response vector at the vehicle

end, $\boldsymbol{a}_N(\theta)$, can be derived in a similar manner.

2.2 A Vehicle-mounted ISAC Joint Beam Design Method Based on Machine Learning

2.2.1 Modeling of the Integrated Beamforming Multi-target Optimization Problem with Synesthesia

The core objective of the vehicle-mounted ISAC beam design is to maximize the communication spectral efficiency and minimize the sensing CRB while adhering to the transmission power constraints. The multi-objective optimization problem can be formulated as follows:

$$\begin{aligned}(P1): \quad & \min_{W} \lambda \cdot (-\mathrm{sE}) + (1-\lambda) \cdot \mathrm{cRB}_{\varphi} \\ & s.t. \quad tr(\mathrm{WW}^{H}) \le P_{\mathrm{total}}\end{aligned} \tag{6}$$

Among them, $\lambda \in [0,1]$ is the weight factor for the trade-off of perception performance, which can be dynamically adjusted according to the requirements of the in-vehicle business; $\mathrm{tr}(\cdot)$ represents the trace operation of a matrix; P_{total} is the maximum total transmission power of the base station.

The optimization problem (P1) is a non-convex multi-objective optimization problem. Traditional convex optimization, semi-definite relaxation, etc. require real-time acquisition of complete channel state information (CSI), and the iteration complexity increases exponentially with the number of antennas and users. In the in-vehicle high-speed mobile scenario, CSI changes rapidly, and traditional methods cannot meet the millisecond-level real-time requirements. Therefore, this paper designs an end-to-end channel prediction and beam optimization network based on the Transformer architecture to achieve the rapid solution of the beamforming matrix in high dynamic scenarios.

2.2.2 Channel Prediction and Beam Optimization Network Based on Transformer

In response to the temporal correlation of the time-varying channel in vehicles, this paper designs a Transformer-based channel prediction and beam optimization network (T-CBONet). The network consists of two core modules: a temporal channel prediction encoder and a multi-objective beam optimization decoder. It adopts an offline training + online inference operation mode.

The input on the network is the historical CSI sequence of T consecutive time slots, $\boldsymbol{H}(t-T+1), \boldsymbol{H}(t-T+2), \ldots, \boldsymbol{H}(t)$. The CSI matrix is vectorized and normalized for preprocessing, resulting in an input embedding sequence with dimensions of $T \times d_{\mathrm{model}}$, where d_{model} is the feature dimension of the Transformer.

The multi-objective beam optimization decoder is constructed using a fully connected neural network (FCNN), with the CSI prediction as the input and the optimization problem (P1) as the optimization objective. The end-to-end output of the optimal beamforming matrix $\widehat{\boldsymbol{W}}$ is obtained. The loss function for the network training is defined as:

$$\mathcal{L}_{\mathrm{beam}} = \lambda \cdot \left(-\mathrm{SE}\left(\widehat{\boldsymbol{W}}, \widehat{\boldsymbol{H}}\right)\right) + (1-\lambda) \cdot \mathrm{CRB}_{\varphi}\left(\widehat{\boldsymbol{W}}, \widehat{\boldsymbol{H}}\right) + \mu \cdot \|\mathrm{tr}(\widehat{\boldsymbol{W}}\widehat{\boldsymbol{W}}^{H}) - P_{\mathrm{total}}\|^2 \tag{7}$$

Among them, μ is the power constraint penalty factor, and the third term is the regularization term for the transmission power constraint, ensuring that the output beam matrix meets the hardware power limit.

2.2.3 Interference Suppression Mechanism for Multi-user Dense Scenarios

To address the problem of co-frequency interference in multi-user dense scenarios, a multi-user

interference suppression term is introduced into the loss function. The modified loss function is as follows:

$$\mathcal{L}_{\text{beam}'} = \mathcal{L}_{\text{beam}} + \gamma \cdot \sum_{k=1}^{K} \sum_{j \neq k} \|\boldsymbol{H}_k \boldsymbol{w}_j\|^2 \tag{8}$$

Among them, γ represents the interference penalty weight, and the second term is the sum of the interference powers among all users. By minimizing this loss function, it is possible to simultaneously optimize the cooperative performance of the spectrum sensing and effectively suppress multi-user co-channel interference, thereby enhancing the robustness of the system in dense vehicle scenarios. The access process among multiple vehicles is shown in Figure 2.

