

Received 26 July 2023, accepted 24 August 2023, date of publication 4 September 2023, date of current version 8 September 2023.

Digital Object Identifier 10.1109/ACCESS.2023.3312021

## RESEARCH ARTICLE

# A Collaborative Control Scheme for Smart Vehicles Based on Multi-Agent Deep Reinforcement Learning

LIYAN SHI<sup>1</sup> AND HAIRUI CHEN<sup>2</sup>

<sup>1</sup>School of Information Engineering and Artificial Intelligence, The Open University of Henan, Zhengzhou 450046, China

<sup>2</sup>Zhongyuan-Petersburg Aviation College, Zhongyuan University of Technology, Zhengzhou 450007, China

Corresponding author: Liyan Shi (slyyy9966@163.com)

**ABSTRACT** With the development of artificial intelligence and autonomous driving technology, the vehicle-road cooperative control system combined with artificial intelligence technology can provide more effective and adaptive traffic control solutions for intelligent transportation systems. Existing research works are confronted with two kinds of challenges. For one thing, traditional recurrent neural networks-based methods cannot model the long-time dependent information in traffic flow sequences. For another, the large sample correlation makes it difficult to optimize the trained strategies. In this paper, we propose a Multi-agent Deep Reinforcement Learning (MADRL)-based intelligent vehicle cooperative control method to deal remedy current gaps. To this end, a closed-loop control system of self-driving vehicles and signal controllers is used as the research object to achieve dynamic scheduling of traffic flow by MADRL. After designing relevant experimental validation, the feasibility of the method is verified in terms of both scheme comparison and operational effect analysis, which is a good aid to traffic signal timing. The simulation results show that the proposal can be well utilized to realize collaborative control of smart vehicles, and there is some performance improvement compared with several typical methods.

**INDEX TERMS** Collaborative control, smart vehicles, deep reinforcement learning, intelligent transportation systems.

## I. INTRODUCTION


### A. IMPACT OF ROAD CONGESTION

Urban road congestion not only seriously affects people's travel efficiency but also is a hidden danger to traffic safety. This restricts urban development and causes incalculable losses to urban development [1]. The rapid growth of urban population and motor vehicle ownership has triggered the rapid growth of urban traffic demand. The contradiction between supply and demand of urban transportation systems is intensifying [2]. And the backward transportation systems have become the main bottleneck to restrict the sustainable development of the city [3]. Traffic signal control is a traffic control method that controls traffic signals and timing schemes by computer [4]. A traffic signal control system

with good reliability and high stability has the advantages of high efficiency [5]. With the rapid development of artificial intelligence, urban traffic signal control scheme recommendations, traffic intersection flow prediction and traffic intersection spatiotemporal data analysis are increasingly popular directions [6].

### B. LIMITATIONS OF EXISTING TRAFFIC SIGNAL CONTROL SYSTEMS

Single intersection signal control scheme recommendation is based on the real-time status of the traffic intersection [3]. The signal control scheme recommendation systems are built by characterizing the traffic state such as the traffic flow, saturation, and queue length of the roadway [7]. However, the traffic flow at traffic intersections is affected by factors that are difficult to be comprehensively counted, and they cannot be accurately modeled for single intersections [8].

The associate editor coordinating the review of this manuscript and approving it for publication was Laura Celentano .

This results in fact that existing signal control scheme recommendation system is still difficult to cope with complex traffic conditions [9]. The current mainstream algorithms are intelligent control methods based on fuzzy control and neural network control [10]. However, the rules of fuzzy control increase with the increase of control intersections. And too many rules will bring certain adverse effects on the operation speed [11]. And it has no self-learning function, yet cannot adapt to nonlinear and randomly changing traffic flow [12].

Traditional recurrent neural networks cannot model long-time dependent information in traffic flow sequences [13]. Deep learning is the process of learning the intrinsic laws and levels of representation of sample data. The information obtained from these learning processes can be of great help in the interpretation of data such as text, images, and sounds [14]. Its ultimate goal is to enable machines to have analytical learning capabilities like humans, capable of recognizing data. The literature [15] estimated the parameters of the construction cost budgeting phase, analyzed theoretically their correctness and reasonableness, programmed in MATLAB using the BP algorithm commonly used in artificial neural networks, and predicted new projects and found that their errors were significantly reduced concerning traditional methods.

### C. POTENTIAL OF DEEP REINFORCEMENT LEARNING

The reinforcement learning was developed to describe and solve problems where agents learn strategies to maximize returns. And this process is supported by the adaptive interactions between the agent and the environment. The application of deep reinforcement learning methods combined with vehicle networking technology to vehicle-road cooperative control decision-making for urban road traffic control is a current research hotspot and frontier [16]. And the optimal control strategy is learned by analyzing the training samples between traffic state changes and control actions [17]. In this paper, based on the multi-intelligence technology, we use the knowledge of game theory to divide three control levels: local level, subarea level, and area level, based on the distribution of traffic flow in the road network [18]. This paper firstly aims to optimize traffic flow control in urban road intersection areas. Then, it uses the closed loop formed by self-driving vehicles and signal controllers as the research object to achieve dynamic scheduling of traffic flow through deep reinforcement learning methods [19]. Through the proposed method, it is expected to improve the transportation efficiency.

### D. OUR CONTRIBUTIONS

For the control objects of each control level, we model regional, sub-regional, and intersection intelligence. Hence, we propose the traffic signal rolling control method based on Multi-agent Deep Reinforcement Learning (MADRL). Specifically, the decision-making ability of the intelligence is trained, the traffic signal control in the whole area of the road network is realized, and the intelligent traffic signal control

optimization system is designed to improve the operational efficiency of the entire traffic network. This also guarantees the overall control performance of the traffic control system. The software system carrying the solution is also able to provide technical service support for the traffic department, and improve the vitality and carrying capacity of urban traffic. Main contributions of this paper can be summarized as five points:

- We comprehensively analyze current challenges in existing researches on artificial intelligence-based vehicle-road collaborative systems.
- This work formulates a traffic signal control algorithm based on intersection clustering.
- This work formulates an Intelligent Vehicle Collaboration Solution based on MADRL
- The above two points constitute main technical framework of this paper: a collaborative control scheme for smart vehicles based on MADRL.
- Some simulations are conducted to verify efficiency of the proposal.

## II. RELATED WORK

### A. TRADITIONAL MATHEMATICAL MODELING-BASED COLLABORATIVE CONTROL APPROACHES

Traffic signal control is a traffic management measure that separates traffic flow rights-of-way in time, solving the problem of traffic flow that cannot be separated in space [20]. The traffic control system is the main facility to ensure traffic order, integrating the management concept and intention of traffic managers [21]. Traffic control technology has gone through the development process from single point control to line control to surface control, from timing control to induction control to adaptive control [22]. The target system of traffic control effect is gradually improved. At the micro level, the traffic control system optimizes the intersection traffic state to make the traffic flow through the intersection with minimum delay. At the level, the traffic control system optimizes the control parameters and joint control of multiple intersections. Hence, the main body of traffic flows smoothly through a cluster of intersections to achieve arterial or regional coordinated control.

At the macro level, the traffic control system adjusts and distributes the traffic flow as a whole, so that the traffic flow has a reasonable distribution on the road network and achieves a dynamic balance of traffic flow. The literature [15] first applied reinforcement learning for traffic signal control, and proposed four classical theories in reinforcement learning. Miao et al. [23] proposed a Markov decision process framework for adaptive control of traffic signals. But the transfer probabilities between traffic states need to be determined in advance. And its practical control applications are limited. A distributed multi-intelligence-based approach for traffic signal control was proposed in the literature [24].

Some researchers have started to use artificial intelligence techniques in the field of traffic signal control. But most of them only consider the control of individual intersections and

are not applied to the area control level. The literature [25] sets up target intersections, uses a cellular transport model to model the target, and sets up a multi-objective optimization algorithm by which various traffic operation metrics are calculated. In the literature [26], for the dynamic and uncertainty of intersection traffic flow, the Markov process is used to predict the occupancy rate of vehicles entering each inlet lane. And then, signal timing control of the intersection is optimized by constructing a reinforcement learning method based on Q-values. Zhang et al. [27] investigates special roundabout intersections by using predictive control ideas, establishing dynamic prediction models for the roundabouts and entrances, and obtaining the optimal timing parameters for roundabout intersection cycle control.

