Simulation Modelling Practice and Theory 148 (2026) 103268

Contents lists available at ScienceDirect

Simulation Modelling Practice and Theory

journal homepage: www.elsevier.com/locate/simpat

# Cooperative decision-making in mixed traffic via conflict-aware Heterogeneous Graph Reinforcement Learning

Guohong Zheng [a,b], Yiyang Chen [a,b], Zhigang Wu [a,b], Zhaocheng He [a,b,c], Haipeng Zeng [a,b,*]

[a] School of Intelligent Systems Engineering, Sun Yat-Sen University, Shenzhen, 518107, Guangdong, China
[b] Guangdong Provincial Key Laboratory of Intelligent Transportation Systems, Shenzhen, 518107, Guangdong, China
[c] Pengcheng Laboratory, Shenzhen, 518107, Guangdong, China

ARTICLE INFO

*Keywords:*
Mixed traffic flow
Intersection management
Conflict-aware modeling
Heterogeneous Graph Reinforcement Learning

ABSTRACT

Intersections are critical nodes of urban road networks where heterogeneous traffic flows converge. Yet, they are also major bottlenecks of efficiency and safety, particularly under mixed autonomy where Connected and Automated Vehicles (CAVs) and Human-Driven Vehicles (HDVs) coexist. Traditional control schemes such as traffic lights or optimization-based scheduling often fail to capture the complex and uncertain interactions in such settings. We propose a Conflict-Aware Heterogeneous Graph-based Reinforcement Learning (HGRL) Framework for decentralized intersection management under mixed traffic flows. In our approach, the traffic environment is represented as a heterogeneous interaction graph, where edges encode both cooperative relations and potential conflicts. Building on this representation, a heterogeneous graph-based reinforcement learning controller enables CAVs to make adaptive and coordinated decisions while explicitly accounting for conflict risks. Comprehensive simulations conducted in SUMO across varying CAV penetration rates demonstrate the effectiveness, robustness, and scalability of the proposed framework. Our method achieves consistent improvements across all key performance indicators compared with strong baselines.

## 1. Introduction

With the acceleration of urbanization and the rapid growth in vehicle ownership, traffic congestion and accidents have emerged as pressing challenges for urban traffic management. According to INRIX (2020) [1], in New York City, drivers lost 100 hours in 2020, costing 1486 per driver and 7.7 billion citywide. PEP (Transport, Health, and Environment Pan-European Programme) reported that transport-related congestion, crashes, and environmental impacts cost Europe nearly €820 million in 2019. Intersections, as critical nodes in urban road networks, concentrate large traffic volumes and complex vehicular maneuvers, thereby becoming hotspots for conflicts and accidents. Data from the U.S. Federal Highway Administration indicate that over 2.8 million intersection-related crashes occur annually, accounting for 44% of all traffic accidents [2]. Consequently, enhancing both the efficiency and safety of traffic flows at intersections has become a central issue in modern traffic management research.

In response to these challenges, national policies increasingly emphasize the deployment of Intelligent Transportation Systems (ITS) and Connected and Automated Vehicles (CAVs). Through vehicle-to-infrastructure (V2I) and vehicle-to-vehicle (V2V) communication, CAV technologies offer the potential to increase throughput, reduce collision risks, improve fuel efficiency, and lower

* Corresponding author at: School of Intelligent Systems Engineering, Sun Yat-Sen University, Shenzhen, 518107, Guangdong, China.
*E-mail addresses:* zhenggh8@mail2.sysu.edu.cn (G. Zheng), chenyy553@mail2.sysu.edu.cn (Y. Chen), wuzhg6@mail2.sysu.edu.cn (Z. Wu), hezhch@mail.sysu.edu.cn (Z. He), zenghp5@mail.sysu.edu.cn (H. Zeng).

https://doi.org/10.1016/j.simpat.2026.103268
Received 11 October 2025; Received in revised form 24 January 2026; Accepted 26 February 2026
Available online 26 February 2026

emissions [3]. For instance, since 2016 the U.S. National Highway Traffic Safety Administration has required all new vehicles to be equipped with internet connectivity [4]. Recent policy documents released by the State Council [5] highlight the strategic importance the Chinese government places on fostering the development of CAVs. Similarly, the U.S. ITS Strategic Plan (2020–2025) underscores the importance of intelligent and connected traffic control technologies, positioning the nation at the forefront of global ITS innovation.

Despite policy advances and technological innovation, effective control strategies for intersections with mixed traffic flows remain underdeveloped. Real-world intersections are characterized by high traffic density, heterogeneous vehicle compositions, and frequent conflict points. In mixed traffic scenarios, where CAVs and human-driven vehicles (HDVs) coexist, variations in human driving behaviors, together with the limited penetration of CAVs undermine the effectiveness of traditional traffic control strategies. Prior approaches—including reservation-based systems [6], predictive scheduling [7], and machine learning-based adaptive controllers [8] have shown promise in efficiency improvement, yet most are designed for fully connected environments or rely on overly simplified assumptions, thereby failing to capture the complex conflict dynamics inherent in mixed autonomy traffic.

Against this backdrop, explicitly addressing conflict modeling at intersections has become essential. While recent studies have begun to integrate conflict analysis into control frameworks — recognizing that safety and efficiency must be considered jointly [9] — existing methods still struggle to capture heterogeneous interactions accurately. Compounding this challenge, the high risks and costs of field experiments render simulation platforms like SUMO [10] indispensable for developing and testing such strategies [11].

Leveraging this simulation paradigm to tackle the core problem of interaction modeling, we propose a novel intersection management framework that is developed and validated within simulation environments to explicitly account for conflict relationships in mixed traffic. Our approach centers on a heterogeneous graph-based reinforcement learning model, which is trained and evaluated using detailed simulation data to capture the complex interactions between CAVs and HDVs at intersections. In the proposed framework, vehicles are represented as nodes in a dynamic interaction graph, where edges encode both cooperative relations and potential conflicts. This graph representation is then used to guide a multi-agent reinforcement learning controller, enabling vehicles to make adaptive, coordinated decisions. By combining graph-based state representation with reinforcement learning, our framework demonstrates significant improvements in intersection efficiency and safety in a simulation environment.

The contributions of this paper are threefold:

- **Modeling framework**: We construct a conflict-aware heterogeneous graph representation of mixed traffic intersections, capable of capturing both cooperative and conflict interactions among heterogeneous vehicles.
- **Control strategy**: We propose a conflict-aware heterogeneous graph reinforcement learning (HGRL) framework that explicitly couples conflict-centric graph representation with sparse attention mechanisms, restricting message passing to safety-critical interactions. By integrating conflict-aware execution constraints into the control loop, the framework achieves robust and interpretable decision-making beyond conventional graph reinforcement learning intersection control methods.
- **Simulation validation**: We conduct simulations under varying CAV penetration rates and show that our framework outperforms benchmark methods in traffic efficiency and safety. We further perform ablation studies to verify the effectiveness of the proposed design.

## 2. Related work

### 2.1. Representation learning for traffic intersections

As critical nodes where traffic flows converge and diverge, intersections are highly susceptible to congestion and accidents. Statistics indicate that over one-third of traffic congestion and more than half of traffic accidents occur at intersections [12–14]. Therefore, effective intersection management requires a precise quantification of their traffic states, which is the foundation for alleviating congestion and reducing safety risks. Consequently, extracting intersection-level features is essential for applications such as traffic network partitioning [15], key node identification [16], speed prediction [17], and signal control [18].

Traditional approaches generally transform road networks into graph structures and analyze them using centrality or connectivity measures [19–22]. While useful, these methods capture only shallow structural information and fail to uncover complex interdependencies between intersections. More recently, network representation learning (NRL) has been introduced to extract latent features while preserving topological and attribute information automatically [23,24]. Classical methods such as DeepWalk and Node2vec [25,26] learn node embeddings via random walks, whereas Struc2vec [27] focuses on structural similarity. However, these approaches are limited in modeling nonlinear or dynamic dependencies. Graph neural networks (GNNs), including GCNs and GATs, have since emerged to jointly leverage node attributes and neighborhood structures [28–30].

In the transportation domain, GNN-based models have shown promise in encoding vehicle interactions under uncertainty [31]. For example, [32] applied GNNs to cooperative autonomous driving, while [33] proposed a spatiotemporal dynamic graph framework for ramp merging. Similarly, [34] modeled trajectory conflicts via graph-based scheduling methods. These studies highlight the potential of graph-based learning but also reveal the need for more effective representations of conflict relationships in mixed traffic flows.

In this context, our work models intersections through a conflict-aware graph representation, which explicitly encodes both cooperative and conflict vehicle interactions. This modeling formulation serves as the foundation of our framework and ensures that key conflict structures are directly incorporated into the learning process.

**Table 1**
Conflict set $C$: each movement and its conflicting directions (right-turns omitted as non-conflicting).

| Movement | Conflicting directions |
|---|---|
| N-L | {E-S, E-L, S-S, W-L} |
| N-S | {E-S, W-L, S-L, W-S} |
| E-L | {E-S, N-L, S-S, W-S} |
| E-S | {W-L, N-S, N-L, S-S} |
| S-L | {N-S, E-L, W-S, W-L} |
| S-S | {N-L, E-S, E-L, W-S} |
| W-L | {S-L, N-S, E-S, N-L} |
| W-S | {N-S, E-L, S-L, S-S} |

### 2.2. Intersection management methods

Coordinating vehicle movements at intersections has traditionally relied on hierarchical control frameworks, where a central controller schedules arrivals and distributed controllers execute them [35]. Various methods, such as fuzzy logic, model predictive control, and optimal control [6,36,37], have been explored. Reservation-based strategies, such as FIFO [38] and batch-based release [39,40], attempt to avoid collisions but rarely achieve global efficiency. Optimization-based methods (e.g., integer programming, dynamic programming) can yield optimal schedules, yet the solution space grows exponentially with traffic volume, limiting scalability [7,41].

Reinforcement learning (RL) has therefore been widely adopted as it circumvents explicit modeling of the scheduling problem. Deep RL further addresses the curse of dimensionality and has shown strong performance in mixed traffic environments [42,43]. However, most RL approaches neglect detailed vehicle interactions, leading to limited cooperative behavior. Recent advances in graph reinforcement learning (GRL) attempt to overcome this by embedding vehicle interactions into graph structures, enabling more effective policy learning [44,45].

