Graph Reinforcement Learning based Resource Allocation Method in RIS-aided Heterogeneous IoV

Wang Zeng*, Huijun Tang†, Pinlong Zhao*, Pengfei Jiao*, Huaifeng Shi‡, Huaming Wu§ and Hongjian Sun†
*School of Cyberspace, Hangzhou Dianzi University, Hangzhou 310018, China
†Department of Engineering, Durham University, Durham DH1 3LE, United Kingdom
‡Nanjing University of Information Science and Technology, Nanjing 210044, China
§Center for Applied Mathematics, Tianjin University, Tianjin 300072, China
Emails: 242270069@hdu.edu.cn, huijun.tang@durham.ac.uk, {pinlongzhao, pjiao}@hdu.edu.cn,
shihuaifeng@nuist.edu.cn, whming@tju.edu.cn, hongjian.sun@durham.ac.uk

ICC 2026 - IEEE International Conference on Communications | 979-8-3195-4209-0/26/$31.00 ©2026 IEEE | DOI: 10.1109/ICC59461.2026.11587894

***Abstract*—Reconfigurable Intelligent Surface (RIS)-assisted vehicular networks entail formidable challenges in resource allocation due to Vehicle-to-Infrastructure (V2I) and Vehicle-to-Vehicle (V2V) communications. However, the heterogeneous roles of vehicles and the spatial correlations among network entities are often overlooked in resource allocation schemes, leading to unstable convergence and suboptimal decisions. To address these issues, we propose a Graph Neural Network (GNN)-enhanced Twin Delayed Deep Deterministic Policy Gradient (TD3) framework for joint optimization of power allocation, channel selection, and RIS phase configuration. A Graph Attention Network (GAT)-based encoder extracts topological dependencies and generates structured state representations for policy learning. Extensive simulations validate the effectiveness of the proposed approach. The results show that GNN-TD3 achieves faster convergence, higher reward stability, and superior robustness across varying learning rates, vehicle densities, and transmission powers. Moreover, it significantly improves Age of Information (AoI) and V2V success probability, demonstrating strong adaptability to fine-grained heterogeneous vehicular environments.**

***Index Terms*—Reconfigurable Intelligent Surface, Vehicle-to-Everything, Graph Neural Network, Deep Reinforcement Learning**

I. Introduction

With the continuous advancement of Intelligent Transportation Systems (ITS) and Vehicle-to-Everything (V2X), V2V and V2I communications play a pivotal role in future autonomous driving and cooperative sensing [1], [2]. Efficient and reliable communication connectivity not only enhances road safety and resource utilization but also provides the foundation for vehicular perception, task offloading, and environmental cooperation [3]. However, real-world vehicular networks exhibit significant fine-grained heterogeneity—vehicles differ substantially in communication capabilities and QoS requirements, while V2I and V2V links demonstrate fundamentally distinct traffic characteristics. This multi-level heterogeneity increases resource allocation difficulty and demands fine-grained coordination between V2I and V2V links [4].

To address complex propagation environments and channel blockage, RIS have been introduced into vehicular networks [5]. RIS can reshape channel propagation characteristics with low power consumption, enhancing signal quality and connectivity reliability [6], [7]. However, RIS introduction complicates spatial coupling relationships, leading to high-dimensional, non-convex, and highly coupled optimization problems [8]–[10]. In fine-grained heterogeneous vehicular networks, where different vehicle types have varying QoS demands and V2I and V2V links require differentiated resource strategies, RIS configurations must dynamically balance multiple objectives at both vehicle-level and link-level granularities.

Deep Reinforcement Learning (DRL) has been widely applied in vehicular networks for power control, resource allocation, and link management [11]–[13]. However, current DRL-based approaches exhibit two significant shortcomings. First, most studies adopt homogeneous vehicle assumptions, ignoring fine-grained heterogeneity at both the vehicle type level and link type level. At the vehicle type level, different vehicle categories exhibit vastly different communication priorities, latency tolerances, and reliability requirements. At the link type level, V2I links prioritize sustained connectivity and stable data rates for infrastructure communication, while V2V links emphasize stringent latency and high reliability for safety-critical direct communication where packet loss can lead to dangerous information gaps. This coarse-grained modeling overlooks differentiated resource allocation needs, leading to poor generalization [14]. Second, traditional DRL state vector representations disrupt the inherent spatial topology among vehicles, RIS, and base stations, obscuring interactive dependencies and resulting in unstable policy learning. To address these issues, we propose a fine-grained optimization framework for RIS-assisted heterogeneous vehicular networks by integrating GNN and DRL. The main contributions are:

- **Role-Aware Heterogeneous System Modeling:** We develop a role-aware heterogeneous system model that explicitly characterizes the diverse roles of vehicles and their fine-grained communication demands. By jointly modeling the distinct V2I and V2V link requirements of different vehicle roles, the proposed framework enables adaptive resource allocation tailored to each communication mode. This modeling design effectively improves AoI for V2I links and enhances the success probability of V2V