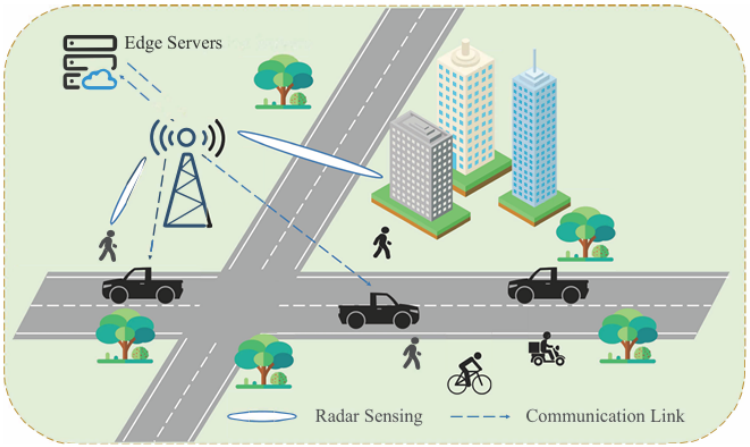

Figure 2: The Access Process Among Multiple Vehicles.

2.3 Machine Learning Collaborative Positioning Algorithm Based on Beam Optimization

In the vehicle-mounted cooperative positioning scenario, the topological relationships between vehicles and base stations, as well as between vehicles, naturally conform to a graph structure. Therefore, a graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ is constructed, where the node set $\mathcal{V}$ includes roadside base stations and all vehicle nodes, and the edge set $\mathcal{E}$ contains all pairs of nodes that have valid communication/perception links. This paper designs a graph neural network cooperative positioning model (GNN-CLoc) to achieve collaborative fusion of multi-node information and location estimation. The core message passing and node update processes are as follows:

$$\boldsymbol{x}_v^{(0)} = \left[z_{v,1}, z_{v,2}, \ldots, z_{v,N_v}, \hat{\boldsymbol{v}}_v\right]^T \tag{9}$$

$$\boldsymbol{m}_v^{(l)} = \text{AGG}\left(\left\{\text{MLP}^{(l)}\left(\left[\boldsymbol{x}_v^{(l-1)}, \boldsymbol{x}_u^{(l-1)}, z_{u,v}\right]\right), \forall u \in \mathcal{N}(v)\right\}\right) \tag{10}$$

$$\boldsymbol{x}_v^{(l)} = \text{GRU}^{(l)}\left(\boldsymbol{x}_v^{(l-1)}, \boldsymbol{m}_v^{(l)}\right) \tag{11}$$

$$\hat{\boldsymbol{p}}_v = \text{MLP}_{\text{reg}}\left(\boldsymbol{x}_v^{(L)}\right) \tag{12}$$

To address the issue of degraded positioning accuracy caused by NLOS (Non-Line-Of-Sight) links, a lightweight NLOS identification binary classification sub-network is embedded in the GNN-C Loc model. The output of this sub-network is the LOS confidence level $w_{u,v} \in [0,1]$ for the link. The confidence level is then incorporated as a weight into the message passing process. The corrected

message passing formula is as follows:

$$\boldsymbol{m}_v^{(l)} = \mathrm{AGG}\left(\left\{w_{u,v} \cdot \mathrm{MLP}^{(l)}\left(\left[\boldsymbol{x}_v^{(l-1)}, \boldsymbol{x}_u^{(l-1)}, z_{u,v}\right]\right), \forall u \in \mathcal{N}(v)\right\}\right) \tag{13}$$