Despite these advances, few studies directly address trajectory planning for CAVs in mixed intersections. Existing works often treat vehicles individually [46,47], or optimize mixed formations [48], but they largely ignore the explicit conflict structures that drive congestion and safety risks.

Due to the expense and safety concerns of on-road experiments, simulation methods have been widely adopted by scholars to examine the potential of vehicle management at intersections. Zhang et al. [36] used SUMO to study the platooning behaviors in fixed-time signalized intersection. Li et al. [49] proposed a vehicle group behavior model based on acceptable gap theory and evaluated collaborative control strategies, with extensive testing conducted using VISSIM simulation software.

To address these limitations, we propose a heterogeneous graph-based reinforcement learning strategy that leverages the structural properties of the conflict-aware interaction graph. By integrating cooperative and conflict relations into the policy-learning process, our approach achieves safe and efficient coordination between CAVs and HDVs. Then, we conduct simulation experiments in SUMO to validate the proposed strategy.

## 3. Problem statement

### 3.1. Intersection traffic

This study considers a prototypical four-leg unsignalized intersection operating under mixed traffic conditions, as illustrated in Fig. 1(a). Each approach (North, East, South, West) is divided into three dedicated lanes corresponding to left-turn (L), through (C), and right-turn (R) movements. For example, a northbound left-turn movement is denoted as N-L, a southbound through movement as S-C. Traffic demand is modeled stochastically: vehicle arrivals on each lane follow random processes, and the type of each arriving vehicle is also probabilistic, comprising either CAVs or HDVs. This configuration captures the heterogeneity and randomness inherent in real-world urban intersections.

In this study, conflicts are operationalized as potential interaction risks, identified when the projected trajectories of two vehicle movements overlap within the conflict area of the intersection. For instance, vehicles traveling on N-S and S-L movements will cross paths in the center of the intersection, thereby forming a conflict pair. Table 1 lists, for each movement direction, the set of other directions with which it conflicts. Conflicts among vehicle movements are a primary source of congestion and accidents at intersections, as they originate from contradictory action decisions between streams. While reward-based penalties discourage unsafe decisions, conflicts may still arise when reinforcement learning vehicles (RVs) attempt to enter the intersection under competing traffic streams. To maintain both safety and flow stability, we employ a conflict-resolution mechanism that post-processes RL outputs before execution. Specifically, an RV may proceed only if the intersection is clear of vehicles from conflicting directions; otherwise, its Go decision is overridden to Stop. When multiple RVs from conflicting streams simultaneously choose Go, priority is given to the stream with the longest waiting time, allowing those vehicles to pass first while others yield.

The following assumptions are made regarding vehicle states and driving behaviors. (1) Full state observability and information sharing. We assume that the positions, velocities, accelerations, and related kinematic attributes of all vehicles are fully observable and shareable across the network. Although HDVs lack onboard sensing and communication capabilities, roadside detection

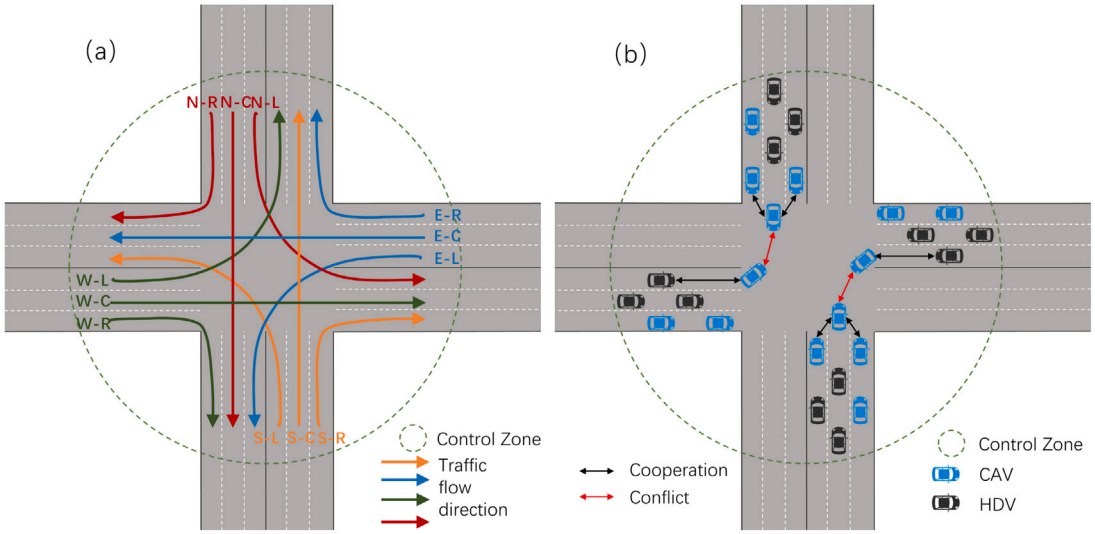


**Fig. 1.** Illustration of the intersection environment and conflict determination. (a) Lane-level movement definitions for each approach, including left-turn (L), through (C), and right-turn (R) movements. (b) Mixed traffic scenario with CAVs (blue) and HDVs (black).

infrastructure (e.g., loop detectors or LiDAR units) is assumed to capture their instantaneous states and broadcast this information to nearby CAVs. (2) HDV driving behavior model. The longitudinal and lateral movements of HDVs are governed by human driving characteristics. Specifically, their trajectories and car-following behaviors are modeled using the Intelligent Driver Model (IDM) [50], which is widely adopted in traffic simulation.

### 3.2. Decentralized RL for mixed traffic

We formulate the mixed-traffic control problem as a partially observable Markov decision process (POMDP), defined by the following 7-tuple:

$$(S, A, O, T, R, \Omega, \gamma) \tag{1}$$

where $S$ is the set of underlying states, $A$ the set of actions, $O$ the set of observations, $T$ the state-transition function, $R$ the reward function, $\Omega$ the observation function, and $\gamma$ the discount factor.

#### 3.2.1. Action space

Since this study focuses on improving intersection throughput, the control objective is to manage CAVs traversing an unsignalized intersection under mixed traffic conditions. Accordingly, the action space of each CAV is restricted to $A = \{Stop, Go\}$. The selected action determines whether the CAV enters the intersection or remains stopped at the stop line.

In simulation, longitudinal acceleration is computed using the IDM. Outside the control zone, all vehicles follow standard IDM dynamics. Within the control zone, if the CAV chooses "Go", it accelerates at its maximum allowable rate $a_{max}$; conversely, if the action is "Stop", it decelerates according to IDM until it halts at the stop line. In the event of a potential collision, the CAV's emergency braking mechanism overrides the nominal control, applying a deceleration greater than the standard requirement to avoid a crash.

#### 3.2.2. Observation space

To enhance generalizability across intersections with different topologies, we define the observation space by separating ego-level and global-level information.

**Ego information.** The ego state of each CAV is encoded as a 4-tuple:

$$S_{ego} = \{x, y, speed, acc\}, \tag{2}$$

where $x$ and $y$ denote the vehicle's center-point coordinates in a Cartesian reference frame; $speed$ represents the scalar magnitude of its velocity, and $acc$ represents the scalar magnitude of its acceleration.

**Global information.** The global information is characterized by two lane-level indicators for each inbound lane of the intersection: the queue length $l_i$ and the average waiting time $w_i$. The value of $l_i$ is computed as the number of vehicles lined up in lane $i$ before reaching the stop line, while $w_i$ is calculated as the average waiting time of these vehicles. Together, $\{l_i, w_i\}$ quantify the spatial and temporal levels of congestion, enabling CAVs to anticipate traffic pressure beyond their own trajectory.

The complete observation for each controlled CAV is thus:

$$O = \{S_{ego}, S_{global}\} \tag{3}$$

This formulation ensures that each agent bases its decision not only on its own kinematic state but also on the broader congestion context of the intersection. In simulation, both ego and global features are directly extracted from the microscopic states of all vehicles. In real-world applications, ego information can be obtained via onboard sensors, while global information may be estimated through V2V communication among CAVs.

#### 3.2.3. Reward function

To balance traversal efficiency, emission, and safety, we define a composite reward function for each RL vehicle as follows, inspired by prior designs in [51]. But we merge jerk and fuel consumption into a unified eco-driving component, which provides a clearer trade-off between smoothness and energy efficiency.

$$r(s_t, a_t, s_{t+1}) = w_w r_w + w_v r_v + w_j p_j + p_c, \tag{4}$$

where $w_w, w_v, w_j$ are weighting coefficients, and $p_c$ are penalties for conflict behaviors. The terms are defined as follows:

- **Waiting-time reward** ($r_w$): Encourages vehicles to minimize excessive idling. If the vehicle chooses to proceed, a positive reward proportional to its reduction in waiting time is given; otherwise, a penalty is applied.

$$\begin{cases} -w^{t+1,j}, & \text{if } a^t = \text{Stop}; \\ w^{t+1,j}, & \text{otherwise.} \end{cases} \tag{5}$$

- **Speed reward** ($r_v$): Normalized instantaneous speed of the ego vehicle,

$$r_v = \frac{v}{v_{\max}}, \tag{6}$$

  where $v$ is the current velocity and $v_{\max}$ is the maximum allowable speed. This term promotes efficient progression through the intersection.
- **Jerk penalty** ($p_j$): Penalizes abrupt acceleration changes to encourage smooth and fuel-efficient driving,

$$p_j = \frac{|a_t - a_{t-1}|}{J_{\max}}, \tag{7}$$

  where $a_t$ and $a_{t-1}$ denote current and previous accelerations, respectively, and $J_{\max}$ is a saturation constant.
- **Conflict penalty** ($p_c$): A fixed penalty of $-1$ is imposed if the vehicle's action results in a potential conflict with another vehicle in the intersection control zone.

In our implementation, the weights are set to $w_w = 1.0$, $w_v = 0.3$, and $w_j = -1.0$, ensuring a balanced trade-off between efficiency, comfort, and safety. The normalized structure of each component ensures comparability in scale and enhances the stability of policy learning.

## 4. Methodology

This section provides a detailed exposition of the decision-making problem formulated using a heterogeneous graph-based reinforcement learning within mixed autonomous traffic flows, including Basic Framework, Heterogeneous Graph Representation, and Heterogeneous Graph Reinforcement Learning (HGRL).