Corresponding author: Huijun Tang

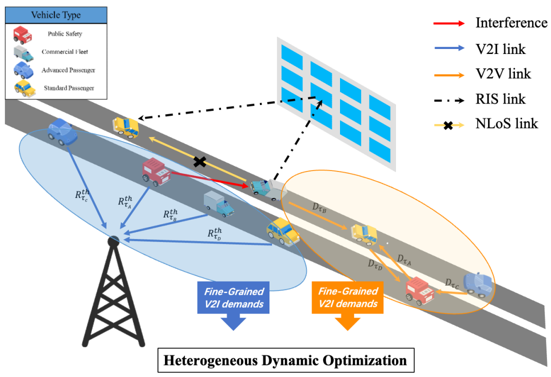

Fig. 1. The System Architecture. Illustration of an RIS-assisted heterogeneous vehicular network with fine-grained modeling, where four vehicle types have differentiated QoS requirements.

transmissions, ensuring balanced performance across the heterogeneous vehicular network services.

- **GNN-DRL Framework for Structured State Encoding and Resource Optimization:** We design a GNN-DRL framework that captures fine-grained spatial structural relationships. By encoding vehicle-level features, link-level characteristics, and topological dependencies, the framework enables differentiated resource management strategies that simultaneously address V2I rate stability requirements and V2V latency and reliability constraints across heterogeneous vehicle categories.
- We systematically verify the framework's advantages over baseline algorithms. The results confirm that fine-grained heterogeneity modeling at both vehicle type and link type levels, combined with GNN-based feature extraction, substantially enhances DRL decision-making capabilities.

II. System Model

A. Network Architecture

As illustrated in Fig 1, we consider a RIS-assisted heterogeneous vehicular network scenario. The system comprises: 1) a multi-antenna base station; 2) an RIS containing N reflectors arranged in a uniform linear array; 3) V single-antenna vehicles. This comprises two link types: **V2I links**: Uplink connections from V vehicles to the base station, with each link allocated a dedicated channel; **V2V links**: Direct connections between K pairs of vehicles, reuse sub-channels from V2I links.

The available spectrum is divided into multiple orthogonal subchannels allocated to vehicles for V2I transmission. V2V links reuse these subchannels for communication, where the binary variable $s_{v,k}^t$ indicates whether the k-th V2V link shares the subchannel of the v-th V2I link at time slot t.

The RIS consists of a uniform planar array with N passive reflecting elements, each capable of adjusting the phase and amplitude of the incident signal. Taking the V2I link as an example, the received signal includes both the direct and RIS-reflected components. The reflection matrix at time slot t is:

$$\Theta_t = \mathrm{diag}\left(b_1 e^{j\theta_1^t}, b_2 e^{j\theta_2^t}, \ldots, b_N e^{j\theta_N^t}\right), \tag{1}$$

where $b_n \in [0,1]$ and θ_n^t denote the amplitude reflection coefficient and phase shift of the n-th element, respectively. Due to hardware limitations, the phase is quantized discretely:

$$\theta_n^t \in \left\{0, \frac{2\pi}{Q}, \ldots, \frac{2\pi(Q-1)}{Q}\right\}. \tag{2}$$

where Q represents the phase quantization resolution controlling the beamforming precision.

B. Heterogeneous Vehicle Classification

To capture the inherent heterogeneity of vehicular networks, vehicles are categorized into four types $\mathscr{T} = \{\tau_A, \tau_B, \tau_C, \tau_D\}$, each with distinct QoS requirements and traffic patterns. Let $\mathscr{V}_\tau$ denote the set of vehicles of type τ.

Category A – Public Safety (τ_A): Emergency vehicles requiring ultra-reliable low-latency communication (URLLC).

Category B – Commercial Fleet (τ_B): Logistics and delivery vehicles demanding enhanced mobile broadband (eMBB) for robust connectivity.

Category C – Advanced Passenger (τ_C): Consumer vehicles equipped with ADAS and infotainment systems for high data-rate services.

Category D – Standard Passenger (τ_D): Basic connected vehicles supporting essential safety.

C. Communication Model

We denote the coordinates of the base station and RIS as (X_B, Y_B, Z_B) and (X_R, Y_R, Z_R), respectively. Each vehicle v of type τ occupies position $(X_{v,t}, Y_{v,t}, Z_{v,t})$ at time t. Following [15], the RIS-assisted vehicular network channel comprises both direct and indirect components. The direct channel gain takes the form:

$$g_{tra,rec}^t = \sqrt{\rho\kappa\left(d_{tra,rec}^t\right)^{-\eta}}\,\xi_{tra,rec}^t, \tag{3}$$

where $tra \in \{k, B\}$ and $rec \in \{v, k, B\}$ denote the transmitter and receiver, ρ represents the path loss constant at reference distance $d_0 = 1$ m, κ captures log-normal shadow fading, $d_{tra,rec}^t$ measures the geometric distance at time slot t, and η specifies the path loss exponent. The term $\xi_{tra,rec}^t$ represents frequency-dependent small-scale fading, which follows an exponential distribution with unit mean, where $d_{tra,rec}^t$ denotes the three-dimensional Euclidean distance between the transmitter and the receiver at time slot t, which is given by $d_{tra,rec}^t = \sqrt{(X_{tra} - X_{rec})^2 + (Y_{tra} - Y_{rec})^2 + (Z_{tra} - Z_{rec})^2}$, where $(X_{tra}, Y_{tra}, Z_{tra})$ and $(X_{rec}, Y_{rec}, Z_{rec})$ denote the Cartesian coordinates of the transmitter and receiver, respectively.