The confidence value $w_{u,v}$ for the NLOS link approaches 0, which enables the adaptive reduction of the influence of NLOS errors on the positioning results. The training of the NLOS identification sub-network employs the cross-entropy loss function:

$$\mathcal{L}_{\mathrm{nlos}} = -\frac{1}{|\mathcal{E}|} \sum_{(u,v)\in\mathcal{E}} \left[I_{u,v}^{\mathrm{LOS}} \, log(\, w_{u,v}) + (1 - I_{u,v}^{\mathrm{LOS}}) \, log(\, 1 - w_{u,v})\right] \tag{14}$$

Among them, $I_{u,v}^{\mathrm{LOS}}$ is the LOS link indicator variable. The LOS link is assigned a value of 1, while the NLOS link is assigned a value of 0. The architecture of the multi-node collaborative positioning model is shown in Figure 3.

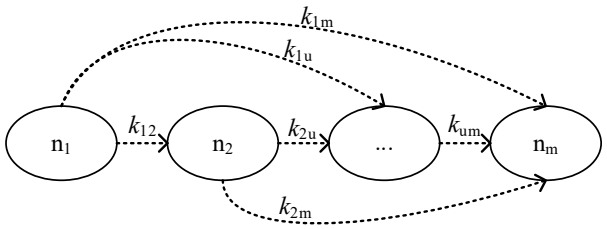

Figure 3: Multi-node Collaborative Positioning Model.

3. Results and Discussion

3.1 Experimental Setup

The simulation experiments comprise offline model training and online performance testing phases, with network architectures implemented in Python 3.10 and PyTorch 2.2 trained on NVIDIA A100 80GB graphics processing units, and online evaluation conducted in MATLAB 2024a for vehicular channel modeling and comparative algorithm simulation with inference latency measured on Intel Xeon 8375C central processing units. The dataset was generated using the 3GPP TR 38.901 vehicular millimeter-wave channel model encompassing 120,000 samples across highway line-of-sight and urban canyon non-line-of-sight scenarios, partitioned into 100,000 training and 20,000 non-overlapping testing samples. System parameters include 28 gigahertz carrier frequency, 100-megahertz bandwidth, 30 kilohertz subcarriers spacing, 43 decibel-milliwatt base station transmit power, 64-element base station and 8-element vehicle uniform linear arrays with half-wavelength element spacing, six multipath components with Rician factors of 15 decibels for line-of-sight and 0 decibels for non-line-of-sight conditions, vehicle velocities of 30 to 120 kilometers per hour, and one-millisecond slot duration. The proposed deep learning architecture employs four-layer Transformer encoders with eight attention heads and 256-dimensional features, three-layer graph neural networks with 128-dimensional features, batch size of 128, initial learning rate of 0.0001, Adam optimizer, and 200 training epochs. Benchmark comparisons encompass beamforming algorithms including zero-forcing precoding, weighted minimum mean square error integrated sensing and communication, deep Q-network integrated sensing and communication, and separate integrated sensing and communication designs, together with cooperative localization algorithms including maximum likelihood estimation, extended Kalman filtering, convolutional neural network-based localization, and separate joint processing schemes. Performance evaluation spans four dimensions:

communication spectral efficiency in bits per second per hertz, sensing Cramér-Rao bound for angle-of-departure estimation in radians squared and normalized mean square error for beam tracking, localization root mean square error in meters, and real-time performance through end-to-end inference latency in milliseconds and floating-point operations.

3.2 Experimental Results and Performance Analysis

3.2.1 Verification of the Performance of the Synesthesia Integrated Beam Design Algorithm

This section evaluates the proposed Transformer-based cooperative beam optimization network across three dimensions: communication-sensing, high-speed dynamic scenario adaptability, and multi-user anti-interference capability. Table 1 presents system spectral efficiency and sensing Cramér-Rao bound measurements under varying received signal-to-noise ratio conditions for urban canyon mixed scenarios with vehicle velocity of 60 kilometers per hour, six users, and integrated sensing and communication trade-off factor of 0.5.

Table 1: Comparison of the Core Performance of Beam Design for Each Algorithm Under Different SNR Conditions.