### 4.1. Basic framework

The basic framework of the GRL-based decision-making system for RVs is illustrated in Fig. 2. The basic framework consists of three parts: the scenario construction module, the heterogeneous graph representation module, the HGRL module, including the HGNN model and the DRL model. Specifically, the scenario construction module provides the dynamic environment of mixed traffic flows, including both CAVs and HDVs, which is defined in Section 3.1; the heterogeneous graph representation module encodes vehicles as graph nodes and their cooperative or conflict relations as edges, thereby transforming raw traffic states into structured graph features; In the HGRL module, the HGNN model serves as the core feature extractor, learning high-dimensional embeddings that capture both local and global interaction patterns; the DRL model leverages the extracted graph embeddings to train a policy network, which outputs discrete driving decisions (Stop/Go) for each controlled CAV through an actor–critic architecture.

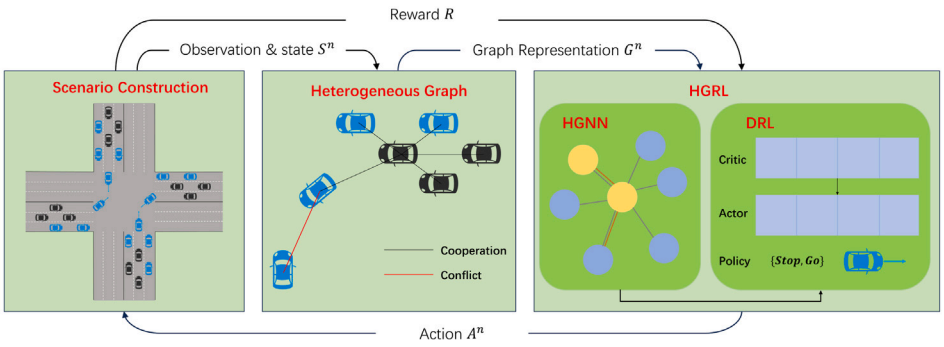


**Fig. 2.** Framework of our method for mixed traffic control. It consists of three main modules: (1) Scenario Construction module, where traffic situations are generated; (2) Heterogeneous Graph module, which models cooperation and conflict relations between vehicles; (3) HGRL module, where the HGNN model encodes graph embeddings and the DRL model learns policies (Stop/Go) based on actor–critic structure.

### 4.2. Heterogeneous graph representation

Building upon the intersection environment defined in Section 3.1, we represent the mixed traffic scenario as a heterogeneous undirected graph $G = \{V, E\}$, where $V = \{v_i, i \in \{1, 2, \dots, n\}\}$ is the set of nodes, each $v_i$ encoding the state of the $i_{th}$ vehicle; $E = \{e_{ij}, i, j \in \{1, 2, \dots, n\}\}$ is the set of edges, where $e_{ij}$ represents the pairwise interaction between vehicle $i$ and vehicle $j$. Specifically, $n$ denotes the number of RVs in the driving loop. The construction of this graph involves three components: node feature matrix, edge attributes, and the adjacency matrix.

**Node feature matrix**. The node feature matrix encapsulates the kinematic observations of all RL vehicles in the control zone. Formally, it is defined as:

$$N = [x_1, x_2, \dots, x_n]^T, \tag{8}$$

where each $x_i$ corresponds to the observation vector of the RL vehicle, ensuring consistency with the reinforcement learning formulation described in formulation (3).

**Edge attributes**. Edges encode pairwise interactions between vehicles and consist of cooperation edges and conflict edges. The definition involves both connectivity judgment (whether an edge exists) and edge values (what information the edge carries).

- Cooperative edges. A cooperative edge is created if the Euclidean distance between vehicles $i$ and $j$ does not exceed the communication threshold $\rho_c$:

$$e_{ij}^{coop} = \begin{cases} d_{ij}, & d_{ij} \le \rho_c, \\ 0, & d_{ij} > \rho_c, \end{cases} \tag{9}$$

  where $\rho_c$ denotes the communication range of CAVs and $d_{ij}$ the Euclidean distance between vehicle $i$ and vehicle $j$. This criterion implies that any two CAVs within the specified communication range can establish a link and exchange information, thereby enabling cooperative interaction. In our experiments, we set $\rho_c = 50$ m.
- Conflict edges. A conflict edge is established if the ego observation of vehicle $i$ identifies vehicle $j$ as a potential conflict partner. Formally,

$$e_{ij}^{conf} = \begin{cases} obs_{ego}^{i}, & \text{if vehicle } i \text{ conflict with vehicle } j, \\ 0, & \text{otherwise}, \end{cases} \tag{10}$$

  In this case, the connectivity is determined by the conflict-detection mechanism, while the edge weight is directly taken from the ego observation vector of vehicle $i$, described in formulation (2).

**Adjacency matrix**. The final adjacency matrix integrates both cooperative and conflict relations:

$$D = D_{coop} + D_{conf}, \tag{11}$$

where $D_{coop} = \{e_{ij}^{coop}\}_{i,j=1}^{n}$ encodes distance-based cooperative interactions, and $D_{conf} = \{e_{ij}^{conf}\}_{i,j=1}^{n}$ encodes conflict interactions derived from ego observations. By jointly incorporating proximity-driven cooperation and safety-critical conflict detection, the adjacency matrix provides a comprehensive representation of inter-vehicle interactions, which is then processed by the HGNN model.

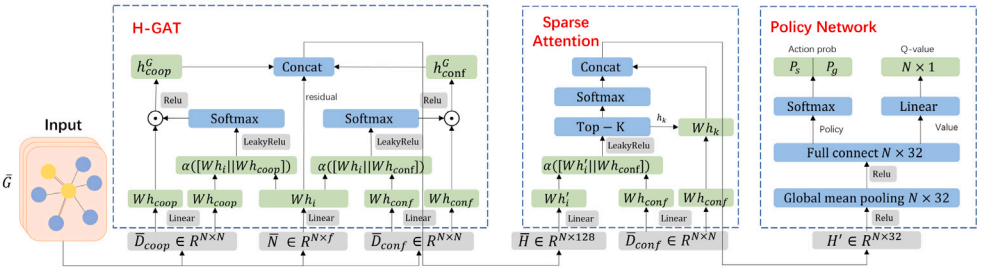


**Fig. 3.** Detailed architecture of the proposed HGNN model, consisting of three components: H-GAT for extracting cooperative and conflict features, sparse attention for refining conflict relations, and a policy network for decision-making.

### 4.3. Heterogeneous Graph Reinforcement Learning (HGRL)

#### 4.3.1. Heterogeneous Graph Neural Network (HGNN) model

The overall architecture of the proposed heterogeneous graph reinforcement learning framework is illustrated in Fig. 3, which consists of three primary components: a heterogeneous graph attention network (H-GAT), a sparse attention mechanism, and a policy network.

(1) Heterogeneous Graph Attention Network (H-GAT): We represent the mixed traffic scenario as a heterogeneous graph $G = (N, D)$, where nodes denote vehicles with kinematic states, and edges capture their *cooperative* or *conflict* relations. The node feature matrix is denoted by $\bar{N} \in \mathbb{R}^{N\times f}$, and the adjacency matrices for cooperative and conflict relations are denoted by $\bar{D}_{coop}, \bar{D}_{conf} \in \mathbb{R}^{N\times N}$.

To capture heterogeneous interactions, we adopt parallel graph attention layers. For each edge $(i, j)$, the attention score is computed as:

$$Attention(e_{ij}) = a^{\top}\,\mathrm{LeakyReLU}\left(W \cdot [h_i \parallel h_j]\right), \tag{12}$$

where $W$ is a learnable weight matrix, $h_i$ and $h_j$ are the features of nodes $i$ and $j$, $[\cdot \parallel \cdot]$ denotes concatenation, and $a^{\top}$ is a learnable attention vector. Separate GAT heads are applied to cooperative and conflict edges, and their outputs are aggregated with residual connections $H \in \mathbb{R}^{N\times 128}$.

This process allows H-GAT to jointly model coordination and competition among vehicles, thereby providing rich contextual node embeddings.

(2) Sparse Attention Mechanism: While cooperative edges typically represent stable and low-risk interactions, conflict edges correspond to safety-critical situations that directly influence intersection decision-making. Directly aggregating information from all conflict edges may still introduce redundancy, especially in dense traffic scenarios where multiple conflicts coexist. To address this issue, we design a Top-K sparse attention mechanism conditioned on conflict relations, which explicitly selects the most informative conflict interactions.

From a modeling perspective, this design introduces a structural prior that emphasizes safety-critical interactions while suppressing less influential conflicts. In mixed traffic intersections, not all conflict relations contribute equally to control decisions. By ranking conflict edges according to their attention scores and retaining only the top-K most relevant ones, the model focuses representation learning on dominant conflict patterns, thereby improving robustness and learning stability.

Given the node embedding $\bar{H}$ from H-GAT and the conflict edge set $D_{conf}$, we apply a transformer-style attention mechanism:

$$h_i' = \sum_{j\in\mathcal{N}_i^{conf}} \alpha_{ij} W_v h_j, \quad \alpha_{ij} = \frac{\exp\left((W_q h_i)^{\top}(W_k h_j)\right)}{\sum_{k\in\mathcal{N}_i^{conf}} \exp\left((W_q h_i)^{\top}(W_k h_k)\right)}, \tag{13}$$

where $W_q, W_k, W_v$ are query, key, and value projections. Only the Top-K conflict neighbors with the highest attention scores are retained for message aggregation, while the remaining conflict edges are masked out.

From a computational perspective, dense attention over all agents scales as $O(N^2)$, whereas the proposed Top-K sparse attention scales as $O(K\cdot|\mathcal{D}_{conf}|)$, with $K \ll |\mathcal{N}_{conf}|$ in typical intersection scenarios. This explicit sparsification significantly reduces redundant message passing and improves scalability, while preserving the most influential conflict interactions.

The resulting conflict-enhanced embeddings are then pooled with a global mean pooling layer to obtain a compact graph-level representation $z \in \mathbb{R}^{B\times 32}$, which is used as input to the downstream reinforcement learning policy network.