The RIS reflection creates an indirect channel:

$$g_{tra,R}^t = \sqrt{\rho\kappa\left(d_{tra,R}^t\right)^{-\eta}}\, e^{-j2\pi\frac{d_{tra,R}^t}{\lambda}}\,\xi_{\mathrm{AoA}}^t, \tag{4}$$

where the arrival array response ξ_{AoA}^t from transmitter to RIS satisfies:

$$\xi_{\mathrm{AoA}}^t = [1, \cdots, e^{-j2\pi\frac{d}{\lambda}(N-1)\sin(\theta_{AoA}^t)}], \tag{5}$$

with λ denoting the subchannel wavelength and θ_{AoA}^t representing the angle of arrival at time slot t. Here, $tra \in \{k, B\}$.

We express the total channel gain for the V2I link of vehicle $v \in \mathscr{V}_\tau$ as $g_{v,\tau}^t = (g_{R,B}^t)^H \Theta_t g_{v,R}^t + g_{v,B}^t$ and for the V2V link pair k where the transmitting vehicle belongs to type τ as $g_{k,\tau}^t = (g_{R,k'}^t)^H \Theta_t g_{k,R}^t + g_{k,k'}^t$, where k and k' designate the V2V transmitter and receiver.

At time t, the SINR for the V2I link of type-τ vehicle v and the V2V link pair k is:

$$SINR_{v,\tau}^t = \frac{P_v^t \left|g_{v,\tau}^t\right|^2}{\sum\limits_{k\in K} s_{v,k}^t P_{k,\tau}^t \left|g_{k,\tau}^t\right|^2 + \sigma^2}, \tag{6}$$

$$SINR_{k,\tau}^t = \frac{P_{k,\tau}^t \left|g_{k,\tau}^t\right|^2}{\sum\limits_{k'\in K, k'\neq k} s_{v,k'}^t P_{k',\tau}^t \left|g_{k',\tau}^t\right|^2 + P_v^t \left|g_{v,\tau}^t\right|^2 + \sigma^2}, \tag{7}$$

where P_v^t denotes the V2I transmission power at time slot t, which is fixed for all vehicle types, $P_{k,\tau}^t$ represents the V2V transmission power for type-τ vehicles at time slot t, and σ^2 denotes the communication noise power. The corresponding transmission rates follow:

$$R_{v,\tau}^t = W \log_2(1 + SINR_{v,\tau}^t), \tag{8}$$

$$R_{k,\tau}^t = W \log_2(1 + SINR_{k,\tau}^t). \tag{9}$$

where W denotes the spectral subband bandwidth.

D. Analysis of Timeliness and Reliability

AoI is directly related to vehicle safety—outdated vehicle positions can lead to collisions, delayed traffic data can result in inefficient route planning, and lagging emergency alerts can hinder timely responses [16]. Heterogeneous applications have differentiated requirements for timeliness: emergency vehicles require sub-second updates for coordinated dispatch, while passenger vehicles can tolerate moderate delays to meet entertainment needs. For vehicle $v \in \mathscr{V}_\tau$ in category τ, the category-adaptive information age evolution process is:

$$A_{v,\tau}^{t+1} = \begin{cases} \Delta t, & \text{if } R_{v,\tau}^t \geq R_\tau^{th}, \\ A_{v,\tau}^t + \Delta t, & \text{otherwise.} \end{cases} \tag{10}$$

The adaptive threshold R_τ^{th} enables differentiated service provisioning across categories, with varying thresholds for different types of vehicles.

The reliability of V2V communication enables coordinated lane changes, collision avoidance, and platooning [17]. Transmission failures can create dangerous information gaps, leading to accidents and traffic disruptions. Emergency vehicles require near-perfect information delivery to coordinate incidents, while ordinary vehicles can tolerate occasional message loss. For V2V link k where transmitting vehicle $v \in \mathscr{V}_\tau$ belongs to category τ:

$$\mathscr{P}_{k,\tau} = \Pr\left\{ \sum_{t=1}^{T} s_{v,k}^t \cdot R_{k,\tau}^t \cdot \Delta t \geq D_\tau \right\}, \tag{11}$$

where $s_{v,k}^t \in \{0,1\}$ denotes whether vehicle v selects channel k at time t, $R_{k,\tau}^t$ is the corresponding achievable transmission rate for type-τ vehicle, Δt is the time slot duration, and D_τ is the data load requirement of type-τ vehicles.

