Distributed Decision-making System for Intelligent Connected Vehicles Based on Federated Learning

Yu Zhao[1], Zhiwei Guan[1], Weiqiang Wang[2], Yinfang Wang[2], Xiaowei Yang[3]
[1]School of Automobile and Rail Transit, Tianjin Sino-German University of Applied Sciences, Tianjin 300350, China
[2]Tianjin Deke Intelligent Control Technology Co., Ltd, Tianjin 300350, China
[3]China Industrial Control Systems Cyber Emergency Response Team, Beijing 100040, China

Corresponding author: Xiaowei Yang (e-mail: ZYTJ523@163.com)

ABSTRACT Current distributed decision-making for intelligent connected vehicles is constrained by model heterogeneity, high communication overhead, and data privacy concerns, which impede efficient multi-vehicle collaboration in wireless vehicular communication environments. As reliable information exchange and low-latency signal transmission become increasingly important in intelligent transportation systems and electromagnetic communication networks, this paper proposes a distributed decision-making framework integrating Federated Distillation (FD) and Graph Attention Networks (GAT). A local GAT dynamically constructs vehicle adjacency graphs for neighboring perception, while FD replaces conventional parameter sharing with soft-label exchange, significantly reducing communication burden, enabling knowledge transfer across heterogeneous devices, and alleviating computational bottlenecks on resource-limited nodes. Evaluated in the most challenging scenario characterized by maximum vehicle density, dynamic communication delays, packet loss, severe data Non-IIDness, and complex multi-intersection interactions, the proposed method demonstrates superior collaborative performance. The average decision latency is maintained at 91 ms, which is 43 ms lower than that of the traditional FedAvg approach, while the number of uncoordinated behavior triggers is reduced by 50% to only three. These results verify that the proposed framework achieves efficient and privacy-preserving collaborative decision-making while providing a scalable solution for distributed intelligence in wireless vehicular networks and communication-intensive electromagnetic environments.

INDEX TERMS federated learning, intelligent connected vehicles, graph attention network, vehicular wireless communication, electromagnetic information transmission

I. INTRODUCTION

As the core component of the new generation of transportation systems, intelligent connected vehicles (ICVs) have the ability to collaborate efficiently and make autonomous decisions, which is of great significance to improving road safety, alleviating traffic pressure, and building a smart travel system. Improving efficiency through autonomous, coordinated systems is a universal goal. This technological push is similar to the efforts in the industry to leverage automated systems to drastically enhance worker safety and production speed while ensuring the continuous, high-quality output of delicate materials. In a multi-vehicle interactive environment, distributed decision-making systems support vehicles to achieve real-time path optimization and behavior selection in dynamic scenarios through information perception and strategy collaboration between nodes [1-3]. Under large-scale deployment, traditional centralized control architectures suffer from communication bottlenecks and central node failure risks, shifting the research focus toward distributed decision-making systems with edge intelligence. This technological evolution holds significant parallels for industrial automation. For example, edge intelligence allows the industry to efficiently manage hundreds of synchronized machines across a vast factory floor, with processing power distributed to the source of the data for real-time control and optimized operational efficiency [4-6]. In this context, building a distributed decision-making method with efficient collaboration capabilities, data privacy protection, and adaptability to heterogeneous environments has become a key direction for the research of intelligent connected vehicle systems. In the proposed system, heterogeneity is characterized by three interrelated dimensions: structural heterogeneity caused by different policy network architectures adopted by vehicles with unequal computing resources, data heterogeneity resulting from distinct local traffic perception and experience distributions, and resource heterogeneity linked to unequal bandwidth and latency constraints. These factors collectively define the heterogeneous environment that the Federated Distillation mechanism is designed to align and unify through soft-label knowledge transfer. Overcoming heterogeneity is particularly vital, and this research offers a framework applicable to industrial settings like the industry, where diverse and often older machinery must collaborate efficiently while protecting proprietary information and adapting to varying material inputs [7,8].

Aiming at the problem of privacy-preserving distributed decision consistency in heterogeneous environments, this paper constructs a distributed collaborative decision-making system for intelligent connected vehicles that integrates FD and GAT. The system architecture consists of a local policy learning module, a neighboring vehicle perception graph construction module, and an FD coordination module. Each vehicle uses its own perception data to train its strategy locally and constructs a dynamic adjacency graph through GAT to extract the interaction weights with surrounding vehicles, thereby achieving local perception enhancement and spatial dependency modeling. At the global level, by uploading soft labels to the server and generating a unified soft distribution based on the global output, knowledge alignment, and distillation synchronization between heterogeneous models are achieved, thereby replacing parameter aggregation in the traditional FedAvg method, reducing communication load and adapting to different network structures. In the process of periodically executing global distillation synchronization and local policy updates, the system gradually converges to individual strategies with collaborative perception capabilities, effectively improving decision consistency and system robustness. Aiming at heterogeneous computing architecture, communication constraints, and multi-vehicle dynamic collaboration requirements, the study proposes a complete distributed decision-making system construction path and method mechanism. The focus on overcoming heterogeneity is particularly valuable as it can be directly applied to industrial settings: providing a system architecture that effectively coordinates machines of different ages and types—a common reality in the industry—to achieve seamless and reliable dynamic production.

II. RELATED WORK

In the study of distributed decision-making mechanisms for intelligent connected vehicles, multi-agent collaboration and distributed optimization methods have become the focus of exploration in recent years. Gan et al. proposed a collaborative driving framework based on multi-agent deep reinforcement learning. Through grouping strategies, partial parameter sharing, and hybrid reward mechanisms, they achieved efficient multi-vehicle collaboration in highway ramp merging scenarios, significantly improving traffic efficiency and reducing conflict risks, and performed well under different traffic densities [9]. In multi-agent systems, FL has gradually become a research hotspot for intelligent connected vehicle decision-making collaboration due to its privacy protection and distributed modeling capabilities. To address the single point failure and scalability bottleneck problems of centralized servers in the Internet of Vehicles environment, Liu et al. built a consensus-based decentralized FL framework, replaced the traditional centralized server with distributed parameter aggregation, and combined the local model consensus mechanism between vehicles to effectively solve the single point failure risk, network scale expansion bottleneck, and aggregation performance degradation caused by redundant data in FL in the Internet of Vehicles scenario [10]. This type of method emphasizes the application of a more flexible consensus mechanism in FL to adapt to the highly dynamic and heterogeneous communication and computing requirements in multi-agent systems, and further promotes the practicality and scalability of decentralized training architectures [11,12]. Carrascosa et al. implemented a decentralized model training architecture by applying a fully distributed consensus mechanism based on a multi-agent system and a new asynchronous consensus protocol in FL, and verified the feasibility and performance advantages of the architecture in actual agent platforms and Australian wind farm cases [13]. The above decentralized mechanism provides a structural basis for efficient collaboration among multiple agents and also creates a more flexible communication and computing environment for integrating complex decision-making models such as reinforcement learning [14,15]. However, in terms of data privacy protection, traditional security aggregation protocols (such as Secure Aggregation and gradient masking) have high communication load and reverse reasoning risks in multiple rounds of training. Ergun et al. pointed out that although sparse encryption can reduce communication, it is still much higher than the soft label transmission mechanism [16]. However, this type of method mainly focuses on heterogeneous management in edge-cloud architectures and has not yet fully solved the problems of strategy sharing and heterogeneous knowledge alignment in vehicle-side multi-agent systems.