(3) Policy Network: The pooled embedding $z$ serves as the input to the policy network, which follows an actor–critic design. The policy network consists of fully connected layers and activation functions to output both action probabilities and value estimates for each CAV. Let $\Phi_{\text{policy}}$ denote the policy operator. The network first projects the aggregated embedding $z$ into a latent representation through a fully connected layer ($N \times 64$). This representation is then divided into two branches: a policy head, which applies a softmax to produce the action probabilities $p_s, p_g$, and a value head, which outputs the state-value $Q(s^{n,T}, a^n)$ via a linear layer:

$$Q(s^{n,T}, a^n) = \Phi_{\text{policy}}(z). \tag{14}$$

**Table 2**
Experimental intersection scenarios and configuration settings.

| Item | Intersection I | Intersection II |
|---|---|---|
| Layout type | Four-leg | T-shaped |
| Intersection size | 30 m × 30 m | 30 m × 30 m |
| Lane length | 100 m | 100 m |
| Traffic demand | 600/900/1200 veh/h | 900 veh/h |
| CAV penetration | 50%–100% | 50%–100% |
| Vehicle assignment | Random | Random |

Here $Q(s^{n,T}, a^n)$ denotes the expected return of taking action $a^n$ under observation $s^{n,T}$. This value serves as a guiding principle for determining the driving behavior of CAVs. Specifically, the formula is given as:

$$\pi(a|s) = \text{Softmax}(W_\pi z + b_\pi), \quad V(s) = W_v z + b_v, \tag{15}$$

where $\pi(a|s)$ denotes the policy distribution and $V(s)$ is the value function. This dual-head design allows the model to jointly optimize safety-oriented decision making and value estimation.

#### 4.3.2. Deep Reinforcement Learning (DRL) model

The graph embeddings generated by the HGNN model are integrated into a reinforcement learning framework based on Proximal Policy Optimization (PPO) [52]. PPO is an actor–critic algorithm that provides stable and sample-efficient policy updates by employing a clipping mechanism to constrain gradient steps. This makes it particularly suitable for large-scale, multi-agent traffic environments.

In this architecture, the actor network corresponds to the policy network, which maps HGNN embeddings into action probabilities $(P_s, P_g)$ for each CAV, representing the decision to Stop or Go. The critic network estimates the value of each state, thereby guiding the policy updates with a learned baseline. To reduce variance in policy gradient estimation, we adopt Generalized Advantage Estimation (GAE).

The clipped surrogate objective for the actor is defined as:

$$L^{CLIP}(\theta) = \mathbb{E}_t[\min(r_t(\theta)\hat{A}_t, clip(r_t(\theta), 1-\epsilon, 1+\epsilon)\hat{A}_t)], \tag{16}$$

where $r_t(\theta) = \frac{\pi_\theta(a_t|s_t)}{\pi_{\theta_{old}}(a_t|s_t)}$ measures the ratio of new to old policy probabilities, $\hat{A}_t$ is the estimated advantage at timestep $t$. The critic network is trained by minimizing the mean squared error:

$$L_{critic}(\phi) = \mathbb{E}_t[(r_t + \gamma V_\phi(s_{t+1}) - V_\phi(s_t))^2], \tag{17}$$

where $r_t$ is the immediate reward and $\gamma$ is the discount factor with a range of $[0, 1]$.

## 5. Experiment and results

### 5.1. Mixed traffic

Training and evaluation of RL algorithms require a suitable simulation environment. We adopted SUMO, an open-source, microscopic traffic simulator that not only incorporates well-established human driving models but also allows for flexible configuration of traffic networks and flows. Moreover, it ensures compliance with traffic rules, safety requirements, and physical constraints, thereby providing a reliable foundation for controlled experiments and reproducible evaluations. The built-in collision avoidance mechanism and human driving model will act as the downstream modules of self-driving software and the human driver, respectively. These mechanisms and models will ensure collision-free driving of a vehicle. This assumption has been widely adopted by previous studies.

To evaluate both the effectiveness and generality of the proposed framework, experiments were conducted on two representative intersection scenarios. The key geometric and traffic configurations are summarized in Table 2. Intersection I is a standard four-leg unsignalized intersection evaluated under three demand levels (600/900/1200 veh/h per entrance). Intersection II is a T-shaped unsignalized intersection with a fixed demand of 900 veh/h, which provides a stable operating condition for assessing adaptability to a non-standard geometry. Unless otherwise specified, both scenarios share the same lane-length settings and CAV market penetration configurations, ranging from 50% to 100%, are evaluated, with vehicle types (CAV or HDV) randomly assigned.

The type of vehicle (CAV or HDV) for each vehicle was randomly assigned. For HDV, the acceleration is computed using IDM. For CAV, when it is outside the control zone, IDM is used to determine the acceleration; when it is inside the control zone, the decisions are determined by the policy, which is computed by RL model. The overall simulation logic is shown in Fig. 4.

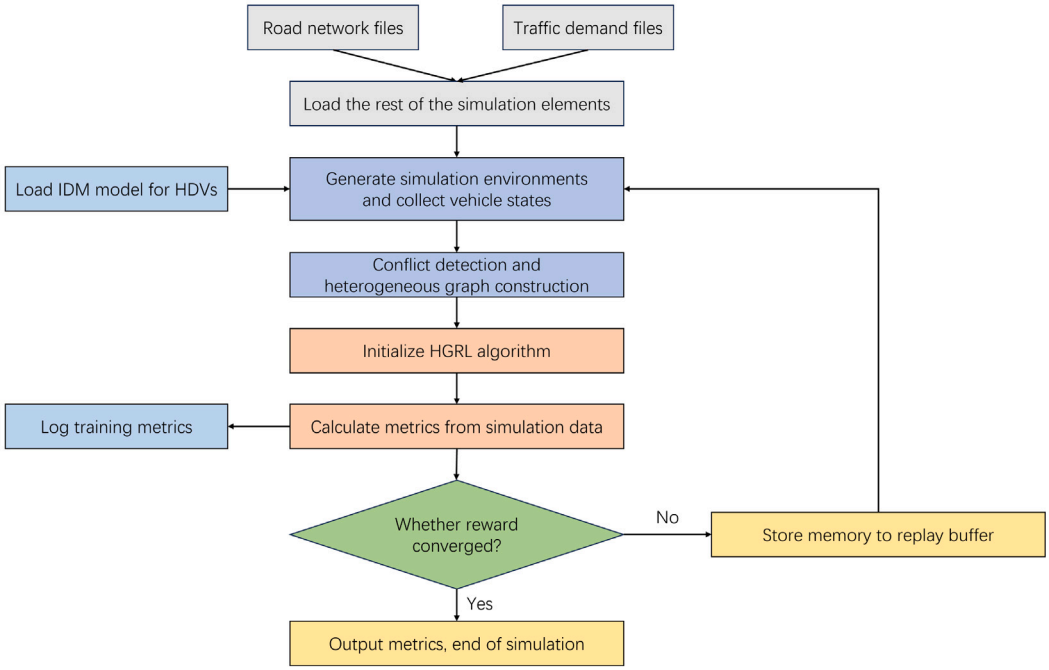


**Fig. 4.** Diagram of the overall simulation logic.

### 5.2. Experiment set-up

Our evaluation metrics contain average waiting time, average travel time, and average fuel consumption of all vehicles, which comprehensively reflect both the operational efficiency of the intersection and the environmental impact of traffic control strategies. These indicators are widely adopted in the literature as they enable a balanced assessment of mobility performance and sustainability outcomes.

We evaluate our method by comparing it to four baselines: (1) TL: the default traffic signal program deployed in the intersection; (2) NoTL: no traffic lights; (3) Wang [51]: the state-of-the-art RL traffic controller with 50% ~100% CAV penetration rate; and (4) Yan [53]: the state-of-the-art RL traffic controller with 100% CAV penetration rate for unsignalized intersections. For the traffic signal baseline (TL), a fixed-time control scheme is adopted for both intersection scenarios to provide a stable and interpretable reference.

For Intersection I, a balanced four-phase signal plan is employed, where each phase serves non-conflicting traffic movements from one approach. For Intersection II, a three-phase signal plan is designed to match the T-shaped geometry, ensuring that conflicting movements are fully separated across phases. In both cases, a green interval of 30 s and a yellow clearance interval of 3 s are applied to each phase. The clearance interval allows vehicles that have entered the intersection to safely clear the conflict area before phase switching, thereby preventing conflicts during transitions.

The selection of a fixed-time strategy and uniform timing parameters follows common traffic engineering practice and SUMO default recommendations. Importantly, these parameters are intentionally chosen to be conservative and non-optimized, so that the traffic signal baseline represents a commonly used real-world configuration rather than an optimized upper bound. This design avoids biasing the comparison in favor of the signal baseline and allows performance differences to more directly reflect the effectiveness of conflict-aware decision-making.

### 5.3. Overall performance

To establish a benchmark for subsequent evaluations under varying CAV penetration rates, we first examine the overall performance of all methods in a fully cooperative setting, where the CAV penetration rate is 100%. This scenario represents the upper bound of achievable performance without interference from human-driven vehicles and thus provides a consistent reference for comparison across different intersection configurations.

The results in Table 3 show that the proposed approach consistently achieves the best overall performance across all traffic demand levels. Compared with TL and NoTL, our method significantly reduces travel time and waiting time, particularly under high-demand conditions, indicating more efficient conflict resolution and smoother traffic progression. Even when compared with learning-based baselines, the proposed framework demonstrates clear advantages in both efficiency and delay reduction, while also

**Table 3**
Comparison of models under averaged metrics in Intersection I (Traffic flow from 600 to 1200).

| CAV rate | Model | Travel time (s) | | | Waiting time (s) | | | Fuel consumption (mL) | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | 600 | 900 | 1200 | 600 | 900 | 1200 | 600 | 900 | 1200 |
| 0.5 | Wang | 72.54 | 89.68 | 106.76 | 3.92 | 48.81 | 40.89 | 772.83 | 795.91 | 789.93 |
| | Ours | 72.01 | 88.28 | 96.31 | 5.24 | 39.30 | 34.84 | 777.17 | 788.40 | 795.14 |
| 0.6 | Wang | 71.98 | 89.85 | 96.37 | 3.17 | 24.47 | 33.18 | 769.67 | 795.11 | 780.02 |
| | Ours | 64.26 | 84.78 | 93.71 | 3.59 | 27.05 | 31.98 | 747.35 | 733.13 | 787.82 |
| 0.7 | Wang | 67.39 | 83.39 | 84.66 | 2.86 | 15.85 | 10.82 | 769.23 | 788.73 | 793.18 |
| | Ours | 48.84 | 75.96 | 88.17 | 0.87 | 11.97 | 15.76 | 724.80 | 769.67 | 786.72 |
| 0.8 | Wang | 68.35 | 85.20 | 96.51 | 3.04 | 12.73 | 13.43 | 768.57 | 787.76 | 788.71 |
| | Ours | 48.85 | 73.47 | 82.23 | 0.88 | 9.93 | 15.16 | 730.32 | 768.43 | 784.67 |
| 0.9 | Wang | 57.91 | 84.29 | 92.22 | 2.50 | 12.91 | 12.08 | 765.03 | 785.48 | 792.25 |
| | Ours | 48.86 | 72.11 | 82.30 | 0.88 | 8.85 | 13.78 | 730.34 | 768.93 | 787.21 |
| 1.0 | TL | 141.46 | 148.00 | 150.07 | 14.54 | 15.61 | 15.59 | 1172.50 | 1131.22 | 1070.02 |
| | NoTL | 68.33 | 171.63 | 129.82 | 5.42 | 27.85 | 52.23 | 1644.80 | 1041.19 | 1033.2 |
| | Yan | 72.46 | 86.44 | 93.63 | 2.57 | 9.29 | 23.45 | 1038.47 | 1068.52 | 1083.92 |
| | Wang | 53.45 | 70.28 | 88.82 | 1.98 | 6.22 | 10.57 | 758.46 | 788.07 | 776.71 |
| | Ours | 48.68 | 61.82 | 75.23 | 0.87 | 4.41 | 8.16 | 732.76 | 745.44 | 774.34 |