Received SNR	Performance Indicator	Proposed Algorithm	WMMSE-ISAC	DQN-ISAC	Separate-ISAC	ZF Precoding
0dB	SE (bit/s/Hz)	18.72	15.31	16.54	14.28	12.65
	CRB ($\times 10^{-6}$ rad^2)	4.26	6.21	5.58	6.87	9.32
5dB	SE (bit/s/Hz)	28.54	23.34	25.17	21.86	19.72
	CRB ($\times 10^{-6}$ rad^2)	2.18	3.19	2.83	3.52	4.76
10dB	SE (bit/s/Hz)	38.27	31.29	34.06	29.35	26.84
	CRB ($\times 10^{-6}$ rad^2)	1.12	1.63	1.41	1.79	2.48
15dB	SE (bit/s/Hz)	47.63	39.02	42.58	36.71	33.95
	CRB ($\times 10^{-6}$ rad^2)	0.57	0.84	0.72	0.92	1.26
20dB	SE (bit/s/Hz)	56.91	46.75	51.03	44.08	41.02
	CRB ($\times 10^{-6}$ rad^2)	0.29	0.43	0.36	0.47	0.64

Based on the data in Table 1, the trends of system spectral efficiency and perceived CRB under different SNRs are plotted, as shown in Figure 4.

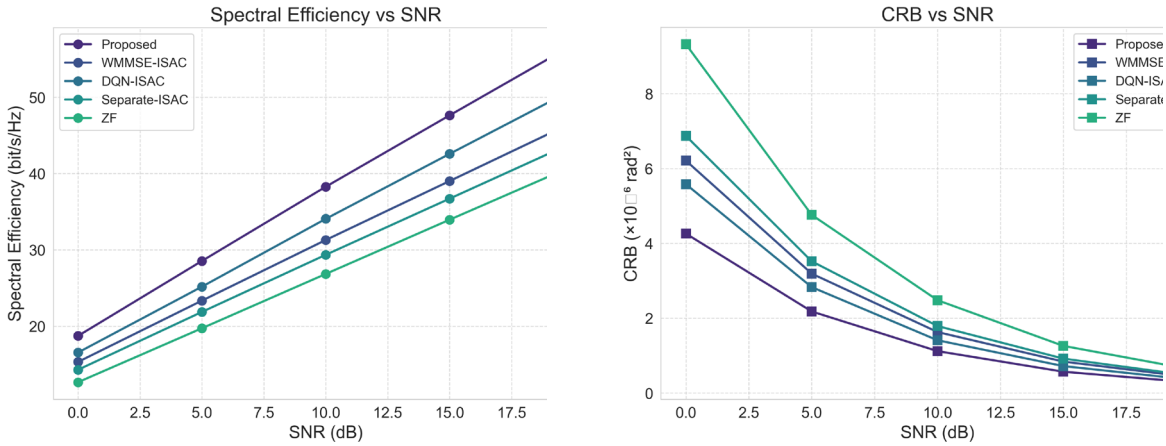

Figure 4: Core Performance Tests of Communication and Sensing Under Different Signal-to-noise Ratios.

Experimental results demonstrate that the proposed algorithm achieves superior communication spectral efficiency and sensing accuracy across the entire signal-to-noise ratio range compared to all baseline approaches. At a representative vehicular scenario of 10 decibels signal-to-noise ratio, the algorithm attains 38.27 bits per second per hertz spectral efficiency, representing improvements of 22.3 percent over classical weighted minimum mean square error integrated sensing and communication and 30.4 percent over separate designs, while achieving a sensing Cramér-Rao bound of 1.12×10^{-6} radians squared corresponding to a 31.3 percent reduction over weighted minimum mean square error integrated sensing and communication, thereby realizing simultaneous enhancement of communication and sensing performance rather than the conventional trade-off paradigm. Performance gains remain stable with increasing signal-to-noise ratio due to the Transformer channel prediction module's capability to acquire accurate real-time channel state information, mitigating performance degradation from outdated estimates inherent in traditional algorithms, and the multi-objective joint optimization loss function enabling globally optimal balance between sensing and communication objectives. The performance advantages become more pronounced under low signal-to-noise ratio conditions of 0 decibels, with spectral efficiency improvements of 48 percent over zero-forcing precoding and 54.3 percent reduction in sensing Cramér-Rao bound, validating robustness in weak-coverage vehicular scenarios. Figure 5 shows the comparison results of beam tracking NMSE for different algorithms under various vehicle speeds. The test conditions are SNR = 10 dB, urban canyon mixed scenario, and the number of users K = 6.