III. SYSTEM DESIGN AND DECISION-MAKING MECHANISM IMPLEMENTATION

OVERALL SYSTEM ARCHITECTURE

The overall system architecture is built around the distributed collaborative decision-making task of intelligent connected vehicles. The functional modules within the system are logically divided into local policy learning modules, adjacency graph construction modules, and distillation coordination modules, and form a closed loop through high-frequency information interaction. Each vehicle node performs local state processing based on its own perception input data, extracts semantic features and sends them to the policy model to generate action distribution, and realizes preliminary local decision-making. The vehicles build adjacency relationships through the state broadcast mechanism to form a local graph structure based on dynamic perception, which is used to characterize the spatial dependency characteristics between multiple agents in the current traffic environment. The graph structure is updated in each training cycle, and the information weights of different neighboring vehicles are dynamically adjusted through the graph attention mechanism. The locally generated action distribution is used as a soft label to participate in the cross-node distillation coordination mechanism. Each node synchronizes the policy update without transmitting the original data, thereby realizing the sharing and integration of decision-making capabilities in a heterogeneous model environment. The above modules maintain efficient collaboration under heterogeneous structures through the FD process, while maintaining a balance between system communication overhead and privacy protection. The entire data flow starts from the perception input and finally completes the policy update process through state encoding, neighboring car mapping, strategy generation, and distillation optimization, as shown in Figure 1.

Figure 1 shows the internal composition and interaction paths of each functional module in the system. The overall structure presents the key processing flow and data flow in a hierarchical nested manner. The system is divided into three subsystems: local learning module, neighboring mapping module, and FD module. The core calculation steps are further refined in each module, and its functional logic is reflected in the form of sub-nodes. The information processing process starts from the perception input and enters the neighboring car state processing and graph relationship modeling process after state feature extraction. The dynamic perception fusion is completed through the graph attention mechanism, and the fusion result is input into the local strategy model to generate action distribution. Then, the cross-node knowledge transfer is completed through soft label upload. The strategy after global distillation aggregation is synchronized to each node to form a closed loop. Each module clearly identifies the data interaction path through directed edges. The system realizes end-to-end collaborative optimization process on the basis of maintaining structural heterogeneity and data localization, which reflects the dynamic modeling and joint learning mechanism of distributed decision-making system in multi-agent interaction scenarios.

DESIGN OF LOCAL POLICY LEARNING MODULE

The local policy learning module uses the state perception data collected by the intelligent connected vehicle itself as input and generates action distribution through reinforcement learning method to guide the vehicle's driving decision at the current moment. The state perception data contains time series features such as vehicle position, speed, acceleration, heading angle, and its historical trajectory, which are mapped into the state vector $s_t \in \mathbb{R}^d$ as the input of the policy network. The temporal information is encoded through state stacking within a fixed observation window, and the resulting state vector is processed by a multi-layer perceptron policy network. The output is the action probability distribution $\pi_\theta(a_t|s_t)$ at the current moment, where a_t represents the vehicle action, and θ is the policy network parameter.

The policy training adopts the proximal policy optimization (PPO) algorithm, combined with the distributed asynchronous update mechanism to improve the stability and convergence efficiency of the strategy. The objective function is designed as follows:

$$\mathcal{L}_{\text{PPO}}(\theta) = \mathbb{E}_t\Big[\min\Big(r_t(\theta)\hat{A}_t, \text{clip}(r_t(\theta), 1-\epsilon, 1+\epsilon)\,\hat{A}_t\Big)\Big] \tag{1}$$

In Formula (1), $r_t(\theta) = \frac{\pi_\theta(a_t|s_t)}{\pi_{\theta_{\text{old}}}(a_t|s_t)}$ represents the probability ratio of the current strategy to the old strategy, reflecting the degree of deviation between the current strategy and the old strategy; $\hat{A}_t$ is the advantage function, which is used to measure the improvement of the current action compared with the baseline strategy; ϵ is the clipping threshold, which limits the amplitude of the policy update to avoid instability caused by drastic changes in the policy. The structure of the minimum function and the clipping strategy in the above objective function can suppress drastic updates while ensuring policy improvement, thereby improving training stability.

The value estimations of the state-action pair use the generalized advantage estimation (GAE), which are calculated as follows:

$$\hat{A}_t = \sum_{l=0}^{\infty}(\gamma_1\lambda)^l\delta_{t+l} \tag{2}$$

$$\delta_t = r_t + \gamma_1 V(s_{t+1}) - V(s_t) \tag{3}$$

In Formulas (2) and (3), γ_1 is the discount factor, which represents the degree of attenuation of future rewards; λ is the factor that balances the deviation and variance; $V(s_t)$ is the value estimation of the current state; δ_t is the single-step temporal difference error, which represents the deviation between the actual reward and the estimated value.

NEIGHBORING VEHICLE DYNAMIC MAPPING AND ATTENTION FUSION MECHANISM

In the multi-agent collaborative decision-making of intelligent connected vehicles, the perception accuracy and fusion quality of neighboring vehicle information have a decisive impact on system performance. To improve the modeling ability of local perception strategies for spatial dependencies, GAT is applied to dynamically model and weight the relationship between vehicles. This section designs a graph structure construction process based on state perception and combines the attention mechanism to achieve weighted fusion of neighboring vehicle features to improve the coordination and perception sensitivity of local strategies in distributed environments.

Each vehicle node in the system constructs an adjacency graph locally to describe its interaction structure with surrounding vehicles at the current moment. The