### B. ARTIFICIAL INTELLIGENCE-BASED COLLABORATIVE CONTROL APPROACHES

The literature [28] analyzes the relationship between the gradient and period under three traffic states of undersaturation, critical saturation, and oversaturation for the constraint relationship between evaluation index and capacity. Then, it uses the similarity function of the two as the optimization objective to time the signal. In [29], based on the time-varying characteristics of intersection traffic flow, a multi-objective signal timing optimization model with the minimization of delay, queue length, and several stops as the objective function is proposed. In this work, the case study shows that the optimization model can significantly improve the applicability and efficiency of signal control. Yu et al. [30] proposed the period and green signal ratio optimization model for single intersection timing control is developed using constraints that satisfy the intersection control in general. In this work, the applicability of the model and algorithm is verified with practical cases. Zhou et al. [31] proposes a single-intersection signal timing optimization method using different control objective functions based on the intersection occupancy and flow ratio relationships. This work verifies the superiority of the method using a real case in Shenzhen.

In [32], the relationship between minimum green time and vehicle delay and traffic safety is analyzed. Then, a single-lane minimum green time optimization model based on risk decision is established. For the multi-lane case, a multi-lane minimum green time optimization model and a maximum green time optimization model are also established. Tang et al. [33] proposed an adaptive signal control method with the control objective of minimizing vehicle travel time in road networks. Its advantages in reducing travel time and stopping times are demonstrated through numerical experiments. In literature [34], an improved priority traffic control method for emergency vehicles at intersections in a coordinated traffic signal timing and speed guidance is proposed. And simulation experiments are conducted using a simulation platform.

Wilson et al. [35] proposed a traffic signal control system using high-quality microscopic data and built an adaptive traffic signal control simulation platform using

deep reinforcement learning methods. The simulation results showed that the system outperformed other control systems. The vast majority of studies use a hypothetical static stochastic environment with fully independent or partially state-cooperative coordination mechanisms for optimal control of local intersections. This constrains the overall effectiveness of network traffic control systems, while there has been a rapid development of action-linkage-based MARL control methods [36].

## III. METHODOLOGY

### A. TRAFFIC SIGNAL CONTROL ALGORITHM BASED ON INTERSECTION CLUSTERING

Unlike traditional traffic signal control methods, reinforcement learning methods can adjust the control strategy as the traffic conditions change and have better adaptability. However, the performance of reinforcement learning-based traffic signal control methods is heavily dependent on the accurate modeling of the traffic environment. Vehicle dynamics information in some traffic networks is usually difficult to obtain in real-time due to the limitations of traffic infrastructure. For example, the number of vehicles in a lane is easy to obtain (usually only camera equipment is needed), while information such as vehicle speed and waiting time is not so easy to obtain. Accurate vehicle speed acquisition relies on the deployment of IoT devices such as speed cameras and millimeter wave radar, while the real-time acquisition of vehicle wait times requires continuous tracking of vehicles.

If accurate vehicle dynamics information cannot be obtained in real-time, existing reinforcement learning algorithms are difficult to adapt to real traffic environments or the learning process converges slowly. Therefore, it becomes a major challenge to rapidly construct high-quality reinforcement learning models for traffic signal control when the observation of vehicle dynamics information is limited. Most existing reinforcement learning-based traffic signal control methods focus on optimizing traffic from the perspective of a single traffic intersection [17]. Even in a multi-junction traffic environment, there is no interaction between the models controlling individual intersections. This design reduces the complexity of the method design and limits the control performance of reinforcement learning methods. Another part of the control methods considering intersection synergy optimizes the control from the perspective of neighboring intersection influence or dividing sub-regions [14].

However, the reinforcement learning modeling is often too complex, which seriously affects the convergence speed of the algorithm. Therefore, accelerating the convergence speed of the algorithm while considering intersection synergy is also a major challenge. Traffic pressure modeling and reinforcement learning modeling are performed with the limited observation of vehicle dynamic information. In addition, unlike other methods that do not consider intersection collaboration or design complex intersection collaboration algorithms, the algorithm in this paper divides all traffic intersections into clusters. Among, the reinforcement learn-

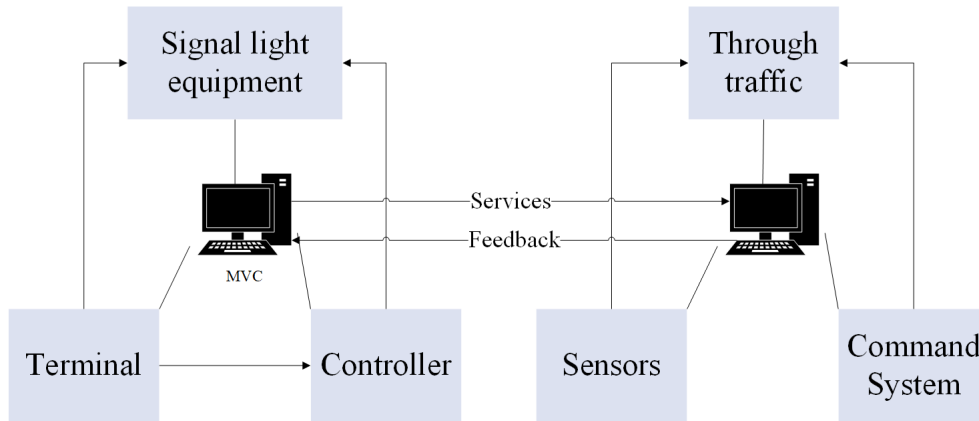


FIGURE 1. The overall architecture of ClusterLight.

ing model is designed to be very simple and effective. Intersections within the same class cluster collaborate by sharing traffic data.

For the case of limited observation of vehicle dynamic information, accurate traffic pressure modeling and reinforcement learning modeling based on the number of vehicles is performed [37]. The modeling scheme performs signal control from the perspective of minimizing intersection pressure. Experimental results in simulated traffic scenarios and real traffic scenarios with different road network sizes show that the algorithm in this paper can reduce the algorithm learning time. Figure 1 shows the overall architecture of ClusterLight in detail. As shown in the figure, the algorithm is mainly divided into two modules: the terminal on the lower side and the cloud on the upper side. Traffic intersections are clustered according to their locations and traffic flows. It can be seen that intersections I1 and I2 are classified into the same class cluster and interact with class cluster 1 in the cloud, while I3 and I4 are classified into the same class cluster and interact with class cluster x in the cloud. The algorithm deploys an edge node at the endpoint for each traffic intersection. As is shown by the blue arrow line, the traffic data collected by the edge node through sensors and other IoT devices.

A portion of this traffic data is transmitted directly to the neural network, which generates control phases and feeds them to the signal devices at the terminal [38]. The other part is transmitted as experience along with the control phases to an experience replay pool for storage, and is periodically removed by the neural network for training. In maximum pressure theory, the definition of traffic pressure is related to the number of vehicles in the lane. But not all vehicles in the lane will cause pressure on traffic during a phase time. When the lanes are long, only some of the vehicles close to the intersection have the opportunity to pass through the intersection. The effective distance is the longest distance a vehicle can drive through in one phase of time, and is defined as:

$$L_{mn} = e^{-\alpha} K' + e^{-\beta} M(x, y) \cdot X_{mn} \cdot Y_{mn} \quad (1)$$

where  $L_{mn}$  denotes long distance,  $e$  is a constant,  $K$  denotes traffic pressure,  $x, y$  denotes the traffic number on the road, and  $X_{mn}$  and  $Y_{mn}$  denote the location of car  $x, y$  to the  $n$ th intersection with  $n$  lanes, respectively.

Lane pressure reflects the traffic pressure exerted on an intersection by the vehicles in the effective section of the lane. Intuitively, the more vehicles in a lane, the more traffic pressure is exerted on the intersection by the vehicles in that lane. When the signal runs out of green time, the model needs to select a new phase that minimizes the traffic pressure on the intersection. ClusterLight defines the action as selecting the best control phase for the intersection  $\varphi_{\rho}(x, y)$ .

$$\varphi_{\rho}(x, y) = \varphi \sum_{y \in \gamma} \sum_{x \in \chi} \left[ \frac{p(x, y)}{\ln p(x, y)} - Ax - Cy \right] + \lambda \quad (2)$$

where  $\rho$  is the distance function and  $A$  and  $C$  are constants.