E. Problem Formulation

The heterogeneous network optimization problem aims to maximize utility while satisfying diverse QoS requirements:

$$\min_{\mathbf{P},\mathbf{S},\Theta} \left\{ \frac{1}{V} \sum_{\tau\in\mathscr{T}} \sum_{v\in\mathscr{V}_\tau} A_{v,\tau}^t - \frac{1}{K} \sum_{\tau\in\mathscr{T}} \sum_{k\in K} \mathscr{P}_{k,\tau} \right\} \tag{12}$$

$$\textbf{s.t.} \quad R_{v,\tau}^t \geq R_\tau^{\text{th}}, \quad \forall v \in \mathscr{V}_\tau, \forall \tau \in \mathscr{T}, \tag{13}$$

$$s_{v,k}^t \in \{0,1\}, \quad \forall v \in V, \forall k \in K, \tag{14}$$

$$\sum_{v\in V} s_{v,k}^t \leq 1, \quad \forall k \in K, \tag{15}$$

$$P_{k,\tau}^t \in [P_{\min}, P_{\max}], \quad \forall k \in K, \forall \tau \in \mathscr{T}, \tag{16}$$

$$\theta_n^t \in \left\{0, \tfrac{2\pi}{Q}, \cdots, \tfrac{2\pi(Q-1)}{Q}\right\}, \quad \forall n \in N. \tag{17}$$

III. Method

We propose a heterogeneous vehicular network resource allocation algorithm based on DRL, combining a GNN encoder and the TD3 algorithm to achieve joint optimization of channel selection, power allocation, and RIS phase shifts, meeting the differentiated QoS requirements of various types of vehicles. TD3 is adopted due to its superior performance in continuous action spaces: compared to DDPG, TD3 mitigates value overestimation through clipped double Q-learning, reduces policy variance via target policy smoothing, and improves training stability with delayed policy updates, making it well-suited for the high-dimensional joint optimization problem in RIS-assisted vehicular networks.

A reinforcement learning framework comprises several fundamental components: an Agent, an Environment, a state space O, an action space A, and an immediate reward function R. At each time step, the base station (Agent) observes the current state o_t from the Environment and executes an action a_t. The Agent then receives an immediate reward r_t corresponding to action a_t and transitions to the next state o_{t+1}. The following sections define the state space O, action space A, and immediate reward function R.

State Space $\mathscr{O}$: The state $o_t \in O$ observed by the Agent at time slot t is concretely represented as follows:

$$o_v^t = \{g_v^t, g_{v,k}^t, I_v^{t-1}, D_v^t, AoI_v^t, R_v^{th}\}, \tag{18}$$

$$O_t = \{s_1^t, s_2^t, \ldots, s_V^t\}, \tag{19}$$

At time t, the state of the v-th vehicle is: g_v^t represents the V2I channel gain, $g_{v,k}^t$ represents the V2V channel gain, I_v^{t-1} represents the V2V interference level experienced by the v-th vehicle at time $t-1$, D_v^t denotes the remaining transmission load of the vehicle, AoI_v^t indicates AoI, and R_v^{th} is the transmission rate threshold for the current type of vehicle.

Action Space $\mathscr{A}$: After observing o_t, the agent takes action a_t, which is specifically represented as follows:

$$a_t = (\mathbf{S}^t,\ \mathbf{P}^t,\ \Theta^t), \tag{20}$$

where $\mathbf{S}^t = \{s_{v,k}^t\}_{v\in\mathscr{V}, k\in K}$ and $\mathbf{P}^t = \{P_{k,\tau}^t\}_{k\in K, \tau\in\mathscr{T}}$ represent the channel selection and power allocation for all V2V pairs, respectively, and Θ^t represents the phase shift configuration of the RIS.

Reward Function $\mathscr{R}$: After the agent takes an action, it receives a reward r_t, which is specifically expressed as follows:

$$r_t = -w_1 \frac{1}{V} \sum_{v=1}^{V} A_{v,\tau}^t - w_2 \frac{1}{K} \sum_{k=1}^{K} \frac{D_{\tau,k}^t}{D_\tau}, \tag{21}$$

where w_1 and w_2 are the balance coefficients used to balance the two parts of the reward, $D_{\tau,k}^t$ represents the remaining load of the k-th V2V vehicle at time t, and D_τ represents the initial load size of the k-th vehicle.

A. GNN Encoder

To process the structured state information, we designed an encoder based on GAT [18] to transform the vehicle network state into a graph representation enriched with spatial relationships.

Graph structure construction: Model the vehicular network as an undirected graph $\mathscr{G} = (\mathscr{V}, \mathscr{E})$, where:

- Node set $\mathscr{V}$: Each vehicle is treated as a node, with node features represented by s_v^t.
- Edge set $\mathscr{E}$: Constructed based on vehicle neighbor relationships, $E = (i,j) | j \in \mathscr{N}(i), i \neq j$.