**Table 4**
Comparison of models under averaged metrics in Intersection II.

| CAV rate | Model | Travel time (s) | Waiting time (s) | Fuel consumption (mL) |
|---|---|---|---|---|
| 0.5 | Wang | 51.12 | 1.32 | 385.77 |
| | Ours | 46.02 | 0.93 | 402.76 |
| 0.6 | Wang | 47.61 | 1.01 | 368.49 |
| | Ours | 45.97 | 0.92 | 394.56 |
| 0.7 | Wang | 53.76 | 1.54 | 386.91 |
| | Ours | 46.28 | 0.95 | 397.06 |
| 0.8 | Wang | 49.92 | 1.22 | 395.30 |
| | Ours | 45.84 | 0.91 | 399.28 |
| 0.9 | Wang | 52.34 | 1.43 | 390.16 |
| | Ours | 45.75 | 0.89 | 389.98 |
| 1.0 | TL | 107.27 | 4.01 | 957.18 |
| | NoTL | 90.37 | 14.43 | 920.39 |
| | Yan | 53.64 | 1.86 | 400.19 |
| | Wang | 48.55 | 1.06 | 395.74 |
| | Ours | 44.42 | 0.83 | 384.70 |

maintaining lower or comparable fuel consumption. These results confirm that the proposed method scales well with increasing traffic demand and remains effective under congested conditions.

To further evaluate the generality of the proposed framework with respect to intersection geometry, Table 4 summarizes the performance comparison in Intersection II, a T-shaped unsignalized intersection, under a fixed traffic demand of 900 veh/h. Despite the asymmetric traffic flows and heterogeneous conflict patterns introduced by this geometry, the proposed method again outperforms all baseline approaches across the evaluated metrics. In particular, it achieves consistently lower travel time and waiting time across different CAV penetration rates, while maintaining competitive fuel consumption.

Taken together, the results from Tables 3 and 4 demonstrate that the proposed conflict-aware HGRL framework delivers robust and superior performance under both varying traffic demands and different intersection layouts. These findings indicate that the proposed method not only improves traffic efficiency and reduces delays but also generalizes well across diverse operating conditions, providing a strong performance baseline for subsequent mixed-traffic evaluations.

To gain deeper insight into the operational characteristics of different methods, we next focus on Intersection I under a representative traffic demand of 900 veh/h, and provide a detailed analysis of travel time, waiting time, and fuel consumption.

#### *5.3.1. Waiting time*

Waiting time measures the average duration that vehicles remain idle before passing through the intersection and serves as a direct reflection of queuing and congestion. Fig. 5 reports the waiting time distributions of different methods under varying penetration rates.

The results show that waiting time decreases as the penetration rate increases and gradually converges to a stable low level. The proposed method achieves reductions ranging from 19.5% at low penetration to over 31.4% at high penetration (0.9–1.0). This trend can be explained by the decreasing proportion of HDVs, which enables cooperative CAVs to coordinate their decisions more

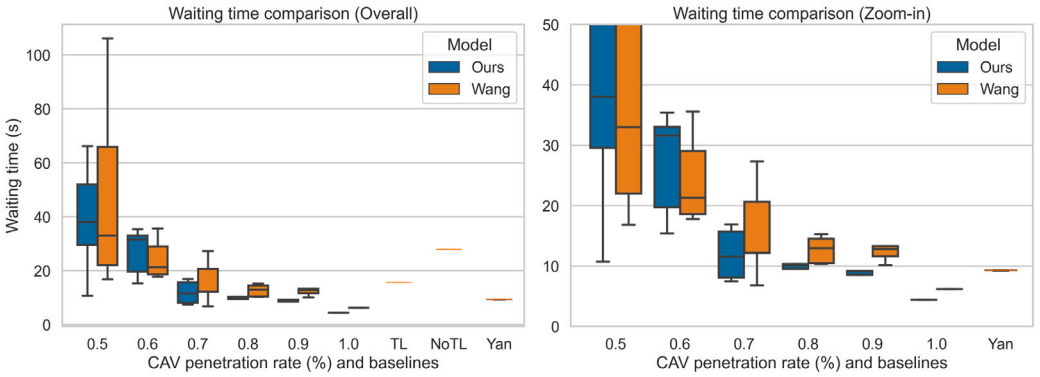


**Fig. 5.** The overall results in average fuel consumption at the intersection. The RIGHT sub-figure displays a zoomed-in version of the LEFT sub-figure, excluding the Yan methods. Our method consistently outperforms the other four baselines.

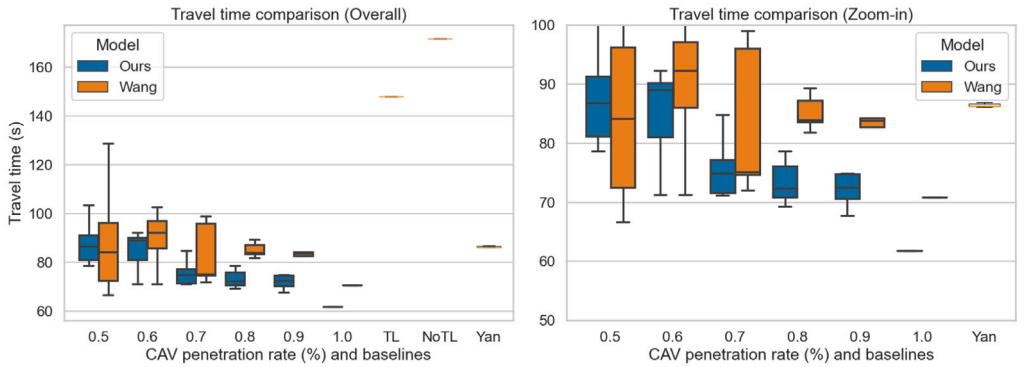


**Fig. 6.** The overall results in average travel time at the intersection. The RIGHT sub-figure displays a zoomed-in version of the LEFT sub-figure, excluding the Yan methods. When the RV penetration rate reaches or exceeds 70%, our method consistently outperforms the other four baselines.

effectively, thereby reducing hesitation and mitigating queue spillback. Our method further amplifies this effect through its graph-based modeling framework, which explicitly captures inter-vehicle interactions and conflict relations. This design allows vehicles to negotiate shared space more efficiently, minimizing idle periods. Consequently, our approach delivers not only shorter but also more stable waiting times, highlighting its robustness under mixed-traffic conditions.

### *5.3.2. Travel time*

Travel time is a fundamental indicator of traffic efficiency. Compared with delay, another widely used metric, travel time also reflects the overall mobility of vehicles at intersections. Therefore, we adopt travel time as the primary measure of efficiency. The distributions of travel time for the different models are shown in Fig. 6.

As illustrated, our method achieves consistently lower travel times across most penetration rates compared with Wang and other baselines. At penetration levels of 0.5 and 0.6, the relative advantage over Wang diminishes, with Wang occasionally achieving marginally shorter travel times. Nevertheless, the variance of our results remains significantly smaller, indicating that the proposed framework provides more stable system performance in mixed-traffic environments dominated by HDV-induced disturbances.

At medium to high penetration rates (0.7–1.0), our approach demonstrates clear superiority, yielding both lower mean travel times and reduced variability. This suggests that the explicit modeling of cooperative and conflict relations enables more efficient coordination among vehicles and mitigates congestion at conflict points. When compared with traditional baselines such as TL, NoTL, and Yan, our method consistently achieves the lowest travel times, underscoring its robustness and scalability across a wide spectrum of traffic compositions.

### *5.3.3. Fuel consumption*

Fuel consumption is an important indicator of environmental sustainability, as it reflects both energy efficiency and emissions reduction. Fig. 7 illustrates the distribution of fuel consumption across different penetration rates for our method and the Wang baseline, with additional comparisons to TL, NoTL, and Yan.

Overall, our method achieves consistently lower fuel consumption, with reductions ranging from 0.9% at low penetration to 5.4% under full penetration compared with Wang. These improvements are less pronounced than those observed for travel time,

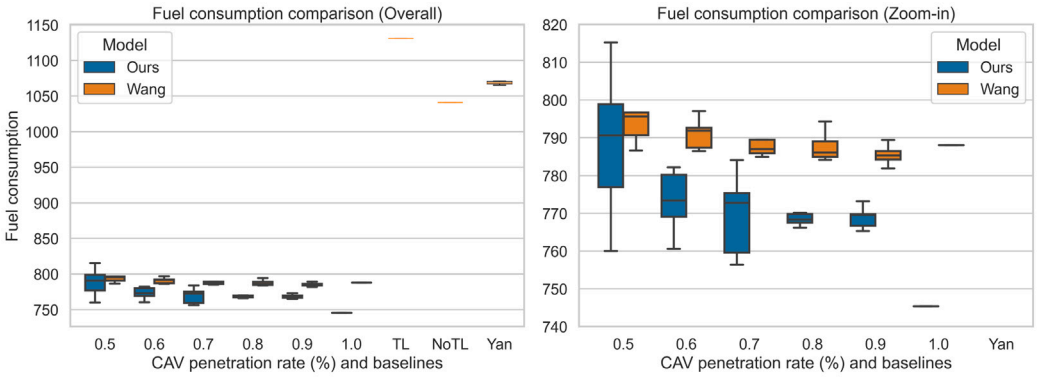


**Fig. 7.** The overall results in average fuel consumption at the intersection. The RIGHT sub-figure displays a zoomed-in version of the LEFT sub-figure, excluding the Yan methods. Our method consistently outperforms the other four baselines.