During the traffic signal control cycle, the first phase allows traffic flow in the NS direction to go straight and prohibits traffic flow in the WE direction. In the traffic signal control cycle, the second phase allows traffic flow in the NS direction to turn left and right and prohibits traffic flow in the WE direction to going straight. In the traffic signal control cycle, phase 3 allows traffic flow in the WE direction to go straight and prohibits traffic flow in the NS direction. In the traffic signal control cycle, the first phase allows the traffic flow in the WE direction to turn left and right and prohibits the traffic flow in the NS direction to go straight. The red light indicates that the east-west direction is prohibited and the north-south direction is allowed. The first phase, and the rest of the phases are analogous according to the traffic rules. The design of the traffic signal phase is one of the key factors affecting traffic congestion. If the phase time is designed too long, it will increase the average delay time of vehicles in opposite directions. If it is designed too short, it is not conducive to balancing the right-of-way of road vehicles and is prone to traffic accidents at intersections [30]. Most of the previous research work used discrete action spaces, where the signal controller selects one phase from all possible phases to act as each step of the simulation. The signal light controller



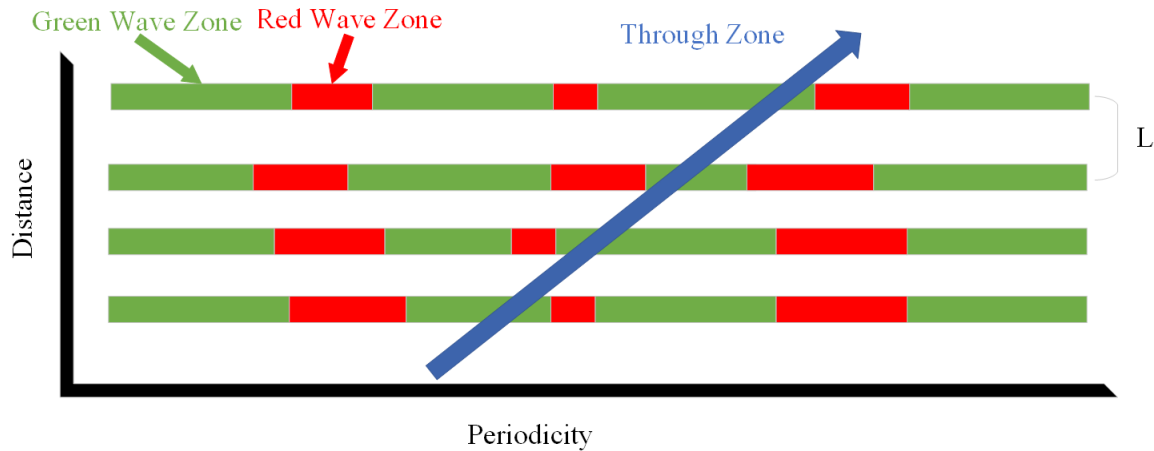


FIGURE 2. Diagram of green waveband control.

can choose to keep the current phase unchanged or switch to the next phase after the minimum phase duration has elapsed.

In this paper, the traffic light controllers in the region are controlled centrally, so the complete set of actions for the smart body traffic signal controller is defined as follows:

$$TL_{\text{action}} = [a_1, a_2, \dots, a_i, \dots, a_n] \quad (3)$$

where  $i$  is the state of a signal and  $n$  is the total number of signals. The signal cycle can be specifically divided into the optimal cycle  $C_0$ , the minimum cycle  $C_{\min}$  and the maximum cycle  $C_{\max}$ . Different cycle lengths are usually used according to the actual operating conditions of the intersection. The desired control effect, and the more commonly used is the optimal cycle, which is calculated by the following formula:

$$C_0 = \frac{aLoss + 5}{1 - Y} \quad (4)$$

where  $Y$  denotes the time loss constant, which is determined by the specific roadway environment. And  $Loss$  denotes the total cycle time lost, which is expressed as:

$$Loss = \sum_{t=0}^n (loss_t + I_t - A_T) \quad (5)$$

where  $loss_t$  denotes the loss of phase at moment  $t$ ,  $I$  denotes the green interval time of the phase, and  $A$  denotes the yellow light time.

Arterial coordinated control, as shown in Figure 2, is to consider multiple adjacent intersections on urban trunk roads as one system for coordinated control. The main control parameters involved in arterial coordinated control include cycle time, green signal ratio, and phase difference [39]. Arterial coordination control can be divided into one-way arterial coordination control and two-way arterial coordination control. Synchronous coordinated control is when the intersection spacing is quite short, and the traffic volume along the arterial direction is much larger than the traffic volume in the intersection direction. when the traffic volume of the arterial road is particularly large, the peak hour traffic

volume is close to the capacity. And the red light vehicle queue at the downstream intersection is likely to cross the upstream intersection, forming these intersections into a synchronous coordinated control system can avoid the traffic congestion situation.

However, in both cases, the use of a synchronized system results in additional stopping time for vehicles on the intersecting streets. In addition, in such systems, the disadvantage of having all green lights displayed ahead can cause drivers to speed up to catch the green light. In an interactive coordination system, signals connecting adjacent intersections in a system display opposite light colors at the same moment. When the vehicle travels between adjacent intersections for a time equal to half of the signal cycle duration, the vehicle can pass through the intersection continuously with the interactive coordination system. The continuous coordination system is based on the average running speed on the trunk line and the intersection spacing, to determine the phase difference of each adjacent intersection [19].

The costly, security and confidentiality issues associated with the acquisition of real traffic data have led to the significant use of traffic simulation in engineering practice. In addition, traffic simulation helps to make an effective evaluation of infrastructure and strategy changes before actually putting them on the road. Therefore, this paper uses an open-source, microscopic, multimodal traffic simulator, SUMO (Simulation of Urban Mobility). This can model multimodal transportation systems such as road vehicles, public transportation, and pedestrians, and contains several tools to support route finding, visualization, network import and emission calculation.

## B. INTELLIGENT VEHICLE COLLABORATION SOLUTION BASED ON MADRL

This algorithm uses the parameter policy to approximate the policy directly and updates the parameters of the characterized policy based on the gradient of the performance metrics [13]. In the Actor-Critic framework, the parametric

strategy is characterized using the parameter  $Sg(x)$  and the value function is characterized using the parameter  $A_i$  [40]. The following analysis of the strategy gradient method algorithm is performed. Taking the parametric strategy as an example, the general form of the update equation in the strategy improvement process can be expressed as:

$$S = \frac{1}{a^n} A_i \begin{pmatrix} n \\ \vdots \\ k \end{pmatrix} (-1)^k g(x - ak) \quad (6)$$

For the parametric strategy  $S_{sigma}$ , in the continuous action control problem, it is generally assumed that action selection obeys a Gaussian distribution. This method can effectively integrate strategy search and strategy  $x(t)$  exploitation [41]. Above for the continuous control problem, the parametric strategy characterized by using Gaussian distribution can be expressed as:

$$S_{sigma} = \frac{1}{t} \int_0^T [x(t) - \gamma(t)] dt + C \quad (7)$$

where  $\gamma(t)$  is the output of the function approximator concerning the parameter  $x(t)$ ,  $C$  denotes the mean of the action distribution and  $\alpha_1$  is the standard deviation of the action distribution associated with state  $s$ . Then Equation (1) can be expressed as follows:

$$L_{local} = \alpha_1 - \frac{(\alpha_i + \alpha_{i+1})(E_n^1 + \lambda_i)}{\left(\sum_{i=1}^k \lambda_i\right)} S_{sigma} \quad (8)$$

Therefore, the estimation of the error signal and the updating accuracy of the policy network are key factors for the algorithm efficiency. In the initial stage of learning, the estimates of the current state  $S_{sigma}$  and state  $S$  value functions  $L_{local}$   $L_{nn}$  are calculated by random initialization of the weights. This in turn makes the error signal difficult to be estimated accurately and eventually affects the update direction. At the same time, the sign of the error signal determines the updating direction of the parameter strategy relative to the action. In reinforcement learning algorithms, the values are generally small, when the actual update result is the direction of the parameter strategy update. When the initial parameter policy is poor, resulting in an intelligent body searching for an action with a poor error signal, this action  $a$  is in the same direction as the optimal action  $a^*$  concerning  $L_{local}$ . In addition, the actual system characteristics limit the executable search strategy, and the intelligent body allows only a limited step of action changes at adjacent moments. Thus, the action search variance can only be limited to a small range. In summary, an effective updating depends on a suitable search strategy with accurate evaluation. The evaluation signal calculated by Equation (8) can be used to evaluate whether the system state is closer to the desired value [42]. However, when the policy network is poorly initialized, the intelligence will get a negative evaluation signal after searching both directions, and cannot determine the correct policy update

direction. Therefore, further correction of the evaluation signals in such cases is needed. The search direction is further evaluated by comparing the difference in the return signals of different search directions. The global signal expression is as follows:

$$L_{global} = \tau_{min} \frac{\partial \gamma}{\partial j} + \frac{1}{\tau_{max}} \sum_{i=1}^n X_i Y_i \quad (9)$$

where  $\tau_{min}$  denotes the minimum response time and  $\tau_{max}$  denotes the maximum response time. Using the evaluation signal  $d$  as supervision, when the relationship of Eq. 10 is available:

$$\text{sign}(\delta) = \text{sign}(d) \quad (10)$$

The following equation is used to update the policy network based on the Bellman equation:

$$d\theta = d\theta + \alpha \delta \log(a | s, \theta) \quad (11)$$

The update direction of the strategy parameter network is opposite to the update (M1) direction of the strategy network based on the Bellman equation. Through the normalized evaluation method described in the previous section, the correct direction of the action search is found through several steps of exploration and comparison. This action search direction is used in the next step of the strategy search. In each batch dataset, the intelligence forces two directions of action search to update based on a priori knowledge. The above design approach makes the reinforcement learning method with less a priori knowledge or training data. The design of each signal such as specific  $dS$  will be specified by the vehicle longitudinal driving policy learning control problem. The driving process is modeled as a Markovian decision process. it is divided into state design, action design, and reward design. The state vector design needs to fully characterize the state. The following mode can be designed as:

$$N_{vi} = \Pi \left( \sum_{i=1}^3 V_i t + \theta \right) + \eta \quad (12)$$

where  $V_i$  denotes the state design, action design, and reward design states, respectively, and we weigh the sum of the global and local rewards, and the final presentation form of the reward function is:

$$N_{global} = EXP \left( \frac{\sum_{i=1}^n N_{vi}}{n} \right) \quad (13)$$

where  $EXP(\cdot)$  denotes the sigmoid function as follows:

$$EXP(x) = \frac{1}{1 + e^{-x}} \quad (14)$$

The communication mechanism allows each intelligence to simulate the global state of the environment, allowing them to make more accurate decisions. To achieve this, a communication module is designed that uses LSTM to encode the previous observations and behaviors, resulting in