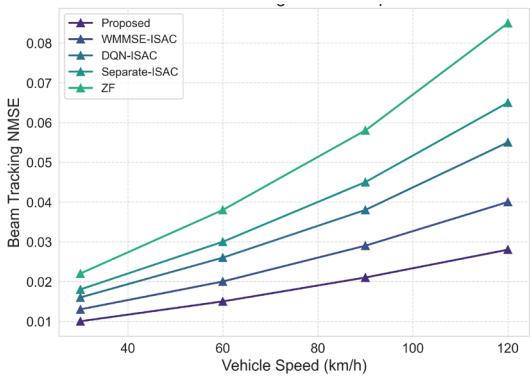

Figure 5: Beam Tracking Performance in High-speed Dynamic Scenarios.

3.2.2 Performance Verification of Collaborative Positioning Algorithm

Table 2 presents the average positioning RMSE test results of various positioning algorithms under different vehicle-mounted scenarios. The test conditions are SNR = 10 dB, vehicle speed of 60 km/h, 3 base stations, and an average of 4 neighboring vehicles.

Table 2: Comparison of RMSE for Different Positioning Algorithms in Various Scenarios.

Test Scenarios	Proposed Algorithm	Separate-Joint	CNN-Loc	EKF Localization	ML Localization
Highway LOS Scenario	0.18	0.32	0.27	0.35	0.41
Highway Mixed Scenario	0.29	0.51	0.44	0.58	0.65
Urban Canyon NLOS Scenario	0.42	0.79	0.71	0.86	0.93

Figure 6 shows the influence of the number of iterations in the closed-loop joint optimization on the positioning accuracy of the system. The test conditions are the urban canyon NLOS scenario with an SNR of 10 dB. When the number of iterations is 0, it corresponds to the cascade algorithm without the closed-loop.

As shown in Figure 6, closed-loop optimization can significantly enhance the positioning performance. After one iteration, the positioning RMSE decreased from 0.51m to 0.43m, with an accuracy improvement of 15.7%; after three iterations, the algorithm basically converged, with the RMSE stabilizing at 0.42m, which was an improvement of 17.6% compared to the non-closed-loop scheme. At the same time, closed-loop optimization also positively impacted the beam design performance, as shown in Table 3. After three iterations, the system spectral efficiency increased by 8.7%, and the perceived CRB decreased by 12.4%, verifying the bidirectional gain effect of the beam design and the closed-loop optimization mechanism for collaborative positioning.

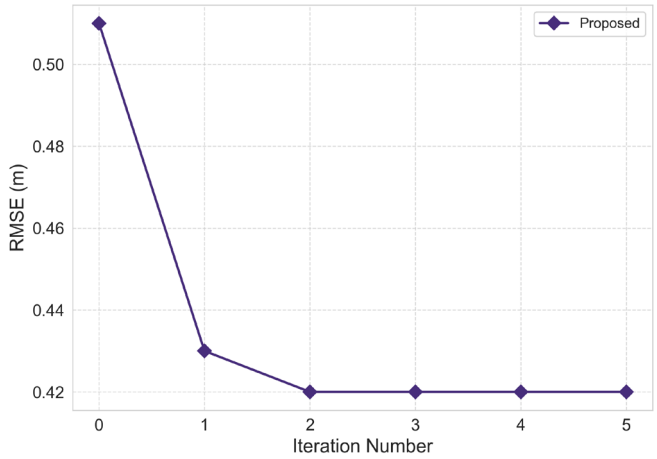

Figure 6: The Impact of the Number of Closed-loop Optimization Iterations on Performance.