Message passing mechanism: Employing GAT for information aggregation:

$$h_i^{(l+1)} = \sigma\left(\sum_{j \in \mathscr{N}(i)} e_{ij} W^{(l)} h_j^{(l)}\right), \tag{22}$$

where $h_i^{(l)}$ is the feature representation of node i at layer l, $\mathscr{N}(i)$ is the set of neighboring nodes of node i, $W^{(l)}$ is the learnable weight matrix at layer l, $e_{ij} \in [0,1]$ is the normalized attention coefficient between node i and its neighbor j, and $\sigma(\cdot)$ is the ELU activation function.

The attention coefficient α_{ij} is calculated as follows:

$$e_{ij} = \frac{\exp(\text{LeakyReLU}(\mathbf{a}^T [W\mathbf{h}_i \parallel W\mathbf{h}_j]))}{\sum_{j \in \mathscr{N}(i)} \exp(\text{LeakyReLU}(\mathbf{a}^T [W\mathbf{h}_i \parallel W\mathbf{h}_j]))}. \tag{23}$$

Graph-level representation learning: Obtaining compact graph-level representations through global average pooling:

$$z_t = \frac{1}{V} \sum_{v=1}^{V} h_v^{(L)}, \tag{24}$$

where $h_v^{(L)} \in \mathbb{R}^d$ is the output feature of node v at the final GNN layer, and $z_t \in \mathbb{R}^d$ is the aggregated graph-level representation. This encoded state z_t serves as the input to the TD3 algorithm, replacing the original high-dimensional state O_t.

B. TD3

We adopt the TD3 algorithm [19] for joint optimization of channel selection, power allocation, and RIS phase configuration. Compared to DDPG, TD3 mitigates value overestimation through clipped double Q-learning, target policy smoothing, and delayed policy updates, ensuring stable training and reliable policy convergence.

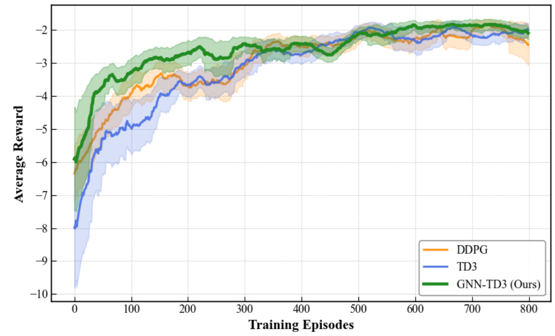

Fig. 2. Reward comparison.

TD3 employs a dual-critic-single-actor architecture: an actor network $\mu(z_t|\theta^\mu)$ generates actions from GNN-encoded states z_t; two critics $Q_1(z_t, a_t|\theta^{Q_1})$, $Q_2(z_t, a_t|\theta^{Q_2})$ independently estimate state-action values; and corresponding target networks μ', Q_1', Q_2' provide stable target computation.

Clipped Double Q-Learning computes the target using the minimum of two critics to reduce overestimation:

$$y = r_t + \gamma \min_{i=1,2} Q_i'\left(z_{t+1}, \mu'(z_{t+1}|\theta^{\mu'}) + \varepsilon \middle| \theta^{Q_i'}\right), \tag{25}$$

where $\gamma \in [0,1]$ is the discount factor. **Target Policy Smoothing** adds clipped noise $\varepsilon \sim \text{clip}(\mathscr{N}(0,\sigma), -c, c)$ to target actions, preventing sharp peaks in the value function. **Delayed Policy Updates** update the actor less frequently than the critics, enhancing stability.

Critic optimization minimizes temporal-difference error:

$$\mathscr{L}(\theta^{Q_i}) = \mathbb{E}_{(z_t, a_t, r_t, z_{t+1}) \sim \mathscr{D}}\left[\left(Q_i(z_t, a_t|\theta^{Q_i}) - y\right)^2\right], \tag{26}$$

where $\mathscr{D}$ is the experience replay buffer. The dual critics and min operation effectively suppress Q-value overestimation.

Actor optimization maximizes expected return:

$$\nabla_{\theta^\mu} J \approx \mathbb{E}_{z_t \sim \mathscr{D}}\left[\nabla_a Q_1(z_t, a|\theta^{Q_1})|_{a=\mu(z_t)} \nabla_{\theta^\mu} \mu(z_t|\theta^\mu)\right]. \tag{27}$$

Gradients from Q_1 alone avoid inconsistencies from the dual-critic structure.

Soft target updates track online parameters via exponential moving average:

$$\theta^{\mu'} \leftarrow \zeta \theta^\mu + (1-\zeta)\theta^{\mu'}, \tag{28}$$

$$\theta^{Q_i'} \leftarrow \zeta \theta^{Q_i} + (1-\zeta)\theta^{Q_i'}. \tag{29}$$

where ζ is the soft update coefficient.