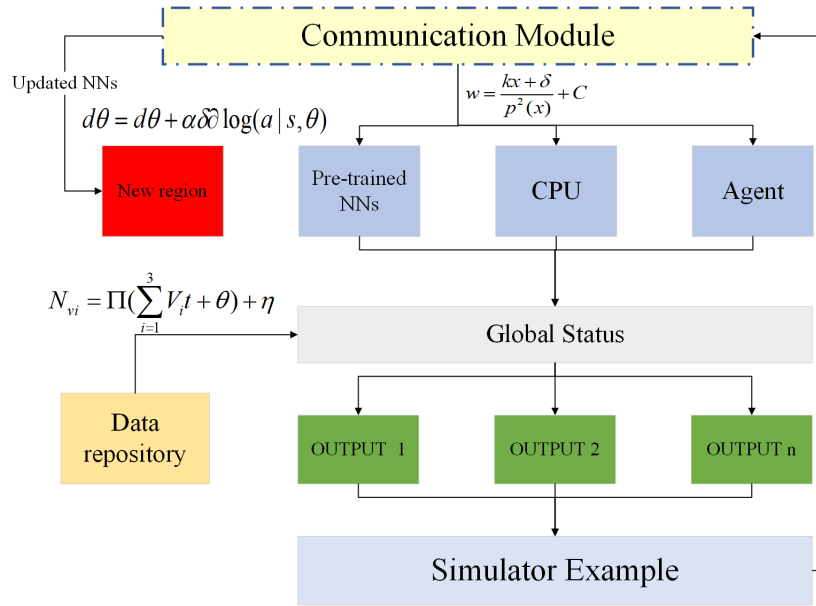


FIGURE 3. Diagram of green waveband control.

a vector form of information. By communicating between the intelligence, the overall state can be approximated as:

$$w = \frac{kx + \delta}{p^2(x)} + C \quad (15)$$

This is because the message  $p(x)$  already contains all previous observations and behaviors. Each agent selects one behavior  $w(t)$ :

$$w(t) = \frac{1.22}{\sqrt{1 + \left(\frac{Kt^2}{T}\right)^2}} + 0.43 \quad (16)$$

The goal is to maximize the overall cumulative return, which is evaluated by  $\frac{Kt^2}{T}$ . As Figure 3 shows the detailed structure of the algorithm, the information based on observations and behavior  $a$  is updated by the communication module, where the red squares indicate information and the blue squares indicate observations. More specifically, at time  $t$ , the agent receives the current observation from the environment. The global state of the environment is shared by all the bits of intelligence, relying not only on the historical state and behavior of each intelligence but also on the current observation.

Once the action returns, the simulator starts again to interact with the environment to generate data, and so on. The system-shared memory array provides fast communication between the action server and the simulator, and the server does not need to access the Tra CI port each time. Synchronous sampling may slow down due to the backward effect equating to the slowest process at each step. The variation in step time arises from different computational loads and other random perturbations in different simulator states. The lagging effect worsens as the number of parallel processes increases, but it is mitigated by stacking multiple independent

simulator instances in each process. Each process executes all simulators for each batch gradient computation.

## IV. EXPERIMENTAL VERIFICATION AND ANALYSIS

### A. EVALUATION ON TRAINING

The algorithm validation process in this paper is based on Spark cloud computing architecture deployed on cloud servers. The Spark cloud architecture consists of a primary node and multiple worker nodes. The red arrows indicate that the primary node broadcasts global learning rates and average parameters to the worker nodes. The blue arrows indicate that the worker nodes aggregate local data to the primary node. The efficiency of the road network before and after the introduction was compared by varying the vehicle penetration rates in four directions of 1600veh/h, 2400veh/h and 3600veh/h with the introduction of autonomous vehicles at 10%, 40% and 75%. The resource management node is dedicated to managing the resource scheduling and monitoring the operation status of the nodes, as shown in Figure 4. In the parallel iterative learning process of the CNN-LSTM prediction model, the local data update is considered the Map process, and the global data update is considered the Reduce process. In the Map phase, all worker nodes compute local gradient sums, and local loss functions and update local parameters in parallel based on a subset of locally cached data. In the Reduce phase, the primary node reconstructs the global learning rate and computes global learning parameters. Primary node broadcasts global data to all worker nodes in the cloud as the next iteration. The primary node broadcasts the global data to all worker nodes in the cloud as the initial values for the next iteration. This parallel iterative computation process continues until a preset number of iterations or prediction accuracy is met.

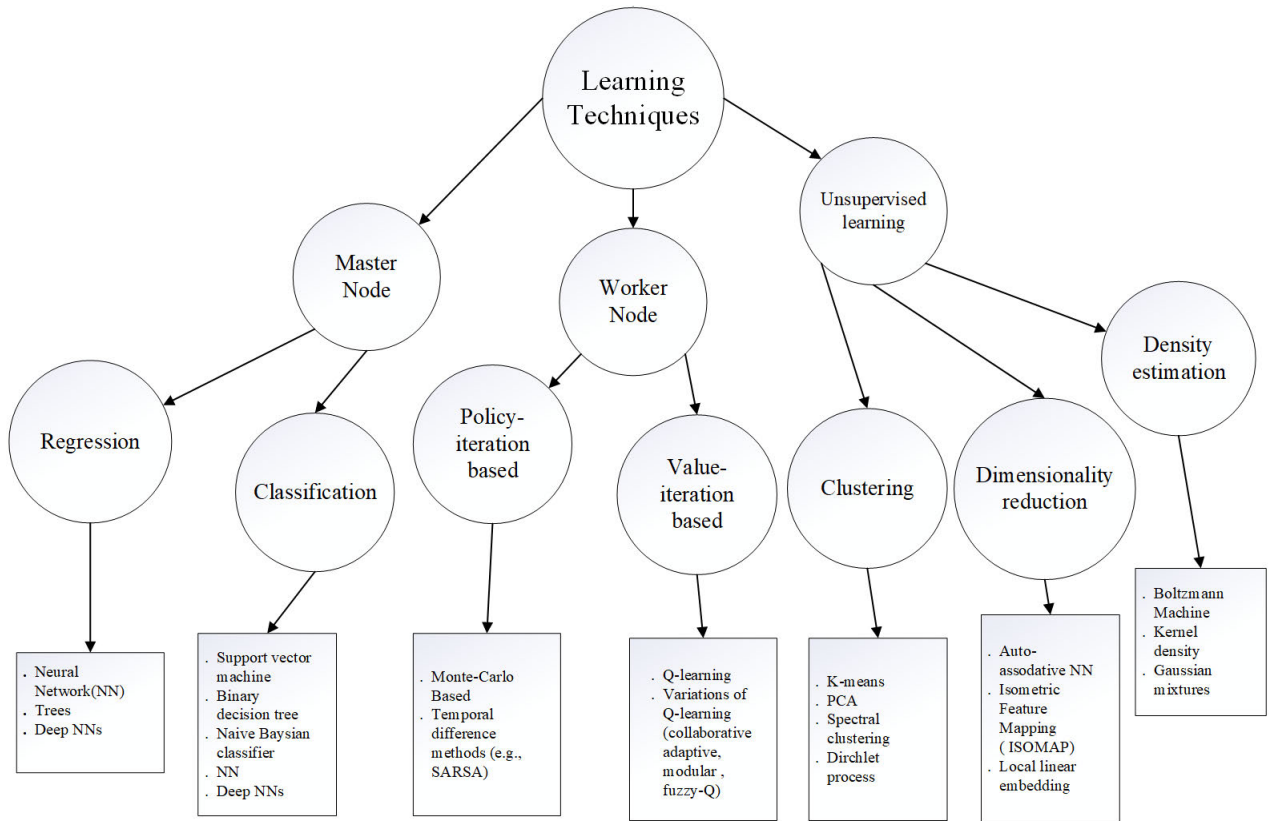


FIGURE 4. The overall structure of cloud parallel training.