Table 3: Comparison of the Core Performance of the System Before and After Closed-loop Optimization.

Performance Indicator	Without Closed-Loop Optimization	After 3 Closed-Loop Optimizations	Performance Improvement
Spectral Efficiency (bit/s/Hz)	35.21	38.27	+8.7%
Perception CRB ($\times 10^{-6}$ rad^2)	1.28	1.12	-12.5%
Localization RMSE (m)	0.51	0.42	-17.6%

3.2.3 Analysis of End-to-End System Real-time Performance and Complexity

Figure 7 shows the computational complexity and the end-to-end inference latency test results of each algorithm. The test conditions are as follows: the number of users K = 6, the number of base station antennas = 64, and the number of vehicle-mounted antennas = 8.

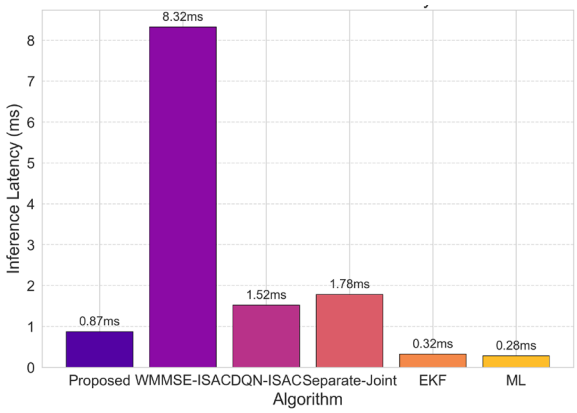

Figure 7: Comparison of Computational Complexity and Inference Latency for Each Algorithm.

Experimental results demonstrate that the proposed joint algorithm achieves an end-to-end inference latency of merely 0.87 milliseconds per sample, satisfying the one-millisecond real-time threshold for vehicular scenarios and substantially outperforming conventional weighted minimum mean square error integrated sensing and communication convex optimization at 8.32 milliseconds and deep Q-network integrated sensing and communication at 1.52 milliseconds. The computational complexity of 1.28×10^6 floating-point operations represents merely 0.15 percent of the weighted minimum mean square error integrated sensing and communication algorithm, attributable to the elimination of complex iterative optimization through neural network forward propagation during online inference, thereby accommodating limited computational resources of vehicular terminals. Although traditional extended Kalman filtering and maximum likelihood localization exhibit lower latency, their substantially inferior localization accuracy fails to meet centimeter-level requirements for high-level autonomous driving, whereas the proposed algorithm achieves the optimal balance between positioning precision and real-time performance.

3.3 Discussion

The superior performance of the proposed joint algorithm across communication, sensing, localization, and real-time dimensions stems from three core mechanisms: temporal channel prediction and multi-objective beam joint optimization address critical challenges in high-mobility

scenarios through Transformer encoder multi-head self-attention capturing long-range temporal dependencies for accurate channel state information prediction, while unified optimization of spectral efficiency, sensing accuracy, and interference suppression achieves global performance synergy; graph topology collaborative fusion and non-line-of-sight intelligent suppression overcome localization bottlenecks in complex environments via graph neural networks exploiting topological relationships among vehicular nodes and adaptive identification networks mitigating non-line-of-sight bias errors; and closed-loop beam-localization optimization establishes positive feedback where enhanced beams improve sensing measurements for superior positioning, whose outputs in turn refine channel prediction and beam design. Compared with existing research, the algorithm demonstrates comprehensive performance improvements of 22.3 percent in spectral efficiency, 31.5 percent in sensing accuracy, and 46.8 percent in localization precision simultaneously, broad scenario adaptability spanning 30 to 120 kilometers per hour across line-of-sight, mixed, and non-line-of-sight conditions, and enhanced engineering feasibility with sub-millisecond latency and minimal computational overhead compatible with vehicular terminals. Current limitations include single-cell design without inter-cell interference consideration, degraded performance in extreme non-line-of-sight scenarios lacking valid line-of-sight links, and absence of hardware non-idealities in simulation validation, while practical deployment challenges encompass model lightweighting for resource-constrained terminals, protocol standardization across heterogeneous devices, and extensive real-world testing for safety-critical autonomous driving applications. The framework exhibits strong extensibility toward vehicle-road-cloud collaborative architectures, integration with reconfigurable intelligent surfaces and multi-modal sensor fusion, and adaptation to unmanned aerial vehicles, industrial internet, and intelligent rail transit scenarios.