**Table 5**
Ablation settings of different model variants.

| Model | Graph structure | Graph encoder | Sparse attention (SAT) |
|---|---|---|---|
| DRL | None | – | – |
| GNN | Homogeneous | GCN | – |
| GRL | Heterogeneous | GCN | – |
| GRL_A | Heterogeneous | H-GAT | – |
| GRL_SAT | Heterogeneous | GCN | ✓ |
| Ours | Heterogeneous | H-GAT | ✓ |

primarily because fuel consumption is influenced not only by cooperation but also by idling and acceleration patterns, which remain affected by HDVs in mixed traffic.

The relatively modest gains highlight that while higher penetration improves coordination and reduces stop-and-go behavior, inefficiencies caused by HDV variability cannot be completely eliminated. Nonetheless, our model delivers noticeable advantages due to the jerk penalty in the reward function, which discourages abrupt acceleration and deceleration. By promoting smoother driving trajectories and reducing idling in queues, our method successfully lowers unnecessary fuel use, achieving superior performance across all penetration rates compared with Wang and other baselines.

### *5.4. Ablation study*

To evaluate the contribution of each component in our framework, we design a comprehensive set of ablation experiments, as summarized in Table 5. The ablated models differ primarily in three aspects: the use of graph structure, the choice of graph encoder, and the inclusion of the conflict-aware sparse attention mechanism. Each variant can thus be regarded as a different version of the proposed HGNN model illustrated in Fig. 3.

Several points about the adopted models should be emphasized as follows:

- The DRL model and GNN model are selected as the baseline models. The DRL model does not employ any graph structure; instead, it aggregates raw observations from the environment as input for decision-making. The GNN model, in contrast, replaces the heterogeneous graph with a homogeneous one, without explicitly modeling conflict relations.
- The GRL model replaces the graph neural network in our model with a standard GCN, to examine the effectiveness of the H-GAT encoder under the heterogeneous graph setting.
- If the model does not incorporate the sparse attention mechanism, which highlights conflict edges, it is replaced with a fully connected layer.

#### *5.4.1. Analysis of results*

The training reward curve is plotted in Fig. 8. Moreover, the evaluation metrics are averaged as the final training results when the model converges, as shown in Table 6.

The DRL and GNN models serve as preliminary baselines to evaluate the effect of incorporating graph structure. Both models lag behind the heterogeneous variants in terms of convergence speed and final reward plateau. The DRL model, which directly aggregates raw observations without any graph representation, converges slowly and remains less stable throughout training. The GNN model, which employs a homogeneous graph, performs slightly better than DRL in terms of training dynamics; however, it still fails to capture heterogeneous relations, such as cooperation and conflict. As reflected in the evaluation result, although its travel time and fuel consumption are not markedly worse than those of GRL, it suffers from the highest waiting time among all

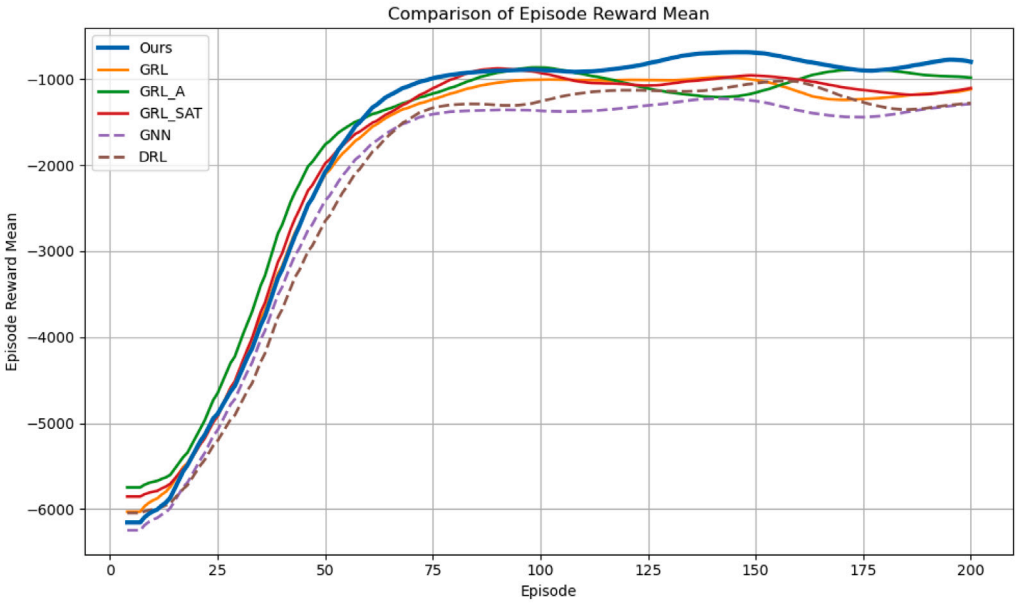


**Fig. 8.** The training reward curve of the implementation models in the ablation study.

**Table 6**
The evaluation result of ablation study.

| Metrics (Avg.) | Models | | | | | |
|---|---|---|---|---|---|---|
| | DRL | GNN | GRL | GRL_A | GRL_SAT | **Ours** |
| Travel time | 69.26 | 65.19 | 67.57 | 67.92 | 62.38 | **61.82** |
| Fuel consumption | 787.97 | 788.86 | 791.94 | 790.57 | 779.47 | **745.44** |
| Waiting time | 7.94 | 8.79 | 5.64 | 6.98 | 5.79 | **4.41** |

models. These results indicate that the explicit construction of heterogeneous graphs is essential for effectively modeling interaction dynamics in mixed-autonomy traffic.

The GRL models exhibits notable deficiencies in stability and efficiency. Although the mean episode reward increases progressively with training, the curve remains characterized by substantial oscillations and fails to converge within a narrow interval. This pattern suggests that a conventional graph convolutional encoder lacks the representational capacity required to adequately capture the intricate interaction structures present in mixed-autonomy traffic, thereby constraining both robustness and decision reliability.

Substituting the GNN with a graph attention network (GRL_A) yields a steeper rise of the reward curve during the initial training phase and accelerates convergence. This indicates that attention mechanisms over nodes and their incident edges enable the policy to prioritize salient local interactions and exploit structural cues more effectively. However, because the attention weights are distributed densely across both critical and non-critical neighbors, the influence of genuinely decisive relations is diluted. As a consequence, fluctuations re-emerge in the mid-to-late stages of training, and the efficiency improvements do not consistently extend to safety or stability.

In contrast, the GRL_SAT variant introduces a sparse attention mechanism restricted exclusively to conflict edges. By concentrating model capacity on those interactions that determine right-of-way and collision risk, this approach attenuates the influence of irrelevant neighbors and enhances discriminative power. Empirically, the corresponding reward curve exhibits smoother dynamics and convergence to a narrower plateau. The model tends to adopt more conservative actions at potential conflict points, thereby sacrificing a degree of instantaneous aggressiveness in exchange for greater reliability and stability of traffic flow.

Finally, the complete model (Ours) integrates the attentive graph encoder with conflict-aware sparse attention in a unified framework. The reward curve rises rapidly and stabilizes at a higher plateau than all ablated configurations, reflecting both faster learning dynamics and improved convergence properties. Consistent with these observations, the model achieves the most favorable trade-off across travel time, fuel consumption, and waiting time. Collectively, these findings demonstrate that while attention is essential for surpassing the limitations of basic GNNs, it is the conflict-aware sparsification of attention that translates representational gains into robust, safe, and efficient decision-making. The synergy of GAT and conflict-edge SAT thus constitutes the critical factor enabling the proposed framework to deliver state-of-the-art performance in mixed-traffic environments.

## 6. Discussion

This paper has introduced a conflict-aware Heterogeneous Graph Reinforcement Learning (HGRL) framework for managing mixed traffic flows of CAVs and HDVs. By explicitly modeling both cooperative and conflict interactions, our approach captures complex heterogeneous dynamics that are often overlooked by conventional graph-based or rule-based methods. The integration of a sparse conflict-aware attention mechanism enables the model to efficiently pinpoint safety-critical patterns. Furthermore, embedding this HGNN representation within a reinforcement learning policy network facilitates adaptive decision-making across varying CAV penetration rates.

Evaluation results demonstrate that our method significantly enhances systemic utility. By adopting an individual CAV perspective, the algorithm ensures not only local operational efficiency but also global intersection performance. This reveals that with access to global information, agents can achieve better coordination with both CAVs and HDVs, leading to more informed decisions before entering conflict zones and a reduction in indecisive or inefficient behaviors.

The promising results presented above underscore the value of the proposed conflict-aware HGRL framework. To deepen the understanding of its contributions and pathways to application, the following discussion is structured around four key themes: the design rationale and efficacy of the conflict-aware mechanism (6.1), the framework's generality and scalability (6.2), a critical reflection on our modeling assumptions and limitations (6.3), and the resulting practical implications for real-world deployment (6.4).

### 6.1. Conflict-aware control mechanism and design rationale

In the proposed framework, the conflict resolution mechanism is applied strictly at the execution level as a safety-oriented conflict resolver. Specifically, it post-processes simultaneous "Go" decisions from conflicting traffic streams to ensure safe and deadlock-free right-of-way assignment. Importantly, this mechanism does not define or alter the optimization objective of the reinforcement learning policy.

The policy itself is driven by a composite reward function that jointly considers waiting time, speed, comfort and eco-driving behavior (e.g., jerk), as well as conflict-related penalties. This reward formulation encourages coordinated and efficient system-level behavior under mixed traffic conditions. The priority rule is activated only in a limited subset of boundary cases where multiple conflicting vehicles concurrently select the "Go" action. In such cases, a fairness-oriented tie-breaking strategy helps prevent starvation and stabilizes traffic progression without biasing the underlying learning objective.

From an engineering perspective, this separation between reward-driven optimization and execution-level safety enforcement reflects a practical trade-off between efficiency and robustness. While conservative interventions may occasionally introduce minor local delays, they play a crucial role in maintaining stable and safe operation under high-conflict scenarios.