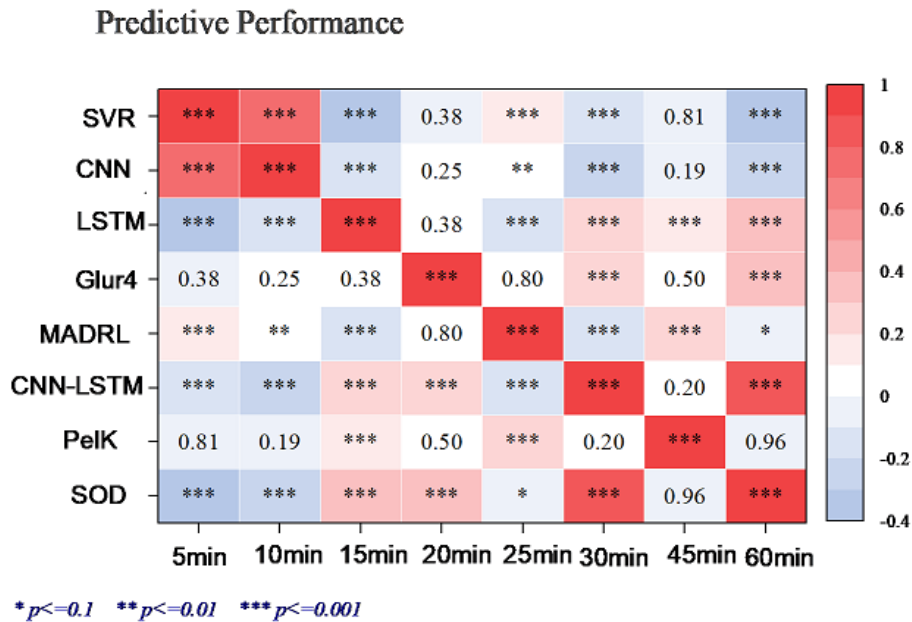
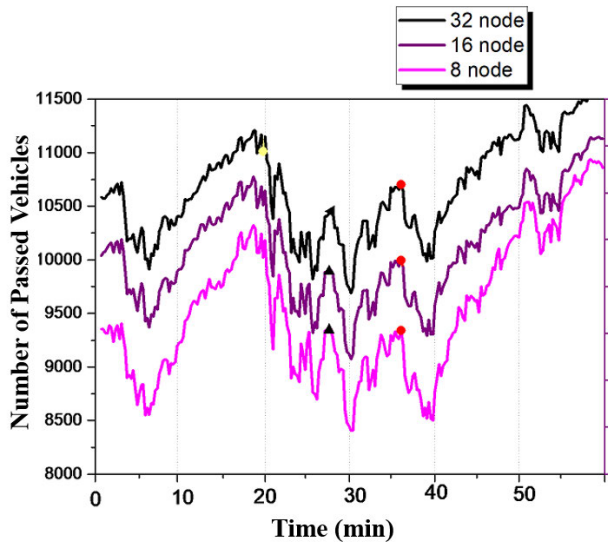


FIGURE 5. Comparison of prediction performance of prediction models in different prediction time domains.

The daily traffic flows in the road network have similar characteristics of random fluctuations, reflecting the relatively stable travel demand and regular traffic flow propagation. It can be observed from the dataset that the traffic

flows on the same day of the week exhibit a time-dependent periodic repetitive characteristic with irregular perturbations. The training dataset embeds the time-series fluctuation characteristics of traffic flow on the same road section and



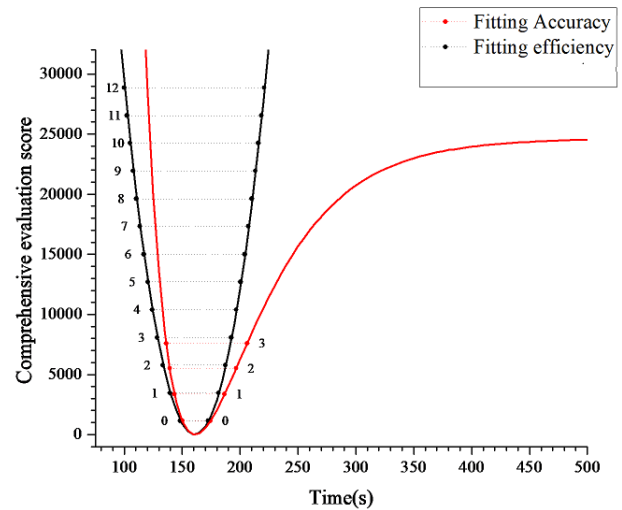


**FIGURE 6.** Number of vehicles passing through all intersections per simulation round.

the spatiotemporal coupling characteristics formed between multiple road sections in the road network. In this paper, 87,798 traffic flow data collected on 14 expressways are used as the data samples for this experiment. The first 10 months of data are used as the training data set with a sample size of 79173, and the remaining month of data is used as the test data set with a sample size of 8625. The entire training data set is decomposed into several subsets and distributed to different computing nodes in the Spark cloud. The global features of the entire dataset are obtained by aggregating the local learning features of the data subsets on multiple compute nodes. The remaining test dataset is used to verify whether the CNN-LSTM model parallel training method can extract the nonlinear spatiotemporal features of the traffic network flow. The future prediction time domains are defined as 5 min, 15 min, 30 min, and 60 min, respectively.

The prediction performance is shown in Figure 5. The experimental results show that the DTR method has the worst performance in traffic network flow prediction because the random perturbation of traffic data affects the stability of the decision tree algorithm. the SVR algorithm reduces the prediction error to some extent, by mapping the uncertain traffic flow data to a high-dimensional feature plane. However, the prediction error is still high because SVR cannot reflect the nonlinear connection between complex inputs and multidimensional outputs in solving the problem of multi-output regression analysis. Compared with the SVR method, the CNN method taps the spatial coupling features between multiple sections by convolution operations and reduces the MAE and RMSE error metrics by an average of 9.42% and 3.25% in different prediction time domains. the LSTM method further improves the prediction accuracy with its unique gate unit and correlation time series network structure.

The CNN-LSTM method based on parallel adaptive training combines the advantages of both CNN and LSTM,



**FIGURE 7.** The computational efficiency of offline training based on the edge computing cloud.

fully extracts the spatial correlation features between multiple sections of traffic network flows and their respective time series features, and further improves the prediction accuracy through the fully connected structure. The CNN-LSTM method is significantly smaller than other prediction methods in terms of MAE and RMSE error performance metrics for traffic network flow prediction tasks in different prediction time domains. Overall, the CNN-LSTM prediction method decreases approximately 29.47%, 25.24%, 17.44%, and 12.11% in MAE error measures, and 28.93%, 16.6%, 13.77%, and 8.43% in RMSE error measures, respectively, compared with the DTR, SVR, CNN, and LSTM methods. These data comparison results show that the CNN-LSTM parallel prediction method can effectively extract the spatiotemporal characteristics of traffic network flow in traffic big data and improve the prediction effect. Compared to CNN-LSTM, MADRL avoids getting caught in the local optimal solution, and the overall prediction accuracy is higher, and there is only some error in the 20min traffic prediction, so MADRL is the most suitable traffic prediction algorithm.

## B. EVALUATION ON COLLABORATIVE CONTROL

The parallel and serial schemes using different numbers of computational nodes have slight oscillations during training, which are caused by the  $\mu$ -greedy greedy search mechanism. In the early stages of training, the multiple intelligences explore actions randomly with a large probability within the constraint, which causes the training curve to fluctuate around a low value. As the number of training episodes increases, the probability of the actor-critic being adopted as a controller gradually increases and the training curve starts to climb upward. At the later stage of training, the training curve gradually converges as multiple intelligences learn to cooperate, relying on feedback rewards and contribution allocation mechanisms. The MADRL parallel training scheme and the serial training scheme based on different degrees of