IV. SIMULATION RESULTS AND DISCUSSIONS

A. Parameters Setup

This section evaluates the performance of the proposed GNN-enhanced TD3 algorithm for resource allocation in RIS-assisted heterogeneous vehicular networks. The simulation parameters are configured based on [11] and shown in TABLE I. We construct a heterogeneous vehicular network in an urban intersection scenario, where the base station is located at $[180, 270]$ m with a height of 25 m. The RIS is deployed at $[290, 375]$ m with a height of 25 m to assist

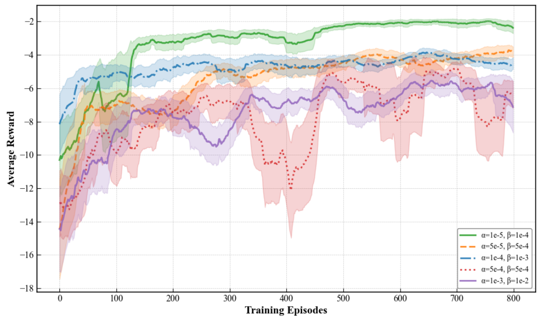

Fig. 3. The Effect of Varying Learning Rates on the Average Reward

vehicles experiencing non-line-of-sight conditions. Vehicles are uniformly distributed along four approaching directions with speeds ranging from $[10, 15]$ m/s. The antenna gains of the base station and vehicles are 8 dB and 3 dB, respectively, while the receiver noise powers are 5 dB and 11 dB, respectively. The carrier frequency is 2 GHz, with each subchannel allocated a bandwidth of 1 MHz. The noise power σ_1 is -114 dBm, the V2I transmission power is 23 dBm, and the V2V transmission power ranges from $[1, 2, \ldots, 23]$ dBm. The path loss models for V2I and V2V channels refer to the literature [20], [21].

The proposed algorithm employs two-layer fully connected neural networks for both Actor and Critic networks, with a hidden layer dimension of 128. GAT encoder adopts a two-layer structure. The learning rate $\mathscr{D}$ is set to 10^{-4}, discount factor $\gamma = 0.95$, experience replay buffer size is 10^6, and soft update coefficient $\zeta = 0.005$. At the end of the training, we test the trained model with each test consisting of 50 episodes, and each result was the average of these 50 episodes.

TABLE I
CONFIGURATION PARAMETERS FOR HETEROGENEOUS VEHICLE TYPES

Vehicle Type	**Distribution Ratio (%)**	**V2I Rate Threshold (Mbps)**	**Payload Size (bits)**	**Typical Application**
Public Safety	3	8.0	16,960	Ambulance, Police
Commercial Fleet	25	5.0	12,720	Delivery, Taxi
Advanced Passenger	50	3.0	8,480	Smart Car
Standard Passenger	22	1.0	6,784	Regular Car

B. Performance Evaluation

This section verifies the performance of the proposed GNN-TD3 algorithm and compares it with four benchmark methods: traditional TD3, DDPG [22], random allocation (Random), and Greedy algorithm. The experiments are conducted under scenarios with varying numbers of vehicles.

Fig. 2 presents the training convergence of the three algorithms. GNN-TD3 converges faster, achieves higher final performance, and exhibits the least fluctuation, demonstrating that the GNN encoder effectively improves learning efficiency and policy quality by capturing spatial topological structures. To identify optimal hyperparameters, we performed a sensitivity analysis on the Actor and Critic learning rates. Fig. 3 shows that learning rates strongly affect training stability and performance: higher rates cause oscillations and instability, while lower rates slow convergence. The notation (α, β) in the legend denotes the learning rate pair for the Actor and Critic networks, respectively.

Fig 4 illustrates the performance comparison of different algorithms under varying numbers of vehicles. As shown in Fig 4(a), GNN-TD3 consistently achieves the lowest violation of diverse QoS requirements across all vehicle densities, demonstrating superior ability to simultaneously satisfy heterogeneous service demands. Fig 4(b) presents the sum of AoI, where GNN-TD3 maintains the freshest information and significantly outperforms all baseline methods. Fig 4(c) shows that GNN-TD3 achieves the highest success probability throughout the entire range, substantially exceeding the random policy. These results validate the effectiveness of our proposed GNN-TD3 algorithm in handling increasing vehicle density while respecting diverse QoS requirements.

Fig 5 demonstrates the impact of V2I transmission power on system performance. As the transmission power decreases, Fig 5(a) shows that violations of diverse QoS requirements increase for all algorithms, with GNN-TD3 maintaining the lowest violation rate throughout. Fig 5(b) reveals that lower transmission power leads to higher AoI values, indicating reduced information freshness. Fig 5(c) presents a counter-intuitive trend where success probability increases with decreasing power, which may be attributed to reduced interference in the network. Overall, GNN-TD3 demonstrates robust performance across different power levels, consistently satisfying diverse QoS requirements better than baseline methods.