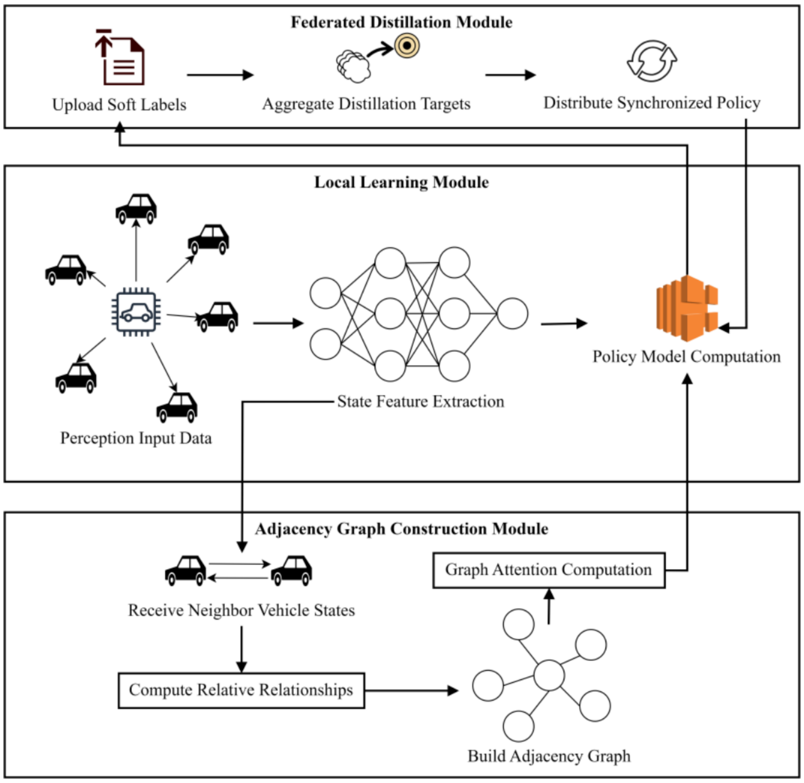

FIGURE 1. System architecture diagram

adjacency graph is constructed solely from the vehicles currently within the communication radius of the target vehicle, and no global connectivity information is assumed. While all vehicles employ an identical GAT architecture to ensure representational consistency, the model parameters are optimized independently based on each vehicle's own interaction experience. No parameter sharing or centralized training is involved, and the coordination effect emerges only through the subsequent distillation process. The graph represents individual vehicles with nodes, and the establishment of edges depends on the relative relationship between nodes in the state space. Assuming that the graph structure is $G_t = (V_t, E_t)$, where V_t is the vehicle set, and E_t is the edge set, the generation of edges is based on vehicle pairs within the threshold radius r. During the construction process, each vehicle node i extracts its relative position Δp_{ij}^t, relative speed Δv_{ij}^t, and current state feature s_j^t with its neighbor j to form a multi-dimensional fusion input.

Figure 2 shows the neighboring vehicle perception graph structure constructed by GAT for the target vehicle in a two-dimensional road scene. Nodes represent vehicle positions; edge connections represent dynamic perception relationships; the color of the edges reflects the distribution of attention weights, which is used to characterize the degree of focus of the model on different neighboring vehicle states before strategy generation.

GAT is applied to perform information fusion and propagation on the constructed graph structure. In GAT, each node i assigns an attention weight a_{ij} to the feature h_j of its neighboring node j, and the calculation method of the attention weight is defined by the following formula:

$$a_{ij} = \frac{\exp\left(\mathrm{LeakyReLU}\left(\vec{a}^T[Wh_i||Wh_j]\right)\right)}{\sum_{k\in\mathcal{N}_i}\exp(\mathrm{LeakyReLU}(\vec{a}^T[Wh_i||Wh_k]))} \tag{4}$$

In Formula (4), h_i and h_j represent the input feature vectors of vehicles i and j, respectively; W is the linear transformation matrix to be learned; $\vec{a}$ is the attention coefficient weight vector; $\mathcal{N}_i$ represents the node i's neighbor set; $||$ represents the vector concatenation operation. This formula constructs an adaptive mechanism for dynamically adjusting the information propagation intensity according to the similarity between features. Through the above calculation, node i obtains the attention weighted feature update value $\hat{h}_i$ from all neighboring nodes, expressed as:

$$\hat{h}_i = \sigma\left(\sum_{j\in\mathcal{N}_i} a_{ij}Wh_j\right) \tag{5}$$

In Formula (5), $\sigma(\cdot)$ represents a nonlinear activation function, usually ReLU. This aggregation mechanism achieves the goal of strengthening the information transmission of key neighboring nodes while maintaining the sparsity of the graph structure.

To improve the modeling capability, the system adopts a multi-head attention mechanism to enhance the stability of feature expression. Assuming the number of attention

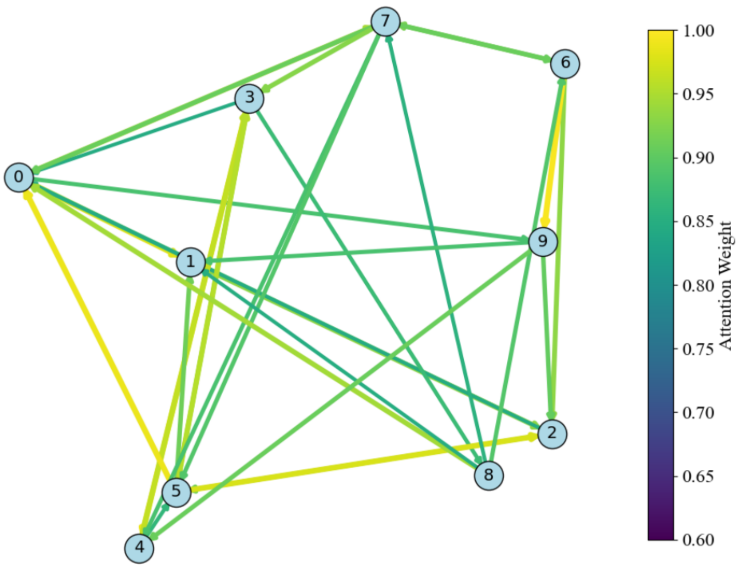

FIGURE 2. Spatial distribution visualization of neighboring vehicle perception relationship graph

heads is K, the final node output $\hat{H}_i$ is composed of the concatenation of the results of each head:

$$\hat{H}_i = ||_{k=1}^{K} \sigma \left(\sum_{j \in \mathcal{N}_i} a_{ij}^{(k)} W^{(k)} h_j \right) \tag{6}$$

In Formula (6), $a_{ij}^{(k)}$ and $W^{(k)}$ are the attention weight and linear transformation matrix of the k-th attention head, respectively. The concatenation operation expands the representation space and effectively improves the perception resolution of the model in complex traffic scenarios.

FD AND HETEROGENEOUS COLLABORATIVE STRATEGY UPDATE

The FD stage is triggered after the communication cycle ends. Before uploading the local soft labels, each vehicle reorders the policy output distributions of its neighboring vehicles according to the attention weights derived from the adjacency graph. Soft labels associated with neighbors receiving higher attention weights are assigned higher fusion priorities and larger aggregation coefficients when forming the global distillation target. In this way, the interaction topology influences the contribution strength of each neighbor in the distillation process, guiding the policy update toward behavior patterns that reflect the dominant local interaction relationships. It is assumed that the vehicle set $\mathcal{K} = \{1, \ldots, K\}$ in the t-th round each infers the policy output vector $\pi_t^k(s)$ on the state s, which is converted into a soft label by the temperature scaling operator $\sigma_T(\cdot)$ and uploaded to the coordination end. Here, the state s does not refer to the instantaneous local physical state of each vehicle. Instead, all vehicles evaluate their policy outputs on a shared proxy state set maintained locally, which is constructed offline from a fixed distribution of abstracted traffic interaction states. Each proxy state encodes normalized relative motion and interaction descriptors rather than absolute kinematic quantities, ensuring semantic consistency across nodes. The global distillation target is therefore computed by aggregating policy outputs evaluated on identical proxy inputs, avoiding direct averaging across heterogeneous local states. The coordination end defines the comprehensive weight $w_k = n_k \rho_k$ according to the node effective sample volume n_k and the link reliability coefficient ρ_k and then calculates the global distillation target distribution $\pi_t^g(s)$:

$$\pi_t^g(s) = \frac{1}{\sum_{k=1}^{K} w_k} \sum_{k=1}^{K} w_k \sigma_T \left(\pi_t^k(s) \right) \tag{7}$$

In Formula (7), $\sigma_T(\cdot)$ is the Softmax transformation controlled by temperature T, and w_k reflects the node contribution. This distribution is sent to all nodes in the same round and cached in the local teacher buffer.