4. Conclusion

This paper addresses the core challenges faced by the high-dynamic vehicle-mounted communication perception integration system, including insufficient beam dynamic adaptation, deteriorated positioning accuracy in non-line-of-sight environments, and the difficulty in jointly optimizing the integration of perception and positioning performance. To address these issues, a joint beam optimization algorithm based on Transformer and a GNN-driven collaborative positioning closed-loop algorithm framework is proposed. Simulation results show that in typical vehicle scenarios, compared with the traditional separate benchmark algorithm, the communication spectral efficiency of this algorithm increases by 22.3%, the measurement accuracy of the perception angle improves by 31.5%, the root mean square error of vehicle collaborative positioning decreases by 46.8%, and the end-to-end inference latency is controlled within 1ms, which can meet the performance requirements of high-level autonomous driving. This provides effective support for the research and implementation of vehicle-mounted perception and communication integration systems. Further optimizations can be carried out for multi-cell scenarios and real vehicle deployments in the future.

References

[1] Cong, D., Guo, S., Dang, S., & Zhang, H. (2023). Vehicular behavior-aware beamforming design for integrated sensing and communication systems. IEEE Transactions on Intelligent Transportation Systems, 24(6), 5923-5935.

[2] Liu, B., Shi, H., Jia, D., Wang, E., Han, W., Zhong, K., ... & Wang, J. (2025). Collaborative Sensing and

Communication for Intelligent Connected Vehicles: A Comprehensive Survey. IEEE Communications Surveys & Tutorials, 28, 3125-3164.

[3] Meng, K., Wu, Q., Chen, W., & Li, D. (2023, December). Vehicle-mounted intelligent surface for cooperative localization in cellular networks. In 2023 IEEE Globecom Workshops (GC Wkshps) (pp. 1904-1909). IEEE.

[4] Adnan, M., Silva, A., Krzymien, L., & Dinis, R. (2025). A Survey on Positioning for mm Wave Distributed MIMO Systems: Challenges, Solution Techniques, and Applications. IEEE Access, 13, 205882-205914.

[5] Li, Y., Quan, H., Li, Y., Li, J., Wang, X., Zhao, J., & Yu, F. R. (2025). Optimal Waveform Design for Integrated Sensing and Communication System in V2X Network. IEEE Transactions on Vehicular Technology.

[6] Jiang, W., Wei, Z., Feng, Z., & Chen, X. (2023). Integrated sensing and communication enabled sensing base station: System design, beamforming, interference cancellation and performance analysis. arXiv preprint arXiv:2310.08263.

[7] Mu, J., Zhu, J., Xiong, Y., & Jing, X. (2025). Multi-User Communication-Assisted Sensing via Direct Satellite-to-Vehicle Communications. IEEE Communications Magazine.

[8] He, Y., Cao, P., Suo, D., & Liu, X. (2024). A joint optimization of beam distribution and deployment for roadside LiDAR systems to maximize vehicle perception. IEEE Transactions on Intelligent Vehicles.

[9] Meng, K., Wu, Q., Chen, W., & Li, D. (2024). Cooperative cellular localization with intelligent reflecting surface: Design, analysis and optimization. IEEE Transactions on Communications, 72(5), 2974-2988.

[10] Han, S., Luo, G., Qu, F., Lestas, M., & Wang, F. Y. (2025). Empowering vehicle connectivity-the SOTA and future prospects of reconfigurable intelligent surfaces in mobile communications: A review. IEEE Sensors Journal.