From the perspective of conflict modeling, in the current implementation, conflict relations are predefined. This design choice simplifies conflict modeling and allows systematic evaluation of the proposed framework. However, we acknowledge that such predefined conflict tables may limit scalability when intersection geometries become highly irregular or when extending the framework to large-scale networks. Importantly, the proposed framework is not inherently restricted to manually specified conflict tables. Conflict relations fundamentally arise from spatial–temporal interactions and potential trajectory intersections between vehicles. Under new geometric configurations, conflict pairs can be automatically identified based on lane connectivity, vehicle trajectories, and predicted path overlaps, without relying on predefined intersection templates.

While automatic conflict relation generation is not implemented in the current study, it represents a natural and practical extension to enhance generalizability and reduce manual intervention. Exploring robust and efficient mechanisms for automatic conflict identification under arbitrary intersection geometries remains an important direction for future engineering-oriented development.

### 6.2. Generality and scalability of the framework

A key strength of our framework is its adaptability to diverse intersection layouts, including non-orthogonal and signalized intersections. This generality stems from the model's reliance on real-time dynamic states — such as vehicle positions, queue lengths, and the conflict relations derived therefrom — rather than on a fixed intersection topology. Consequently, irregular geometries do not pose a limitation, as conflict identification remains consistent given basic positional data. This generality is empirically supported by the additional evaluation conducted on a T-shaped unsignalized intersection. Compared with the standard four-leg configuration, the T-shaped intersection introduces asymmetric traffic flows and heterogeneous conflict patterns. Despite this structural difference, the proposed framework maintains consistent performance gains, indicating that the conflict-aware graph representation and decision mechanism are not tied to a specific intersection topology but generalize effectively across different geometric layouts.

This principle also underpins the method's strong scalability to multi-intersection networks. The model's inputs — vehicle states and conflict definitions — are inherently independent of a single intersection's physical boundaries. The agent-based design ensures that decisions remain local, while the consistency of decision rules and the ability to incorporate broader global information (e.g., adjacent intersections' states) enable system-wide coordination. This allows the framework to naturally account for network-level phenomena, such as upstream–downstream coupling and congestion spillback.

### 6.3. Modeling assumptions and practical limitations

The proposed framework is developed and evaluated under several simplifying assumptions, including full state observability, accurate position information, and deterministic human-driven vehicle (HDV) behavior modeled using the Intelligent Driver Model (IDM). These assumptions represent an idealized setting that allows us to isolate and evaluate the effectiveness of the proposed conflict-aware graph representation and control strategy without introducing confounding factors at the sensing or communication layers.

In practical deployments, sensing noise, communication delays, and partial observability may affect the timeliness and accuracy of conflict detection, as widely discussed in studies on V2X communication reliability and security [54–57]. Such imperfections primarily influence the state construction and conflict identification stages rather than the policy optimization process itself. As discussed in Section 6.1, the conflict resolution mechanism operates strictly at the execution level and does not modify the reinforcement learning objective. Since the proposed framework relies on aggregated, conflict-relevant interaction features rather than exact instantaneous states, moderate delays or information loss are expected to result in more conservative decision timing rather than fundamentally altering learned coordination behaviors. Moreover, the use of sparse attention restricts information propagation to safety-critical interactions, which helps limit the impact of noisy or delayed information to localized regions of the interaction graph.

Regarding human driving behavior, the IDM provides a widely adopted baseline model but does not capture the full variability of real-world human driving, such as differences in reaction times, aggressiveness, or compliance. Introducing richer behavioral variability may increase uncertainty in interaction patterns and conflict emergence. However, the proposed conflict-aware design focuses on relative spatial and temporal relationships among vehicles, which are less sensitive to specific car-following parameters as long as interaction dynamics remain observable. As a result, the framework is expected to maintain stable performance trends under heterogeneous driving behaviors, albeit with potential degradation in efficiency under highly aggressive or unpredictable conditions.

### 6.4. Practical implications

Building upon the demonstrated generality and scalability, the proposed conflict-aware HGRL framework is well suited for real-world traffic management scenarios involving mixed traffic with partial CAV penetration. By explicitly modeling conflict and cooperative interactions, the framework can support more informed decision-making at intersections where human-driven and automated vehicles coexist.

The reliance on dynamic vehicle states and conflict relationships reduces the need for intersection-specific rule design or manual reconfiguration when traffic layouts or demand patterns change. From an operational perspective, this property simplifies system deployment and maintenance, particularly in urban environments where intersection geometries and traffic conditions are highly heterogeneous. The agent-based formulation further supports network-level operation by enabling local decision-making while accommodating broader traffic interactions, such as congestion spillback and upstream–downstream coupling.

In addition, the proposed approach is compatible with existing sensing and V2X infrastructures, as it primarily requires vehicle-level state information that is increasingly available in modern traffic systems. This compatibility suggests that the framework can be incrementally deployed, for example by prioritizing CAVs as decision-making agents while coexisting with legacy traffic control mechanisms, thereby providing a feasible pathway toward real-world adoption.

## 7. Conclusion

Mixed traffic intersections present considerable challenges for safe and efficient operation, as CAVs must coordinate not only with each other but also with HDVs under complex and dynamic conditions. To address this, we have introduced a conflict-aware Heterogeneous Graph Reinforcement Learning (HGRL) framework that explicitly models vehicle interactions to enable cooperative and efficient decision-making.

Our main contributions are threefold. First, we introduced a conflict-aware heterogeneous graph representation for mixed-traffic intersections, which explicitly models both cooperative and conflicting interactions among heterogeneous vehicles. This representation provides a unified modeling framework for analyzing the complex dynamics between CAVs and HDVs. Second, we developed an HGRL framework that integrates a heterogeneous graph neural network with a deep reinforcement learning policy, facilitating safe and efficient multi-agent coordination. This design ensures that individual CAVs enhance their own performance while simultaneously improving system-wide efficiency and safety. Third, we conducted comprehensive simulations across varying CAV penetration rates, demonstrating the superiority of our approach over established baselines. Ablation studies further validated the necessity of each component within the proposed framework.

Despite the promising results, several limitations warrant further investigation. First, HDVs are modeled using IDM, which captures basic car-following behavior but does not fully represent the diversity and uncertainty of real human driving. Second, the experiments assume idealized sensing and communication conditions (e.g., accurate positions and strong observability), whereas real-world deployments may face delays, packet loss, and partial observability that affect state construction and conflict detection. Third, predefined conflict relations simplify controlled evaluation but may limit scalability when extending to highly irregular geometries and large road networks, where automated conflict identification would be desirable.

Addressing these points provides a clear pathway for future work. In particular, our future efforts will primarily focus on: (1) Behavioral realism. We will integrate real-world trajectory data and richer HDV behavior models to better capture human variability and improve the fidelity of mixed-traffic interactions. (2) Robustness under non-ideal information. We will evaluate robustness under sensing noise, delayed or partial V2X information by introducing communication delay and information loss into the simulation, and refine the state construction and reward design accordingly. (3) Generalizability and scalability. We will develop automated conflict relation generation based on lane connectivity and predicted path overlaps to reduce manual configuration across diverse intersection geometries, and conduct large-scale network simulations to assess system-level performance and computational demands under realistic traffic conditions. In addition, we will explore practical extensions — including vehicle heterogeneity and priority requirements, more fine-grained (continuous or hybrid) action spaces, and energy evaluation for electric vehicle scenarios — to further enhance engineering applicability.

## Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## Acknowledgments

This work is supported by grants from the National Key R&D Program of China (2023YFB4301900) and the Science and Technology Planning Project of Guangdong Province, China (2023B1212060029). The authors declare no competing financial interests in this paper.

## Data availability

The authors do not have permission to share data.

## References

[1] B. Pishue, 2021 inrix global traffic scorecard, Vol. 1, INRIX, Washington, 2021, USAM Scorecard Rep.
[2] J. Li, C. Yu, Z. Shen, Z. Su, W. Ma, A survey on urban traffic control under mixed traffic environment with connected automated vehicles, Transp. Res. Part C: Emerg. Technol. 154 (2023) 104258.
[3] X. Di, R. Shi, A survey on autonomous vehicle control in the era of mixed-autonomy: From physics-based to AI-guided driving policy learning, Transp. Res. Part C: Emerg. Technol. 125 (2021) 103008.
[4] National Highway Traffic Safety Administration, Vehicle-to-Vehicle Communication Technology for Light Vehicles, Tech. Rep. FMVSS 150, US Dep. Transp., 2016.
[5] The State Council, National comprehensive three-dimensional transportation network planning outline, 2021.
[6] X. Huang, P. Lin, M. Pei, B. Ran, M. Tan, Reservation-based cooperative ecodriving model for mixed autonomous and manual vehicles at intersections, IEEE Trans. Intell. Transp. Syst. 24 (9) (2023) 9501–9517.
[7] X. Pan, B. Chen, S. Timotheou, S.A. Evangelou, A convex optimal control framework for autonomous vehicle intersection crossing, IEEE Trans. Intell. Transp. Syst. 24 (1) (2022) 163–177.
[8] J. Zhang, S. Li, L. Li, Coordinating CAV swarms at intersections with a deep learning model, IEEE Trans. Intell. Transp. Syst. 24 (6) (2023) 6280–6291.
[9] D. Zhou, P. Hang, J. Sun, Reasoning graph-based reinforcement learning to cooperate mixed connected and autonomous traffic at unsignalized intersections, Transp. Res. Part C: Emerg. Technol. 167 (2024) 104807.
[10] M. Behrisch, L. Bieker, J. Erdmann, D. Krajzewicz, SUMO–simulation of urban mobility: an overview, in: Proceedings of SIMUL 2011, the Third International Conference on Advances in System Simulation, ThinkMind, 2011.
[11] A. Mustapha, A.M. Abdul-Rani, N. Saad, M. Mustapha, Advancements in traffic simulation for enhanced road safety: A review, Simul. Model. Pract. Theory 137 (2024) 103017.
[12] N. Zainuddin, S. Shah, M. Hashim, M. Roslam, L. Tey, Comparison of operational performance before and after improvement: Case study at pengkalan weld, pulau pinang, in: AIP Conference Proceedings, Vol. 2020, AIP Publishing LLC, 2018, 020027.
[13] E. Namazi, J. Li, C. Lu, Intelligent intersection management systems considering autonomous vehicles: A systematic literature review, IEEE Access 7 (2019) 91946–91965.
[14] J. Yu, L. Wang, X. Gong, Study on the status evaluation of urban road intersections traffic congestion base on AHP-TOPSIS modal, Procedia Soc. Behav. Sci. 96 (2013) 609–616.
[15] Q. Yu, W. Li, D. Yang, H. Zhang, Partitioning urban road network based on travel speed correlation, Int. J. Transp. Sci. Technol. 10 (2) (2021) 97–109.
[16] I. Jayaweera, K. Perera, J. Munasinghe, Centrality measures to identify traffic congestion on road networks: A case study of Sri Lanka, IOSR J. Math. 13 (02) (2017) 13–19.
[17] D. Li, J. Lasenby, Spatiotemporal attention-based graph convolution network for segment-level traffic prediction, IEEE Trans. Intell. Transp. Syst. 23 (7) (2021) 8337–8345.
[18] H. Wei, N. Xu, H. Zhang, G. Zheng, X. Zang, C. Chen, W. Zhang, Y. Zhu, K. Xu, Z. Li, Colight: Learning network-level cooperation for traffic signal control, in: Proceedings of the 28th ACM International Conference on Information and Knowledge Management, 2019, pp. 1913–1922.
[19] S. Porta, P. Crucitti, V. Latora, The network analysis of urban streets: a primal approach, Environ. Plan. B: Plan. Des. 33 (5) (2006) 705–725.
[20] Y. Zhang, X. Wang, P. Zeng, X. Chen, Centrality characteristics of road network patterns of traffic analysis zones, Transp. Res. Rec. 2256 (1) (2011) 16–24.
[21] W. Zhang, S. Wang, X. Tian, D. Yu, Z. Yang, The backbone of urban street networks: Degree distribution and connectivity characteristics, Adv. Mech. Eng. 9 (11) (2017) 1687814017742570.
[22] Y. Yue, A.G.-O. Yeh, Spatiotemporal traffic-flow dependency and short-term traffic forecasting, Environ. Plan. B: Plan. Des. 35 (5) (2008) 762–771.
[23] D. Zhang, J. Yin, X. Zhu, C. Zhang, Network representation learning: A survey, IEEE Trans. Big Data 6 (1) (2018) 3–28.
[24] W.L. Hamilton, R. Ying, J. Leskovec, Representation learning on graphs: Methods and applications, 2017, arXiv preprint arXiv:1709.05584.