When the vehicle performs policy updates on the local data batch $(s, a, r, s') \sim \mathcal{D}_k$, the distillation consistency loss is applied to form a joint optimization with the original reinforcement learning objective function, as shown in Formula (8):

$$\begin{aligned} \mathcal{L}_k = {} & \mathbb{E}_{s \sim \mathcal{D}_k} \left[\mathrm{KL} \left(\pi_t^g(s) \| \sigma_T(\pi_t^k(s)) \right) \right] \\ & + \lambda \mathbb{E}_{(s,a,r,s') \sim \mathcal{D}_k} \left[\left(r + \gamma_2 V_{\theta_k}^{-}(s') - Q_{\theta_k}(s,a) \right)^2 \right] \end{aligned} \tag{8}$$

In Formula (8), the Kullback–Leibler divergence term measures the consistency between the local policy output distribution and the global distillation target distribution. The loss weight controls the contribution of the distillation constraint relative to the PPO policy optimization objective. The overall training process follows a unified actor–critic PPO framework, where the policy network and value network are jointly optimized without any value-based Q-learning components. The gradient is calculated by $\nabla_{\theta_k}\mathcal{L}_k$. Then, the Adam optimizer is used to update the parameter θ_k, and $\theta_k^- \leftarrow \theta_k$ is synchronized every τ^{-1} rounds. The heterogeneous structure adaptation strategy is based on the model-independent characteristics of the distillation mechanism: since only the normalized policy distribution is transmitted instead of the weight tensor in the communication, policy networks with different capacities or different hierarchical combinations can also extract knowledge from the global teacher. To further alleviate the bandwidth pressure, quantization coding is used on the node side to logarithmize $\sigma_T(\pi_t^k(s))$, and the quantization step size is adaptively adjusted according to the node's historical communication error; the coordination end implements edge aggregation after decoding and only calculates the nodes with a change rate $\|\pi_t^k(s) - \pi_{t-1}^k(s)\|_1$ higher than the threshold through Formula (7), thereby avoiding redundant upload.

Table 1 summarizes the differences between distillation collaboration and traditional parameter aggregation in terms of communication overhead, model adaptability, and privacy constraints. The average data transmission volume of the distillation scheme in each round of communication is 6.5MB, and the jitter during synchronization is controlled within 3ms. While maintaining privacy in-situ storage, it significantly compresses the scale of cross-node information interaction and maintains consistent convergence of decision performance under different network scales and model architecture combinations.

The communication volume reported in this table refers to one-way transmission per node per round. The faster convergence observed in the distillation-based collaboration is attributed to the reduction of cross-node gradient interference under heterogeneous data distributions. By transmitting normalized policy distributions rather than full parameter tensors, the distillation process filters out low-quality updates caused by severe Non-IID local experiences. This mechanism stabilizes the optimization trajectory and reduces redundant oscillations during early training, resulting in fewer effective convergence rounds under identical synchronization intervals.

IV. EXPERIMENTAL ENVIRONMENT AND SETTINGS

SIMULATION PLATFORM AND ENVIRONMENT CONSTRUCTION

The construction of the simulation platform needs to consider traffic flow, road geometry, and communication link quality to ensure that the decision algorithm is fully verified in representative scenarios. To systematically cover these factors, this paper defines five sets of differentiated configurations in the CARLA environment. Table 2 lists the specific parameters.

The S5 configuration also combines the widest Non-IID distribution of driving behaviors due to the dense and variable traffic flow, creating the most heterogeneous training conditions in both data and network dimensions. This composite complexity makes S5 the most rigorous environment for validating the scalability and stability of the proposed FD+GAT system. Table 2 divides the scenarios into three discrete categories: traffic flow, road structure type, and communication condition configuration. The traffic flow level increases from low to extremely high, with low flow being 3 vehicles per lane, medium flow being 6 vehicles per lane, high flow being 9 vehicles per lane, and extremely high flow being 10 vehicles per lane. The road structure covers straight roads, intersections, roundabouts, mixed ramps, and urban multi-intersection grids. The communication configuration involves different conditions from zero-delay ideal links to high-fluctuation noise channels. The diverse combinations ensure complete test coverage and provide reliable data support for subsequent algorithm performance evaluation.

All training and test data in this paper are constructed and collected in the CARLA 0.9.X version platform. The Python API (Application Programming Interface) is used for automated simulation script control. For each scenario S1–S5, 300 vehicle trajectory data are generated. The trajectory sampling frequency is set to 10 Hz, and a total of more than 100 hours of vehicle interaction data are generated. The trajectory data includes vehicle position, speed, acceleration, relative distance, neighboring vehicle status, etc., as state input; the instant reward value is automatically calculated based on the custom reward function for reinforcement learning target modeling. In the data generation process, CARLA's built-in traffic module and random disturbance mechanism are combined to ensure the diversity of scene coverage and the rationality of behavior. All data are simulated synthetic data and do not involve real vehicle or personnel privacy information.

PARAMETER SETTING

To ensure that the system can achieve efficient collaborative learning in a multi-node heterogeneous environment, the training phase needs to unify parameter settings in multiple dimensions and clarify the coordination mechanism. Parameter selection is directly related to the strategy convergence speed, distillation synchronization quality, and overall communication stability, and needs to be finely configured in combination with the characteristics of the experimental platform and the strategy model structure. Table 3 lists the key training and distillation collaborative settings used in this study.

Table 3 covers the core training elements such as initial learning rate, communication cycle, distillation temperature, policy network structure, and loss weight. The communication cycle controls the distillation trigger frequency; the

TABLE 1. Comparison of distillation collaboration and parameter aggregation methods

Metric	Federated Distillation Collaboration	Traditional Parameter Aggregation
Average One-way Communication per Round	6.5 MB	28.7 MB
Synchronization Jitter	3 ms	11 ms
Heterogeneous Model Support	Direct soft label transfer, structure-independent	Requires weight homogeneity, high adaptation cost
Privacy Leakage Risk	Only exposes probability distributions, hard to reverse engineer data	Directly shares gradients/weights, prone to model inversion
Local Convergence Rounds	27 rounds	34 rounds

TABLE 2. Overview of simulation scenarios and communication configurations

Scene ID	Traffic Flow	Road Structure Type	Communication Condition Configuration
S1	Low (three vehicles per lane)	Two-lane one-way straight road	Ideal link zero latency no packet loss
S2	Medium (six vehicles per lane)	Four-way intersection	Fixed latency 10 ms no packet loss
S3	High (nine vehicles per lane)	Multi-entry roundabout	Random latency 5–30 ms packet loss rate 1%
S4	Medium (six vehicles per lane)	Curved and sloped mixed segment	Latency 15 ms packet loss rate 3% bandwidth constrained
S5	Very high (ten vehicles per lane)	Urban multi-intersection grid	Latency fluctuates 10–50 ms packet loss rate 5% noisy channel

TABLE 3. Federated training and distillation strategy parameter settings

Parameter Description	Value	Notes
Initial learning rate	0.0003	Applied for policy network optimization
Communication interval	Every 10 rounds	Determines how often federated distillation occurs
Distillation temperature	2.5	Controls the smoothness of the soft label distribution
Policy network structure	3-layer MLP (Multilayer Perceptron) with ReLU	Input dimension is 32, output matches action space size
Distillation loss weight	0.4	Balances the distillation loss with reinforcement learning objective

temperature parameter adjusts the soft label distribution of the policy output to adapt to different model structures; the policy network adopts a multi-layer perceptron structure combined with nonlinear activation to achieve the mapping from state to action; the loss term weight is used to balance the optimization direction between the local reinforcement learning goal and the collaborative distillation task. The parameter configuration can ensure that each vehicle-side node can achieve cross-node knowledge fusion and maintain the stability of policy convergence while maintaining local characteristics.