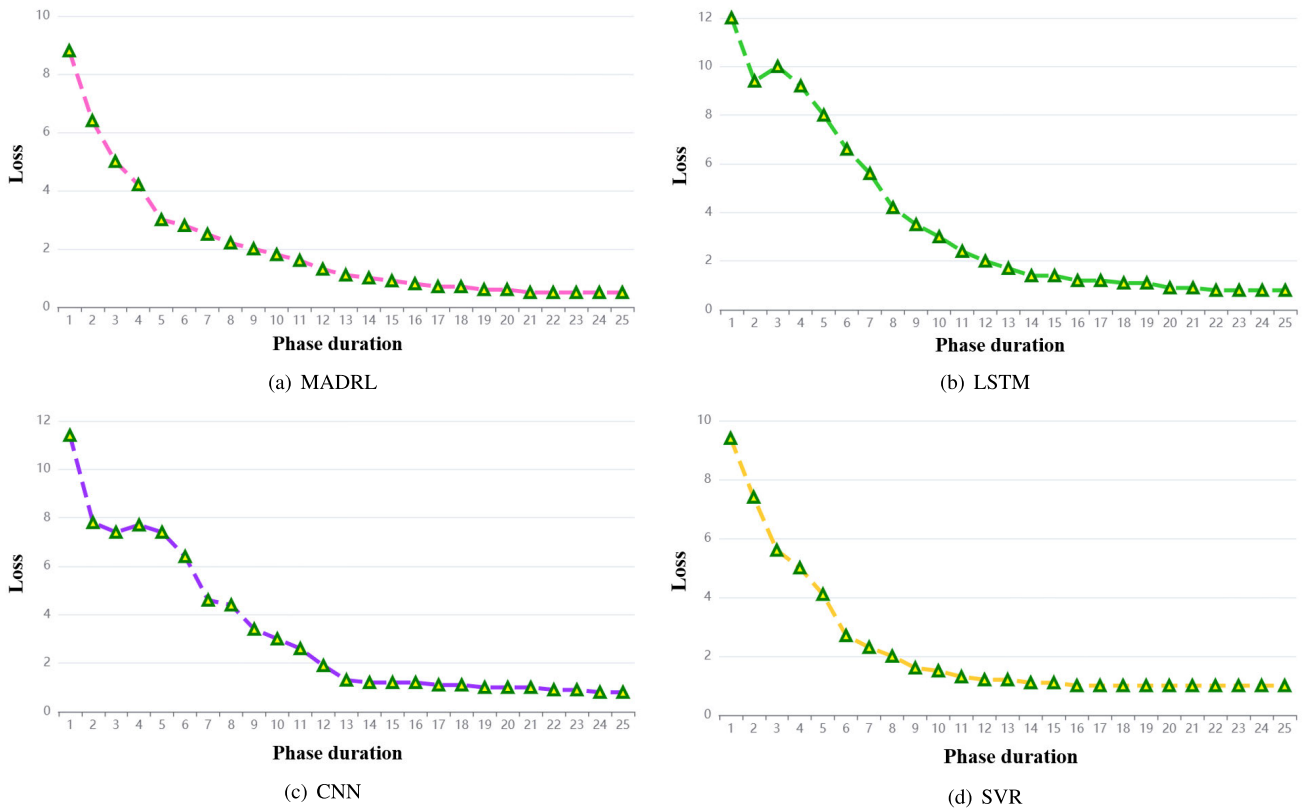


FIGURE 8. Tendency of loss values with respect to experimental methods (key points-based curves).

parallelism eventually converge to a near-stable threshold, as shown in Figure 6. This indicates that the MADRL parallel computing method does not degrade the accuracy of the serial training scheme.

The performance comparison of the computation time consumed by CMAC training using a different number of computation nodes is shown in Figure 7. In Figure 7, the training time decreases gradually with the number of computational nodes. The traditional serial training scheme that requires a large amount of computation time spent and memory consumption. However, for those parallel training schemes that use different numbers of computational nodes, the more the number of computational nodes, the less the number of actor-critic intelligence is allocated to each node, and the less computational time and memory are required. This is because large computational loads, such as parameter updates, action generation, and value function computation tasks, are computed synchronously in a parallel manner by multiple computational nodes assigned to the cloud. In particular, the efficiency of MADRL parallel computing is maximized when only one actor-critic intelligent body decision task is assigned to one node of edge computing. As shown in the figure, the acceleration ratio curve of the MADRL parallel training method shows an increasing trend with the number of computational nodes. However, for Spark clusters of different sizes deployed in the cloud, the load generated by resource scheduling, task allocation, communication, and synchronization among compute nodes

is uneven, so the increasing trend of this speedup ratio is non-linear.

Taking MADRL, LSTM, CNN and SVR as an example, their training process is visualized via two format of charts. The changing tendency of loss values is selected as the core metric for presentation. It is displayed via two types of charts: key points-based curves and continuous smooth curves. The former is shown as Figure 8, and the latter is shown as Figure 9. They both have four subfigures which correspond to circumstances of MADRL, LSTM, CNN and SVR. For each subfigure, its X-axis denotes the phase duration changing from 1 to 25, and its Y-axis denotes the loss values. The two kinds of charts can well display the changing tendency of training process via different visualization effect. It can be seen from the figures that experimental methods can tend to converge after some iterative rounds. In addition, we also make some comparison between MADRL and other three methods in terms of control quality, which is shown as Figure 10. For MADRL, the shorter the phase duration, the better the control quality of the strategy, because when the phase duration is reduced, less green time is wasted. The traffic network under MADRL control not only always has the smallest average travel time, but also learns faster than the MADRL-based one, which starts converging in the first ten rounds of learning.

Comprehensive experiments were conducted under four simulated traffic datasets and two real traffic datasets to validate the effectiveness of MADRL. Compared with

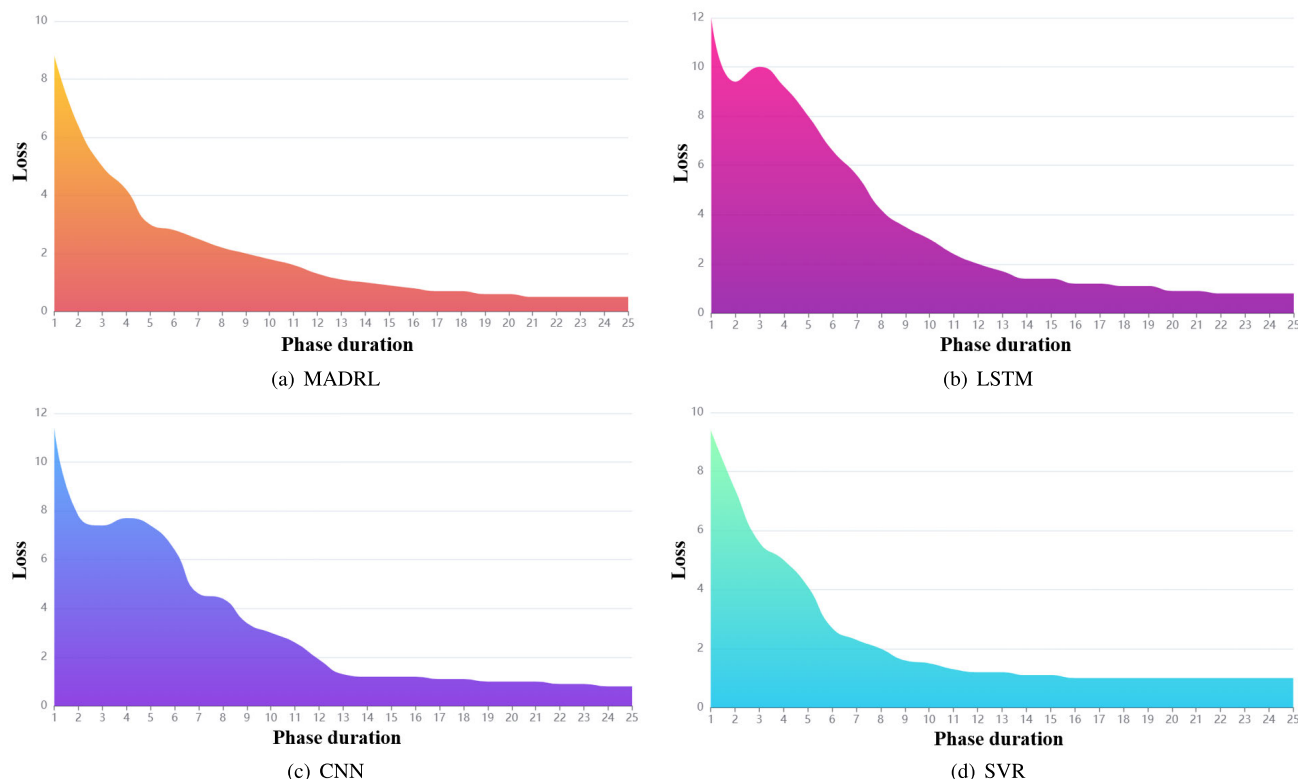


FIGURE 9. Tendency of loss values with respect to experimental methods (continuous smooth curves).

traditional signal control algorithms and three advanced reinforcement learning control algorithms, MADRL can improve the global traffic network efficiency while avoiding a few vehicles waiting all the time and enhancing people’s travel experience. At the same time, MADRL has a faster convergence rate and can learn the optimal control strategy quickly. The state information acquisition device serves the control algorithm integration deployed in the cloud, and the traffic light device receives the control action to ensure the traffic operation. After receiving the status information, the algorithm integration part performs calculations from the status information and returns the calculated control commands to the traffic light devices. The road condition information acquisition equipment and the average vehicle travel time acquisition equipment transmit traffic statistics to the cloud visualization platform in real time for traffic monitoring. Meanwhile, the vehicle information acquisition device records the vehicle id, the intersection, the lane, the specific location, and the waiting time. After the vehicle information is uploaded to the cloud, the visualization platform will show the details of the four vehicles with the longest waiting time.