V. CONCLUSION

This paper proposed a GNN-enhanced DRL framework for joint power allocation, channel selection, and RIS phase optimization in heterogeneous vehicular networks. By integrating a GAT encoder with the TD3 algorithm, the framework captures spatial correlations among vehicles, BS, and RIS while addressing diverse QoS requirements. Simulation results confirm that the proposed method outperforms conventional DRL and greedy baselines in convergence speed, information freshness, and V2V reliability. Future work will explore multi-agent extensions, adaptive GNN structures, and joint communication–sensing optimization.

ACKNOWLEDGMENT

This work is supported by the National Natural Science Foundation of China (No. 62401190), the Zhejiang Provincial Natural Science Foundation of China (LQN25F020016), the Zhejiang Province Key R&D Program Project (2024C01102), the Natural Science Foundation of Tianjin (No. 25JCY-BJC01540) and the Engineering and Physical Sciences Research Council grant (EP/X040518/1 and EP/Y037421/1).

REFERENCES

[1] C. Yang, C. F. Kwong, D. Chieng, P. Kar, K.-L. A. Yau, and Y. Chen, "Navigating the road ahead: a comprehensive survey of radio resource allocation for vehicle platooning in c-v2x communications," *IEEE Communications Surveys & Tutorials*, 2024.

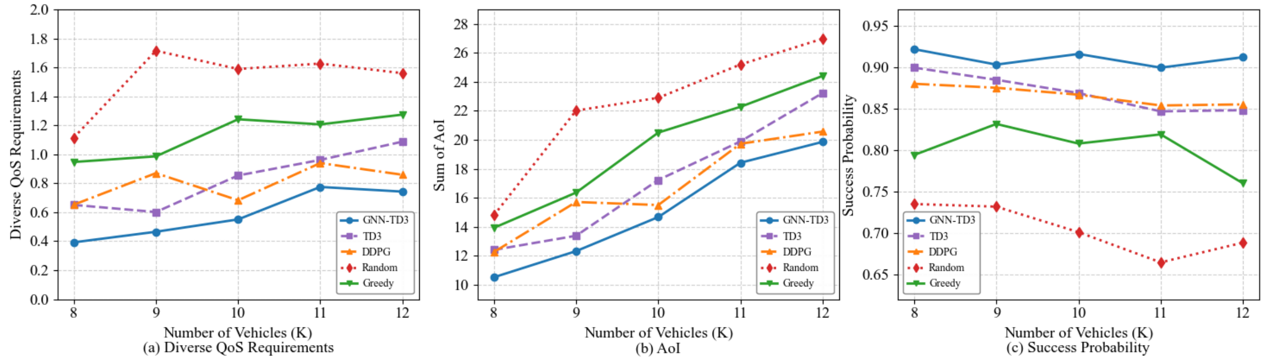

Fig. 4. Performance Comparison of Different Algorithms under Varying Number of Vehicles.

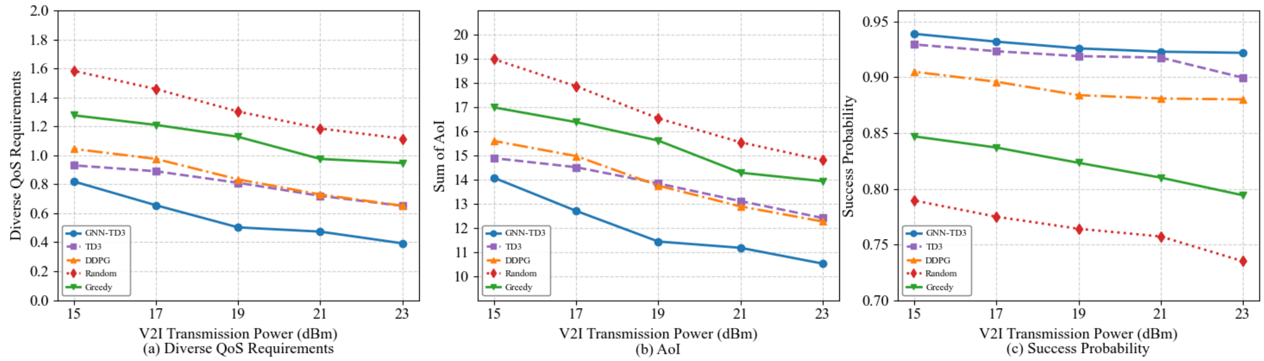

Fig. 5. Impact of V2I Transmission Power on System Performance.

[2] W. Fan, Y. Su, J. Liu, S. Li, W. Huang, F. Wu, and Y. Liu, "Joint task offloading and resource allocation for vehicular edge computing based on v2i and v2v modes," *IEEE Transactions on Intelligent Transportation Systems*, vol. 24, no. 4, pp. 4277–4292, 2023.

[3] F. Dettinger, M. Weiß, D. Dittler, J. Stümpfle, M. Artelt, and M. Weyrich, "A survey on performance, current and future usage of vehicle-to-everything communication standards," *arXiv preprint arXiv:2410.10264*, 2024.