V. RESULTS

DECISION DELAY

To evaluate the real-time response capability of the proposed distributed decision-making system in multi-vehicle collaborative scenarios, five typical simulation environments covering different traffic densities and communication conditions are constructed experimentally, and the average decision delay performance of the three collaborative strategies in each scenario is compared and analyzed. Among the designed comparison methods, the FD+GAT method applies the graph attention mechanism to construct a dynamic perception map between vehicles to extract the influence of neighboring vehicles and uses the FD method to exchange only the policy soft labels, which effectively alleviates the communication bandwidth pressure. The traditional FedAvg method adopts a unified model structure and periodically transmits model parameters. It lacks an adaptation mechanism for heterogeneous models and has a high communication load. The centralized model aggregates the information of each vehicle through a central coordination node and issues a unified strategy. Although the model structure is consistent, the system response relies too much on the central node under communication delay and packet loss conditions. The three strategies are compared in five different scenarios to obtain their respective average decision response times. The experimental results are shown in Figure 3.

Figure 3 shows that in the low-density ideal communication scenario S1, the decision delay of the FD+GAT method is 62ms; FedAvg is 75ms; the centralized model is 81ms. The delay difference is mainly caused by differences in communication load and synchronization mechanism. After the fixed delay is applied in the S2 scenario, FD+GAT increases to 68ms. Compared with 87ms of FedAvg and 94ms of the centralized model, its response advantage is still stable, indicating that the distillation mechanism still maintains better synchronization efficiency under the condition of fixed communication delay. In the S3 scenario, the communication delay fluctuates, and there is packet loss. The FedAvg delay rises to 112ms; the centralized model rises to 129ms; the FD+GAT is controlled at 79ms. The gap has widened significantly. The reason is that the FD structure reduces the dependence on high-frequency communication and reduces the central waiting time through the local perception structure. The road structure of the S4 scenario is complex, and the communication conditions are limited. The FD+GAT is 83ms; the FedAvg is 106ms; the centralized

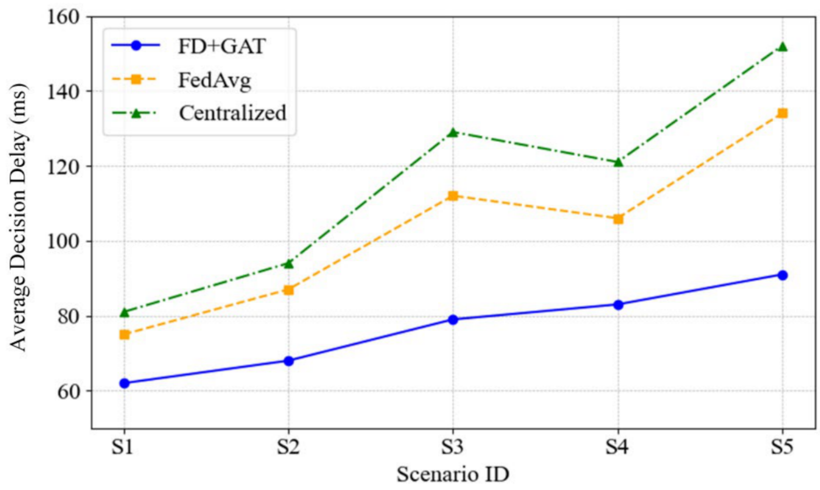

FIGURE 3. Comparison of average decision delays of different methods

model is 121ms, which still shows high robustness. The S5 scenario is the most complex. The centralized model delay rises to 152ms; the FedAvg rises to 134ms; the FD+GAT delay is controlled at 91ms. This is mainly attributed to its use of soft label low-bandwidth transmission and heterogeneous model decoupling strategy, which effectively reduces decision blocking and synchronization conflicts in high-density scenarios. The results in the figure show that the FD+GAT method significantly improves the real-time response performance of distributed collaborative systems in complex communication and road environments.

STRATEGY CONVERGENCE SPEED

In the strategy synchronization mechanism, distillation frequency and graph neural structure have a significant impact on the local strategy convergence efficiency. To analyze the effects of various mechanism combinations on the convergence performance during training, the experiment sets different synergistic strategy combinations. Starting from four configurations, a joint strategy containing FD and GAT graph modeling, a distillation strategy only, a GAT structure only, and a control group without any synergistic mechanism are constructed. The training parameters and initial strategy are fixed, and the same environment distribution and task configuration are used. The recorded loss corresponds to the joint PPO optimization objective, consisting of the clipped policy loss, value regression loss, entropy regularization, and the distillation consistency loss. The horizontal axis represents local training rounds, where each round includes one local PPO update cycle, and distillation synchronization is triggered every predefined communication interval. The results are shown in Figure 4.

Figure 4 shows that the initial loss of the FD+GAT group in round 0 is 1.02, and it drops to 0.02 after training to round 50. The overall decrease is 1.00, and the convergence process is stable, indicating that the synergy of structural modeling and strategy distillation strengthens the strategy's ability to capture global features and shortens the approach path to the optimal solution. The loss of the FD group drops from 0.94 to 0.03 in the same number of rounds, with an overall decrease of 0.91. Although it also achieves effective convergence, there are local fluctuations during the 20th to 40th rounds, reflecting that in the absence of structural associations, policy distillation faces the problem of unstable knowledge transfer. The initial loss of the GAT group is 0.97 and finally drops to 0.07, with a decrease of 0.90. The convergence speed is significantly slower than that of the FD+GAT group. Although structural guidance enhances the information flow between local nodes, the lack of a synchronous convergence mechanism for centralized targets leads to a delay in the convergence path. The loss of the group without a collaborative mechanism drops from 1.00 to 0.18, with an overall decrease of only 0.82. The convergence is slow during training and is in a high loss state for a long time, indicating that the collaborative mechanism plays a core role in compressing the local loss space. The combination of the distillation mechanism and the graph structure shows better convergence performance in policy optimization.