[25] B. Perozzi, R. Al-Rfou, S. Skiena, Deepwalk: Online learning of social representations, in: Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2014, pp. 701–710.
[26] A. Grover, J. Leskovec, Node2vec: Scalable feature learning for networks, in: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2016, pp. 855–864.
[27] L.F. Ribeiro, P.H. Saverese, D.R. Figueiredo, Struc2vec: Learning node representations from structural identity, in: Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2017, pp. 385–394.
[28] A. Pareja, G. Domeniconi, J. Chen, T. Ma, T. Suzumura, H. Kanezashi, T. Kaler, T. Schardl, C. Leiserson, Evolvegcn: Evolving graph convolutional networks for dynamic graphs, in: Proceedings of the AAAI Conference on Artificial Intelligence, Vol. 34, 2020, pp. 5363–5370.
[29] A. Sankar, Y. Wu, L. Gou, W. Zhang, H. Yang, Dysat: Deep neural representation learning on dynamic graphs via self-attention networks, in: Proceedings of the 13th International Conference on Web Search and Data Mining, 2020, pp. 519–527.
[30] S. Abidi, P. Mathieu, A. Nongaillard, Analyzing communication policies in cooperative multi-agent reinforcement learning for traffic signal control: A simulation-based study, Simul. Model. Pract. Theory 141 (2025) 103100.
[31] S. Cheng, S. Qu, J. Zhang, Transfer-mamba: Selective state space models with spatio-temporal knowledge transfer for few-shot traffic prediction across cities, Simul. Model. Pract. Theory 140 (2025) 103066.
[32] M. Klimke, B. Völz, M. Buchholz, Cooperative behavior planning for automated driving using graph neural networks, in: 2022 IEEE Intelligent Vehicles Symposium, IV, IEEE, 2022, pp. 167–174.
[33] Q. Liu, Y. Tang, X. Li, F. Yang, X. Gao, Z. Li, SIF-STGDAN: A social interaction force spatial-temporal graph dynamic attention network for decision-making of connected and autonomous vehicles, in: 2024 IEEE Intelligent Vehicles Symposium, IV, IEEE, 2024, pp. 376–383.
[34] C. Chen, Q. Xu, M. Cai, J. Wang, J. Wang, K. Li, Conflict-free cooperation method for connected and automated vehicles at unsignalized intersections: Graph-based modeling and optimality analysis, IEEE Trans. Intell. Transp. Syst. 23 (11) (2022) 21897–21914.
[35] B. Xu, X.J. Ban, Y. Bian, J. Wang, K. Li, V2I based cooperation between traffic signal and approaching automated vehicles, in: 2017 IEEE Intelligent Vehicles Symposium, IV, IEEE, 2017, pp. 1658–1664.
[36] J. Zhang, H. Li, Y. Ma, C. Zhang, L. Qin, N. Chen, Modeling and optimization of platooning behaviors in fixed-time signalized intersection entrance areas, Simul. Model. Pract. Theory 132 (2024) 102900.
[37] W. Xie, X. Peng, Y. Liu, J. Zeng, L. Li, T. Eisaka, Conflict-free coordination planning for multiple automated guided vehicles in an intelligent warehousing system, Simul. Model. Pract. Theory 134 (2024) 102945.
[38] K. Dresner, P. Stone, Multiagent traffic management: A reservation-based intersection control mechanism, in: Autonomous Agents and Multiagent Systems, International Joint Conference on, Vol. 3, IEEE Computer Society, 2004, pp. 530–537.
[39] R. Tachet, P. Santi, S. Sobolevsky, L.I. Reyes-Castro, E. Frazzoli, D. Helbing, C. Ratti, Revisiting street intersections using slot-based systems, PLoS One 11 (3) (2016) e0149607.
[40] C. Ren, L. Lu, X. Liu, F. Fu, L. Cheng, Multi-intersection platoon ecological speed planning strategy and method for autonomous driving simulation testing, Simul. Model. Pract. Theory (2025) 103166.
[41] Z. Yao, H. Jiang, Y. Jiang, B. Ran, A two-stage optimization method for schedule and trajectory of CAVs at an isolated autonomous intersection, IEEE Trans. Intell. Transp. Syst. 24 (3) (2023) 3263–3281.
[42] D. Li, F. Zhu, T. Chen, Y.D. Wong, C. Zhu, J. Wu, COOR-PLT: A hierarchical control model for coordinating adaptive platoons of connected and autonomous vehicles at signal-free intersections based on deep reinforcement learning, Transp. Res. Part C: Emerg. Technol. 146 (2023) 103933.
[43] R. Zhao, Y. Li, K. Wang, Y. Fan, F. Gao, Z. Gao, Centralized cooperation for connected autonomous vehicles at intersections by safe deep reinforcement learning, IEEE Trans. Mob. Comput. 23 (12) (2024) 12830–12847.
[44] S. Shen, Y. Fu, H. Su, H. Pan, P. Qiao, Y. Dou, C. Wang, Graphcomm: A graph neural network based method for multi-agent reinforcement learning, in: ICASSP 2021-2021 IEEE International Conference on Acoustics, Speech and Signal Processing, ICASSP, IEEE, 2021, pp. 3510–3514.
[45] S. Munikoti, D. Agarwal, L. Das, M. Halappanavar, B. Natarajan, Challenges and opportunities in deep reinforcement learning with graph neural networks: A comprehensive review of algorithms and applications, IEEE Trans. Neural Netw. Learn. Syst. 35 (11) (2023) 15051–15071.
[46] Y. Zhang, C.G. Cassandras, An impact study of integrating connected automated vehicles with conventional traffic, Annu. Rev. Control. 48 (2019) 347–356.
[47] P.-C. Chen, X. Liu, C.-W. Lin, C. Huang, Q. Zhu, Mixed-traffic intersection management utilizing connected and autonomous vehicles as traffic regulators, in: Proceedings of the 28th Asia and South Pacific Design Automation Conference, 2023, pp. 52–57.
[48] S. Jiang, T. Pan, R. Zhong, C. Chen, X.-a. Li, S. Wang, Coordination of mixed platoons and eco-driving strategy for a signal-free intersection, IEEE Trans. Intell. Transp. Syst. 24 (6) (2022) 6597–6613.
[49] H. Li, J. Zhang, Y. Li, Z. Huang, H. Cao, Modeling and simulation of vehicle group collaboration behaviors in an on-ramp area with a connected vehicle environment, Simul. Model. Pract. Theory 110 (2021) 102332.
[50] M. Treiber, A. Hennecke, D. Helbing, Congested traffic states in empirical observations and microscopic simulations, Phys. Rev. E 62 (2) (2000) 1805.
[51] D. Wang, W. Li, L. Zhu, J. Pan, Learning to control and coordinate mixed traffic through robot vehicles at complex and unsignalized intersections, Int. J. Robot. Res. 44 (5) (2025) 805–825.
[52] J. Schulman, F. Wolski, P. Dhariwal, A. Radford, O. Klimov, Proximal policy optimization algorithms, 2017, arXiv preprint arXiv:1707.06347.
[53] Z. Yan, C. Wu, Reinforcement learning for mixed autonomy intersections, in: 2021 IEEE International Intelligent Transportation Systems Conference, ITSC, IEEE, 2021, pp. 2089–2094.
[54] A.A. Almazroi, M.H. Alkinani, M.A. Al-Shareeda, S. Manickam, A novel ddos mitigation strategy in 5g-based vehicular networks using Chebyshev polynomials, Arab. J. Sci. Eng. 49 (9) (2024) 11991–12004.
[55] A.A. Almazroi, M.A. Alqarni, M.A. Al-Shareeda, M.H. Alkinani, A.A. Almazroey, T. Gaber, FCA-VBN: Fog computing-based authentication scheme for 5G-assisted vehicular blockchain network, Internet Things 25 (2024) 101096.
[56] M.A. Al-shareeda, M. Anbar, I.H. Hasbullah, S. Manickam, N. Abdullah, M.M. Hamdi, Review of prevention schemes for replay attack in vehicular ad hoc networks (vanets), in: Proceedings of the 2020 IEEE 3rd International Conference on Information Communication and Signal Processing, ICICSP, IEEE, 2020, pp. 394–398.
[57] A.A. Almazroi, E.A. Aldhahri, M.A. Al-Shareeda, S. Manickam, ECA-VFog: An efficient certificateless authentication scheme for 5G-assisted vehicular fog computing, PLoS One 18 (6) (2023) e0287291.