We used TraCI to assign different ports to run the traffic instances of this experiment in SUMO, using 8 CPU parallel processes to run, where the information about the intelligence interacting with multiple environments can communicate with each other, increasing the randomness of the training samples and facilitating the learning of the experiences gathered from different environments. The average rates of

the road network trained by introducing 10%, 40%, and 75% autonomous vehicles under vehicle penetration (VP, vehicles penetration) of 1600veh/h, vehicle penetration of 2400veh/h, and vehicle penetration of 3600veh/h are shown in Figure 11, where the penetration rate indicates the one-way penetration rate in four directions. The horizontal axis indicates the simulation time (s), and the vertical axis indicates the regional average rate (km/h). From the analysis of the training results, the average rate increase is the largest when the vehicle permeability is 1600veh/h and the road network is less saturated, and the average rate increase becomes smaller as the road network becomes more and more saturated, but the average rate is still significantly higher than the signal control. By the analysis of the experimental conclusion of the fixed timing scheme, we know that the system balance depends on the uniformity of the vehicle rate, the more uniform the vehicle rate, the more balanced the system. Under the above three groups of different road network saturation conditions, the MADRL model with the introduction of autonomous vehicles all shows more stable growth, while the average rate variance of the road network under signal control is larger, indicating that the system is in an unbalanced state.

Figure 12 shows the critic network loss under the two control strategies after smoothing, from the figure it can be seen that the signal light control model converges too early and may fall into local optimum, while the model converges slowly under the vehicle-road cooperative control model, here it is because in the vehicle-road cooperative control

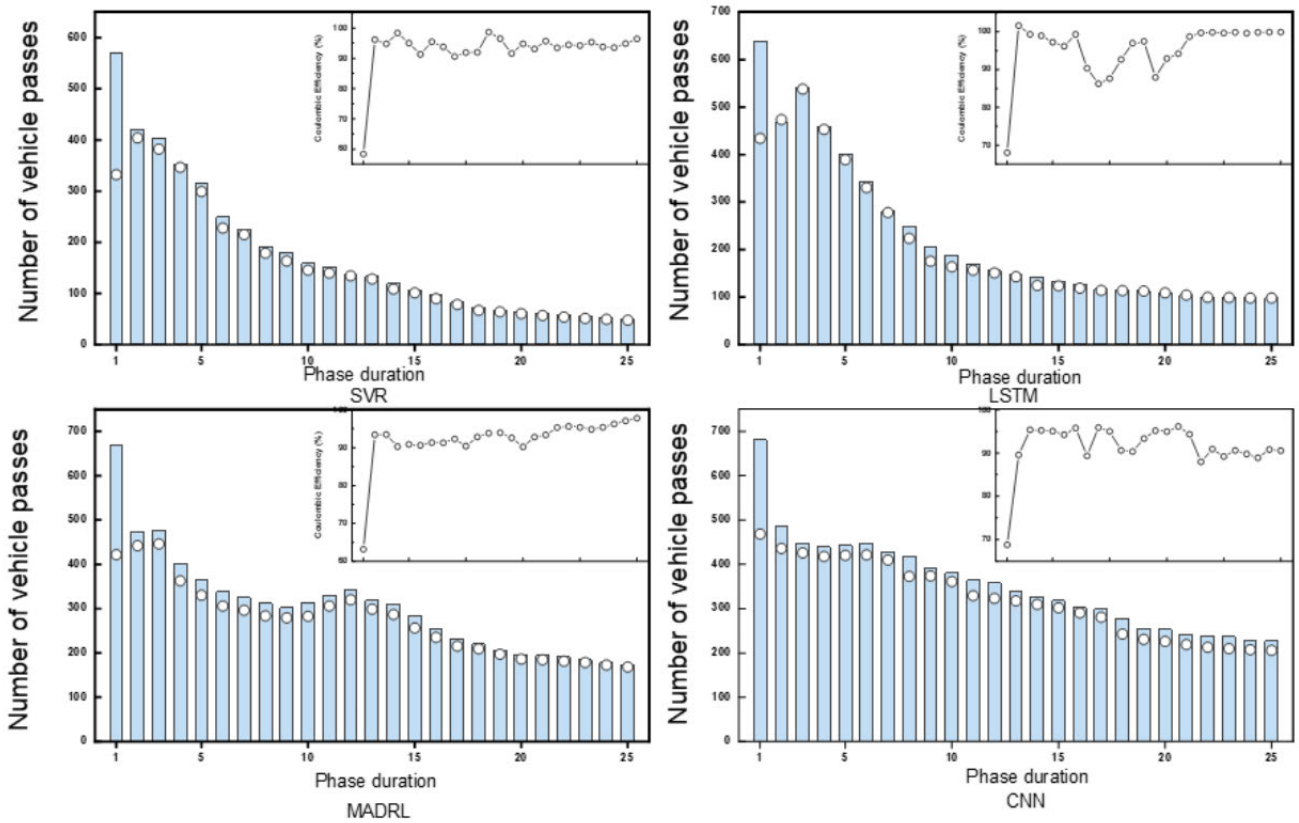


FIGURE 10. Comparison among experimental methods with respect to control quality and algorithm convergence.

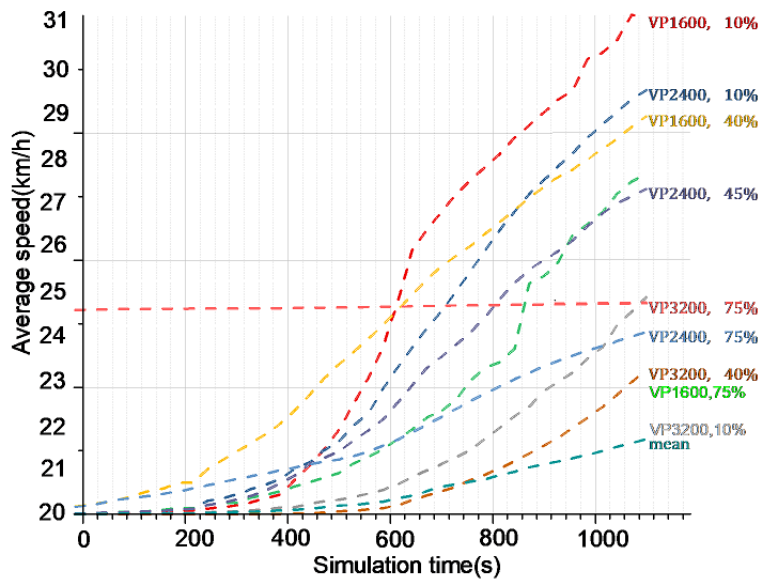


FIGURE 11. Comparison experiment under different vehicle permeability.

model, TLcontroller and AVcontroller share an objective function and the two bits of intelligence Considering each other's current part of observable state information, the two parties establish a cooperation mechanism, and at the same time, the sample data is random and the strategy uses joint optimization to avoid falling into local optimum. The

experimental results show that in the high saturation state of the road network, the regional traffic throughput is improved by 23.6% on average and the average speed is improved by 30.7% compared with the single-signal control, and the parallel process makes the model training time significantly optimized compared with the signal control.

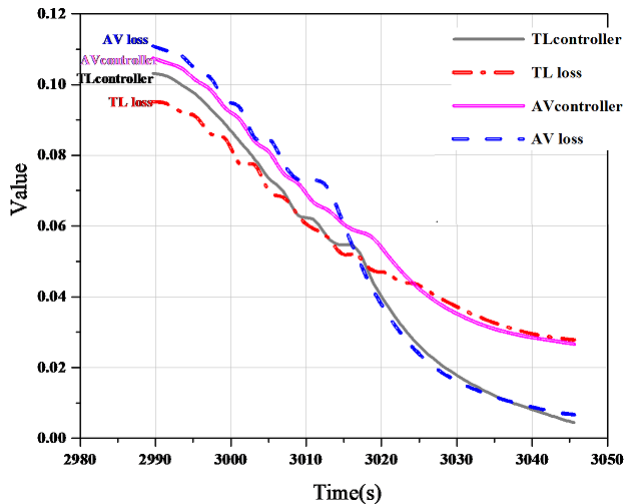


FIGURE 12. Critic network losses under different control strategies.