COLLABORATIVE CONSISTENCY EVALUATION

To further quantify the collaborative consistency performance of each distributed collaborative strategy during task execution, three indicators, namely reward volatility, strategy similarity, and frequency of incoordinated behavior, are designed, and a test set containing five typical traffic scenarios is constructed for comparative evaluation. Reward volatility is computed as the variance of cumulative episode rewards across all vehicles over a fixed window of 50 evaluation episodes. Strategy similarity is calculated by averaging cosine similarity between policy output distributions evaluated on the same proxy state set across different vehicles at the same training stage. An

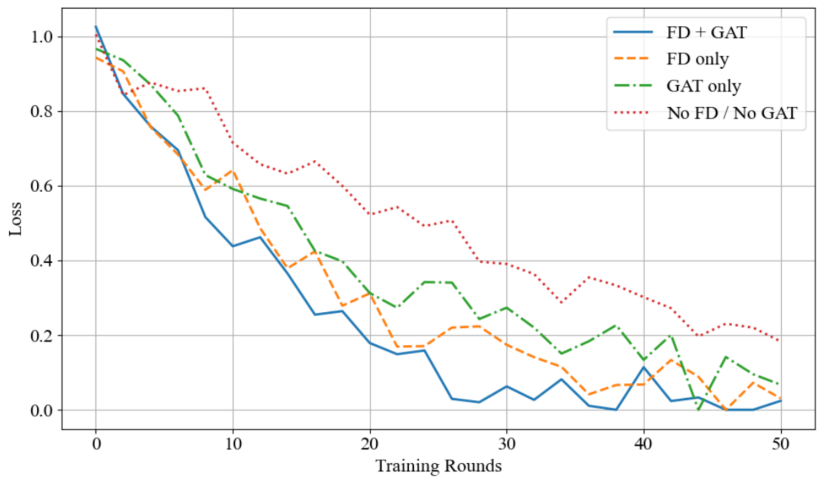

FIGURE 4. The influence of strategy synergy mechanism on convergence speed

incoordinated behavior is recorded when two neighboring vehicles simultaneously select conflicting actions under identical interaction conditions, based on a predefined safety constraint used uniformly across all methods. The Centralized method of centralized joint training relies on global observation to perform centralized decision-making, has complete information input, and has a fast strategy convergence speed. FedAvg, as a federated average strategy, achieves a limited degree of strategy synchronization through parameter aggregation on the basis of maintaining local training privacy. The Local-only method without collaborative learning has each node learning independently, without a communication mechanism, relying entirely on local observations and training independently, lacking a global perspective, and with significant strategy differences. The FD+GAT method integrates the functional decoupling mechanism with GAT, applies structured communication modeling while maintaining distributed deployment, and realizes higher-dimensional strategy feature sharing. The above four methods are deployed and trained in five typical traffic scenarios, and the collected evaluation indicators are quantified. The results are shown in Figure 5.

In terms of reward distribution variance, the variance of the Local-only strategy in the S5 scenario is 0.518, which is much higher than 0.153 of FD+GAT. The reason is that Local-only does not share global information, which leads to the divergence of strategies of each node. The Centralized strategy uses global data input, and its variance in S1 to S5 is maintained between 0.163 and 0.267. However, due to the transmission bottleneck of the centralized architecture, FedAvg is slightly better than Centralized in all scenarios. In S3, its variance is 0.204 while that of Centralized is 0.231. FD+GAT effectively suppresses strategy fluctuations by applying graph structure modeling of collaborative relationships, so that the variance in all scenarios is controlled below 0.153. In terms of policy similarity, the average cosine similarity of FD+GAT in the S1 scenario reaches 0.968, which is higher than 0.936 of FedAvg and 0.917 of Centralized. The reason is that the graph attention mechanism enhances the policy's responsiveness to the behavior of key adjacent nodes and improves policy alignment. In terms of inconsistent behavior, Local-only is triggered 17 times in S5, while FD+GAT is only 3 times, which is 50% less than FedAvg. This shows that FD+GAT achieves more stable decision generation, thereby significantly reducing path conflicts and abnormal control. Comprehensive analysis shows that FD+GAT performs best in all indicators of collaborative consistency.

COMPARISON OF COMMUNICATION LOAD AND EFFICIENCY

To deeply evaluate the collaborative performance of the proposed method between communication overhead and policy update efficiency, the experiment constructs a comparison system of four algorithm schemes, namely FedAvg with traditional parameter aggregation, FedAvg+GAT with graph attention mechanism, FD with distillation mechanism, and FD+GAT with joint graph modeling and distillation strategy. Each method runs independently in a multi-node environment. By counting the average communication load of each node after a certain training cycle and the number of convergence rounds of the local strategy, the trade-off between the communication efficiency of the collaborative update and the strategy optimization ability is characterized. The communication cost is quantitatively evaluated using the average data transmission volume per node in megabytes per synchronization round, including both upload and download volumes. This metric provides a direct measurement of the total communication burden incurred during each aggregation cycle and serves as the basis for comparing communication efficiency across different methods. The reduction ratio between the total communication volume of the proposed FD+GAT method and that of FedAvg is used to indicate compression efficiency, ensuring reproducibility

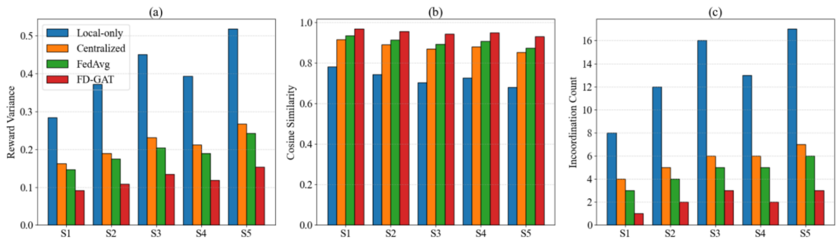

FIGURE 5. Comparison of collaborative consistency; Figure 5 (a). Comparison of reward distribution variance; Figure 5 (b). Comparison of strategy distribution similarity; Figure 5 (c). Comparison of the number of times incoordinated behavior is triggered

of the communication evaluation process. All experiments are based on a unified scenario setting and are conducted under the same initial conditions and network topology to ensure the effectiveness and fairness of the comparison. The node results collected in the experiment are shown in Figure 6.

In Figure 6, the FD+GAT scheme shows a clear trend of gathering in the low value area on both the horizontal and vertical axes, reflecting its dual advantages in communication control and strategy update speed. This phenomenon is mainly attributed to the communication compression effect brought about by the application of soft label transmission by the FD mechanism, and the dynamic modeling of the neighboring vehicle relationship by the graph attention mechanism, which enhances the effectiveness of information fusion and accelerates the strategy learning process. In contrast, the FedAvg method is distributed in a higher value range in both dimensions. This is because it uses complete model parameter synchronization, resulting in high communication frequency and large capacity, and lacks structural perception ability, affecting the efficiency of local strategy optimization. The point distribution of the FedAvg+GAT and FD schemes is in the middle range, which is limited by the unoptimized communication and unmodeled structural relationship, respectively. The overall results show that in the distributed multi-agent system, the combination of distillation mechanism and structural modeling significantly improves the synergy of communication efficiency and strategy convergence.

The comparison of communication volume under identical synchronization intervals uses the average data transmission per node in one round. Table 4 presents the measured upload volume, download volume, and total communication volume for the proposed method and FedAvg.

TABLE 4. Communication volume per node per synchronization round

Method	Upload (MB)	Download (MB)	Total (MB)
FD (proposed)	6.5	6.5	13.0
FedAvg	13.7	13.7	27.4

The proposed method transmits action probability distributions, while FedAvg transmits gradient parameter tensors. Probability distributions contain fewer elements than parameter tensors. The proposed method therefore reduces the total communication volume by 52.6% under the same synchronization frequency.