[4] J. Zhao, X. Xiong, Q. Zhang, S. Ren, J. Chen, W. Xu, and D. Wang, "Boosting vehicular connectivity through resource allocation algorithm based on heterogeneous agent proximal policy optimization," *Vehicular Communications*, vol. 50, p. 100856, 2024.

[5] Y. U. Ozcan, O. Ozdemir, and G. K. Kurt, "Reconfigurable intelligent surfaces for the connectivity of autonomous vehicles," *IEEE Transactions on Vehicular Technology*, vol. 70, no. 3, pp. 2508–2513, 2021.

[6] M. Deng, M. Ahmed, A. Wahid, A. A. Soofi, W. U. Khan, F. Xu, M. Asif, and Z. Han, "Reconfigurable intelligent surfaces enabled vehicular communications: A comprehensive survey of recent advances and future challenges," *IEEE Transactions on Intelligent Vehicles*, 2024.

[7] H. Tang, J. Zhang, Z. Zhao, H. Wu, H. Sun, and P. Jiao, "Joint optimization based on two-phase gnn in ris- and df-assisted miso systems with fine-grained rate demands," *IEEE Transactions on Wireless Communications*, vol. 24, no. 12, pp. 9989–10 002, 2025.

[8] S. Pala, M. Katwe, K. Singh, T. A. Tsiftsis, and C.-P. Li, "Robust transmission design for ris-aided full-duplex-rsma v2x communications via multi-agent drl," *IEEE Transactions on Vehicular Technology*, 2024.

[9] S. Abeywickrama, R. Zhang, Q. Wu, and C. Yuen, "Intelligent reflecting surface: Practical phase shift model and beamforming optimization," *IEEE Transactions on Communications*, vol. 68, no. 9, pp. 5849–5863, 2020.

[10] Z. Lu, Y. Zhao, X. Li, H. Wu, C.-Z. Xu, and Q. Xue, "Hybrid Reconfigurable Intelligent Surface for Integrated Cooperative Localization and Communication for 6G V2X System ," *IEEE Transactions on Mobile Computing*, no. 01, pp. 1–15, Dec. 2025.

[11] K. Qi, Q. Wu, P. Fan, N. Cheng, W. Chen, J. Wang, and K. B. Letaief, "Deep-reinforcement-learning-based aoi-aware resource allocation for ris-aided iov networks," *IEEE Transactions on Vehicular Technology*, 2024.

[12] B. Wang, L. Liu, and J. Wang, "Actor-critic based drl algorithm for task offloading performance optimization in vehicle edge computing," in *2023 International Wireless Communications and Mobile Computing (IWCMC)*. IEEE, 2023, pp. 93–98.

[13] Q. Chen, X. Song, T. Song, and Y. Yang, "Vehicular edge computing networks optimization via drl-based communication resource allocation and load balancing," *IEEE Transactions on Mobile Computing*, 2025.

[14] J. Tian, Q. Liu, H. Zhang, and D. Wu, "Multiagent deep-reinforcement-learning-based resource allocation for heterogeneous qos guarantees for vehicular networks," *IEEE Internet of Things Journal*, vol. 9, no. 3, pp. 1683–1695, 2021.

[15] C. Guo, L. Liang, and G. Y. Li, "Resource allocation for high-reliability low-latency vehicular communications with packet retransmission," *IEEE Transactions on Vehicular Technology*, vol. 68, no. 7, pp. 6219–6230, 2019.

[16] A. Muhammad, M. Elhattab, M. A. Arfaoui, A. Al-Hilo, and C. Assi, "Age of information optimization in ris-assisted wireless networks," *IEEE Transactions on Network and Service Management*, vol. 21, no. 1, pp. 925–938, 2023.

[17] L. Liang, H. Ye, and G. Y. Li, "Spectrum sharing in vehicular networks based on multi-agent reinforcement learning," *IEEE Journal on Selected Areas in Communications*, vol. 37, no. 10, pp. 2282–2292, 2019.

[18] P. Veličković, G. Cucurull, A. Casanova, A. Romero, P. Lio, and Y. Bengio, "Graph attention networks," *arXiv preprint arXiv:1710.10903*, 2017.

[19] S. Fujimoto, H. Hoof, and D. Meger, "Addressing function approximation error in actor-critic methods," in *International conference on machine learning*. PMLR, 2018, pp. 1587–1596.

[20] P. Merias, "Study on lte-based v2x services (release 14) technology," *Technical Re-1164 port 3GPP TR*, vol. 36, 2016.

[21] I. Winner, "Winner ii channel models," *IST-4-027756 WINNER II, D. 1. 1. 2 V1. 2*, 2007.

[22] T. P. Lillicrap, J. J. Hunt, A. Pritzel, N. Heess, T. Erez, Y. Tassa, D. Silver, and D. Wierstra, "Continuous control with deep reinforcement learning," *arXiv preprint arXiv:1509.02971*, 2015.