MODEL ROBUSTNESS AND SCENARIO ADAPTABILITY

This paper sets six measurement indicators and conducts experimental evaluation from the decision delay, convergence rounds, cross-node reward variance, task success rate, retraining frequency, and communication interruption recovery time, so as to deeply explore the stability and adaptability of different FL methods in multi-scenario intelligent agent systems. The experimental environment is set as the five simulation scenarios (S1 to S5) shown in Table 2. By comparing the changes in the indicator values of the four methods (FedAvg, FedAvg+GAT, FD, FD+GAT) in each scenario, six sets of radar charts are constructed to show the comprehensive performance of each method under different task complexities in multiple dimensions. The results are shown in Figure 7.

In terms of average decision delay, the delay of FD+GAT from S1 to S5 is 62ms, 68ms, 79ms, 83ms, and 91ms, respectively, which is always better than other methods. This is related to the application of graph attention mechanism in state modeling to strengthen the weight of key nodes, thereby reducing the delay caused by redundant communication; while FedAvg reaches 134ms in S5. The higher delay is due to the frequent conflicts caused by ignoring the heterogeneity of nodes in the global aggregation process. The number of convergence rounds also shows a consistent trend. FD+GAT converges in the 32nd round in the S5 scenario, which is 14 rounds less than FedAvg. This is due to the reduction of interference from low-quality gradients in its distributed collaborative structure and the increase in the effective update frequency. The significant convergence of cross-node reward variance (FD+GAT is 0.153 in S5) is related to the application of edge feature selection in local strategy migration. This mechanism reduces the propagation of mismatched strategies between adjacent nodes. In terms of task success rate, FD+GAT maintains 85% in the S5 scenario, which is significantly better than FedAvg's 72%.

DATA PRIVACY PROTECTION CAPABILITY EVALUATION

In the process of building a multi-agent collaborative decision-making system, the reversibility of communication content has a direct impact on the risk of data privacy leakage. To verify the ability of the model to resist in-

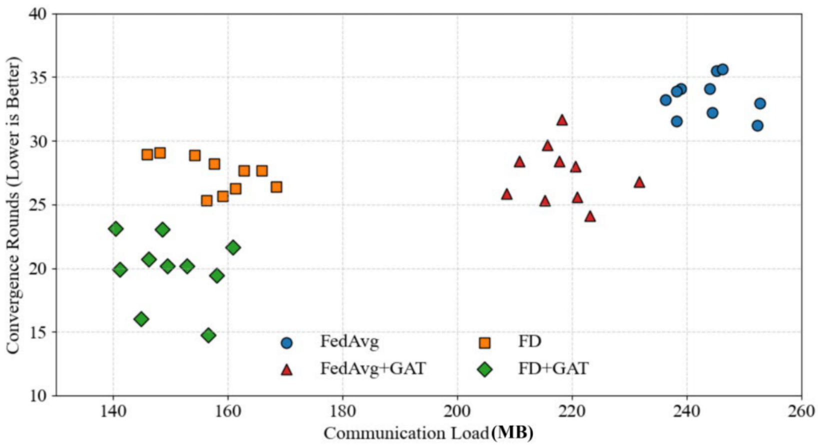

FIGURE 6. Multi-method comparison distribution of communication load and strategy convergence efficiency

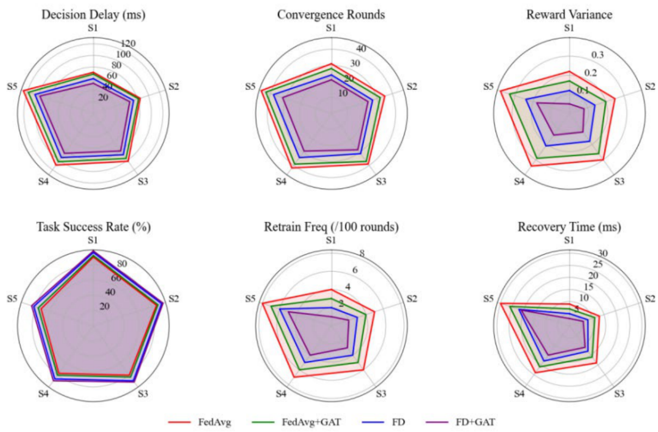

FIGURE 7. Comparison of different methods in multiple scenarios

formation inversion attacks after adopting the soft label distillation mechanism, this paper sets up an inversion attack experiment. By optimizing the inversion method, the input state image disguised by the attacker is constructed, and the output results of FedAvg based on the parameter sharing mechanism and the FD scheme based on the soft label transmission mechanism under 10 rounds of attack are statistically analyzed by the structural similarity index (SSIM) and normalized mean squared error (NMSE). The attack model uses gradient descent to approximate the most likely state feature map representation that is deterministically reshaped from the original normalized numerical state vector before policy inference, rather than any intermediate network activation or learned latent representation. The reconstructed state is represented as a structured feature map derived from the normalized state vector, enabling structural similarity evaluation. The communication content accessible to the attacker is strictly limited to the soft policy output distributions transmitted during federated distillation synchronization. These distributions are produced by the final output layer of the policy network after temperature scaling and do not contain any model parameters, gradients, value estimates, or internal feature representations. Each transmitted distribution is a low-dimensional probability vector whose dimension equals the discrete action space size and whose elements are constrained to a normalized simplex, which fundamentally limits the amount of recoverable information about the original state space. Figure 8 shows the dynamic trends of SSIM and NMSE under the change of the number of attack rounds, which characterizes the degree of exposure of the original state of the two methods in the face of attacks.

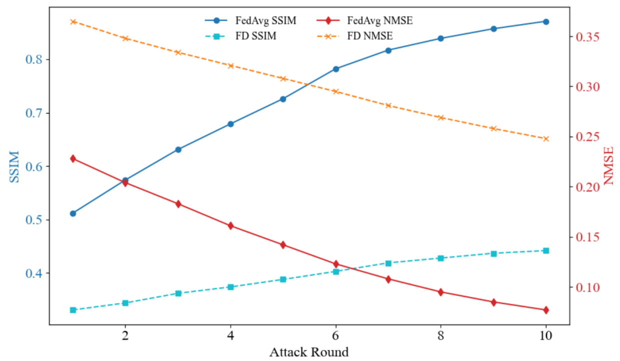

FIGURE 8. Comparison of communication content privacy protection capabilities in intelligent connected vehicle collaborative strategies

The data in Figure 8 shows that when the number of attack rounds increases from 1 to 10, the SSIM of FedAvg increases rapidly from 0.512 to 0.871, and the NMSE decreases from 0.228 to 0.077, which indicates that the exposed model parameters contain a large number of reversible features, allowing the attacker to reconstruct the original state with high precision. In the same attack process, the SSIM of the FD method is always controlled between 0.331 and 0.442, and the NMSE is maintained in the range of 0.365 to 0.248, with a gentle trend change, indicating that the policy distribution of the soft label output lacks explicit mapping characteristics to the input state. The fundamental reason for this result is that the parameter upload of FedAvg contains high-dimensional representations of gradients and weights, and there is an optimizable function path between these tensors and the input state, while the soft label uploaded by the FD method is only the output probability distribution, which does not have the ability to effectively restore the input space. It can be seen that the federated distillation mechanism adopted in this paper significantly improves the privacy protection ability of communication content in the face of inversion attacks while ensuring collaborative performance.

VI. CONCLUSION

This paper builds a heterogeneous collaborative decision-making system integrating FD and GAT to meet the distributed decision-making needs in the multi-agent collaboration of intelligent connected vehicles. The proposed architecture offers a strong methodological analogy for industrial automation and coordinating heterogeneous equipment, such as in the industry. It realizes dynamic neighbor perception by using a local Graph Attention Network (GAT) for weighted information aggregation. The core collaboration mechanism, Federated Distillation (FD), replaces the transmission of large model parameters with lightweight soft labels. This strategy completes strategy alignment between heterogeneous models, significantly reducing communication load while maintaining privacy. Experimental results demonstrate substantial synergistic advantages: in the complex S5 urban scenario, the average decision delay of 91ms is 61 ms lower than that of centralized methods. Policy convergence is highly stable, with local loss dropping from 1.02 to 0.02. Crucially, policy divergence is suppressed, with a cross-node reward variance of 0.153 in S5—far lower than the Local-only baseline of 0.518. The FD+GAT method successfully balances heterogeneous model support, low-bandwidth communication, privacy protection, and structured information fusion, substantially boosting the distributed collaborative efficiency and robustness of intelligent connected vehicles in dynamic environments.

This enhanced robustness and efficiency have excellent engineering application prospects, providing a scalable model for managing decentralized systems in sectors like the industry, where coordinating diverse machinery while maintaining data privacy is vital for reliable, continuous, and high-quality production.

AUTHOR CONTRIBUTIONS

Conceptualization-Yu Zhao, Xiaowei Yang, Zhiwei Guan, Weiqiang Wang and Yinfang Wang; methodology-Yu Zhao, Xiaowei Yang, Zhiwei Guan, Weiqiang Wang and Yinfang Wang; formal analysis-Yu Zhao, Xiaowei Yang, Zhiwei Guan, Weiqiang Wang and Yinfang Wang; investigation-Yu Zhao; resources-Yu Zhao; writing-Yu Zhao, Xiaowei Yang, Zhiwei Guan, Weiqiang Wang and Yinfang Wang. All authors have read and agreed to the published version of the manuscript.

CONFLICTS OF INTEREST

The authors declare no conflict of interest.

FUNDING

This research was funded by Tianjin Municipal Education Commission Scientific Research Project, "Research on Autonomous Lane-changing Strategy of Intelligent Connected Vehicles for Complex Traffic Flow", grant number 2021KJ101.

ACKNOWLEDGEMENTS

Not applicable.

REFERENCES

[1] D. M. Manias and S. Abdallah, "Making a case for federated learning in the internet of vehicles and intelligent transportation systems," IEEE network, vol. 35, no. 3, pp. 88-94, 2021, doi: 10.1109/MNET.011.2000552.

[2] S. S. Shinde and D. Tarchi, "Joint air-ground distributed federated learning for intelligent transportation systems," IEEE Transactions on Intelligent Transportation Systems, vol. 24, no. 9, pp. 9996-10011, 2023, doi: 10.1109/TITS.2023.3265416.

[3] C. Li, Y. Zhang, and Y. Luo, "A federated learning-based edge caching approach for mobile edge computingenabled intelligent connected vehicles," IEEE Transactions on Intelligent Transportation Systems, vol. 24, no. 3, pp. 3360-3369, 2022, doi: 10.1109/TITS.2022.3224395.

[4] S. Zhang, J. Li, L. Shi, M. Ding, D. C. Nguyen, and W. Tan, "Federated learning in intelligent transportation systems: Recent applications and open problems," IEEE Transactions on Intelligent Transportation Systems, vol. 25, no. 5, pp. 3259-3285, 2023, doi: 10.1109/TITS.2023.3324962.

[5] V. P. Chellapandi, L. Yuan, C. G. Brinton, S. H. Zak, and Z. Wang, "Federated learning for connected and automated vehicles: A survey of existing approaches and challenges," IEEE Transactions on Intelligent Vehicles, vol. 9, no. 1, pp. 119-137, 2023, doi: 10.1109/TIV.2023.3332675.

[6] R. Zhang, J. Mao, H. Wang, B. Li, X. Cheng, and L. Yang, “A survey on federated learning in intelligent transportation systems,” IEEE Transactions on Intelligent Vehicles, pp. 1 - 17, 2024.

[7] G. Wang, F. Xu, H. Zhang, and C. Zhao, “Joint resource management for mobility supported federated learning in Internet of Vehicles,” Future Generation Computer Systems, vol. 129, pp. 199-211, 2022, doi: 10.1016/j.future.2021.11.020.

[8] Z. Qu, Y. Tang, G. Muhammad, and P. Tiwari, “Privacy protection in intelligent vehicle networking: A novel federated learning algorithm based on information fusion,” Information Fusion, vol. 98, pp. 101824-101832, 2023, doi: 10.1016/j.inffus.2023.101824.

[9] Q. Gan, B. Li, Z. Xiong, Z. Li, and Y. Liu, “Multi-Vehicle Cooperative Decision-Making in Merging Area Based on Deep Multi-Agent Reinforcement Learning,” Sustainability, vol. 16, no. 22, pp. 9646-9658, 2024, doi: 10.3390/su16229646.

[10] X. Liu, Z. Dong, Z. Xu, S. Liu, and J. Tian, “Enhanced decentralized federated learning based on consensus in connected vehicles,” 2022; 22(09):10722-10730.

[11] H. Liu, S. Zhang, P. Zhang, X. Zhou, X. Shao, and G. Pu, “Blockchain and federated learning for collaborative intrusion detection in vehicular edge computing,” IEEE Transactions on Vehicular Technology, vol. 7, no. 06, pp. 6073-6084, 2021, doi: 10.1109/TVT.2021.3076780.

[12] Chen Jin, M. Chen, G. Zeng, and J. Weng, “BDFL: A byzantine-fault-tolerance decentralized federated learning method for autonomous vehicle,” IEEE Transactions on Vehicular Technology, vol. 70, no. 9, pp. 8639-8652, 2021, doi: 10.1109/TVT.2021.3102121.

[13] C. Carrascosa, A. Pico, M. M. Matagne, M. Rebollo, and J. A. Rincon, “Asynchronous consensus for multi-agent systems and its application to federated learning,” Engineering Applications of Artificial Intelligence, vol. 135, pp. 108840-108853, 2024, doi: 10.1016/j.engappai.2024.108840.

[14] X. Zhou, W. Liang, A. Kawai, K. Fueda, J. She, and K. I. K. Wang, “Adaptive segmentation enhanced asynchronous federated learning for sustainable intelligent transportation systems,” IEEE Transactions on Intelligent Transportation Systems, vol. 25, no. 7, pp. 6658-6666, 2024, doi: 10.1109/TITS.2024.3362058.

[15] X. Xu, R. Li, Z. Zhao, and H. Zhang, “The gradient convergence bound of federated multi-agent reinforcement learning with efficient communication,” IEEE Transactions on Wireless Communications, vol. 23, no. 1, pp. 507-528, 2023, doi: 10.1109/TWC.2023.3279268.

[16] I. Ergun, H. U. Sami, and B. Guler, “Sparsified secure aggregation for privacy-preserving federated learning,” 2021; 21(12):12872-12899.