### C. DISCUSSION

For the technical methodology, we utilize the deep reinforcement learning to improve the collaborative control efficiency for transportation signal systems. The above two subsections have well verified efficiency of the proposed collaborative control method for transportation signal systems. It can be observed from the experiments that the proposal can converge to a relatively stable status after some iterations. It can be also observed that proper control effect can be achieved. However, it is noted that the proposal still suffers from two aspects of challenge. The first challenge is the real-time updating on large-scale data stream. In realistic large-scale data stream, the dynamic parameter is important for the models. The larger updating frequency is, the better dynamics the models have. The second challenge is the balance between model complexity and sample diversity. In realistic engineering scenarios, the statistical characteristics of samples is changing dynamically. But general deep learning models are with fixed structures. Will the models have the ability to adaptively adjust their model structures according to time-varying samples? This is also a future research point.

### V. CONCLUSION

Along with the development of urbanization, traffic congestion is becoming more and more prominent. Traffic congestion not only has an impact on the environment and economy but also leads to a huge waste of time and seriously reduces people's travel experience. As a controlled key device in the traffic system, intelligent control of signals is crucial to reduce traffic congestion. A traffic signal control algorithm based on reinforcement learning and intersection clustering is proposed for traffic scenarios with the limited observation of vehicle dynamic information. In this paper, traffic pressure is modeled based on the number of vehicles in the lane under the restricted observation of vehicle dynamic information, and the state, action, and reward of the reinforcement learning method are designed based on the traffic pressure. In addition, this paper clusters traffic intersections based on location

information and traffic flow to form multiple reinforcement learning models for centralized control. Through accurate reinforcement learning modeling and centralized control of intersections, the algorithm can learn high-quality signal control strategies quickly. In this paper, a dynamic selection strategy for phase duration is designed based on the real-time traffic state. Combining the traffic intensity modeling and the design of dynamic phase duration, this paper can significantly reduce the average vehicle travel time. Meanwhile, the waiting time consideration can avoid vehicles from waiting for a long time at intersections, which improves people's travel experience. Experimental results on real traffic datasets and simulated traffic datasets with different road network sizes show that the method proposed in this paper can significantly reduce the average vehicle travel time while shortening the control strategy learning time compared with other state-of-the-art traffic signal control methods.

### REFERENCES

- [1] Q. Li, L. Liu, Z. Guo, P. Vijayakumar, F. Taghizadeh-Hesary, and K. Yu, "Smart assessment and forecasting framework for healthy development index in urban cities," *Cities*, vol. 131, Dec. 2022, Art. no. 103971.
- [2] Z. Guo, K. Yu, A. Jolfaei, G. Li, F. Ding, and A. Beheshti, "Mixed graph neural network-based fake news detection for sustainable vehicular social networks," *IEEE Trans. Intell. Transp. Syst.*, early access, Jul. 7, 2022, doi: 10.1109/TITS.2022.3185013.
- [3] Z. Guo, K. Yu, K. Konstantin, S. Mumtaz, W. Wei, P. Shi, and J. J. P. C. Rodrigues, "Deep collaborative intelligence-driven traffic forecasting in green Internet of Vehicles," *IEEE Trans. Green Commun. Netw.*, vol. 7, no. 2, pp. 1023–1035, Jun. 2023.
- [4] B. Liang, S. Zheng, C. K. Ahn, and F. Liu, "Adaptive fuzzy control for fractional-order interconnected systems with unknown control directions," *IEEE Trans. Fuzzy Syst.*, vol. 30, no. 1, pp. 75–87, Jan. 2022.
- [5] Y. Xu, J. Lin, H. Gao, R. Li, Z. Jiang, Y. Yin, and Y. Wu, "Machine learning-driven APPs recommendation for energy optimization in green communication and networking for connected and autonomous vehicles," *IEEE Trans. Green Commun. Netw.*, vol. 6, no. 3, pp. 1543–1552, Sep. 2022.
- [6] C. Chen, J. Huang, D. Wu, and X. Tu, "Interval type-2 fuzzy disturbance observer-based T-S fuzzy control for a pneumatic flexible joint," *IEEE Trans. Ind. Electron.*, vol. 69, no. 6, pp. 5962–5972, Jun. 2022.
- [7] L. Zhao, Z. Bi, A. Hawbani, K. Yu, Y. Zhang, and M. Guizani, "ELITE: An intelligent digital twin-based hierarchical routing scheme for software-defined vehicular networks," *IEEE Trans. Mobile Comput.*, vol. 22, no. 9, pp. 5231–5247, Sep. 2023.
- [8] J. Zhang, Q. Yan, X. Zhu, and K. Yu, "Smart industrial IoT empowered crowd sensing for safety monitoring in coal mine," *Digit. Commun. Netw.*, vol. 9, no. 2, pp. 296–305, Apr. 2023.
- [9] Z. Ning, M. C. Zhou, Y. Yuan, E. C. H. Ngai, and R. Y. Kwok, "Guest editorial special issue on collaborative edge computing for social Internet of Things systems," *IEEE Trans. Computat. Social Syst.*, vol. 9, no. 1, pp. 59–63, Feb. 2022.
- [10] Z. Liu, J. Yu, and H.-K. Lam, "Passivity-based adaptive fuzzy control for stochastic nonlinear switched systems via T-S fuzzy modeling," *IEEE Trans. Fuzzy Syst.*, vol. 31, no. 4, pp. 1401–1408, Apr. 2023.
- [11] Q. Zhang, Z. Guo, Y. Zhu, P. Vijayakumar, A. Castiglione, and B. B. Gupta, "A deep learning-based fast fake news detection model for cyber-physical social services," *Pattern Recognit. Lett.*, vol. 168, pp. 31–38, 2023.
- [12] J. Yang, F. Lin, C. Chakraborty, K. Yu, Z. Guo, A.-T. Nguyen, and J. J. P. C. Rodrigues, "A parallel intelligence-driven resource scheduling scheme for digital twins-based intelligent vehicular systems," *IEEE Trans. Intell. Vehicles*, vol. 8, no. 4, pp. 2770–2785, Apr. 2023.
- [13] A. Alzubaidi, A. S. Al Sumaiti, Y.-J. Byon, and K. A. Hosani, "Emergency vehicle aware lane change decision model for autonomous vehicles using deep reinforcement learning," *IEEE Access*, vol. 11, pp. 27127–27137, 2023.



- [14] H.-B. Choi, J.-B. Kim, Y.-H. Han, S.-W. Oh, and K. Kim, "MARL-based cooperative multi-AGV control in warehouse systems," *IEEE Access*, vol. 10, pp. 100478–100488, 2022.
- [15] Z. Wang, S. Wang, Z. Zhao, and M. Sun, "Trustworthy localization with EM-based federated control scheme for IIoTs," *IEEE Trans. Ind. Informat.*, vol. 19, no. 1, pp. 1069–1079, Jan. 2023.
- [16] R. S. Sutton and A. G. Barto, *Reinforcement Learning: An Introduction*. Cambridge, MA, USA: MIT Press, 2018.
- [17] T. Haarnoja, A. Zhou, P. Abbeel, and S. Levine, "Soft actor-critic: Off-policy maximum entropy deep reinforcement learning with a stochastic actor," 2018, *arXiv:1801.01290*.
- [18] S. Pareek, H. J. Nisar, and T. Kesavadas, "AR3n: A reinforcement learning-based assist-as-needed controller for robotic rehabilitation," *IEEE Robot. Autom. Mag.*, early access, Jun. 21, 2023, doi: 10.1109/MRA.2023.3282434.
- [19] I. Grondman, L. Busoniu, G. A. D. Lopes, and R. Babuska, "A survey of actor-critic reinforcement learning: Standard and natural policy gradients," *IEEE Trans. Syst., Man, Cybern., C Appl. Rev.*, vol. 42, no. 6, pp. 1291–1307, Nov. 2012.
- [20] J. Yang, M. Xi, J. Wen, Y. Li, and H. H. Song, "A digital twins enabled underwater intelligent internet vehicle path planning system via reinforcement learning and edge computing," *Digit. Commun. Netw.*, early access, 2022, doi: 10.1016/j.dcan.2023.07.003.
- [21] L. Yan, X. Chen, Y. Chen, and J. Wen, "A cooperative charging control strategy for electric vehicles based on multiagent deep reinforcement learning," *IEEE Trans. Ind. Informat.*, vol. 18, no. 12, pp. 8765–8775, Dec. 2022.
- [22] J. Yang, J. Huo, M. Xi, J. He, Z. Li, and H. H. Song, "A time-saving path planning scheme for autonomous underwater vehicles with complex underwater conditions," *IEEE Internet Things J.*, vol. 10, no. 2, pp. 1001–1013, Jan. 2023.
- [23] Q. Miao, H. Lin, J. Hu, and X. Wang, "An intelligent and privacy-enhanced data sharing strategy for blockchain-empowered Internet of Things," *Digit. Commun. Netw.*, vol. 8, no. 5, pp. 636–643, Oct. 2022.
- [24] B. Liu, X. Jiang, X. He, L. Qi, X. Xu, X. Wang, and W. Dou, "A deep learning-based edge caching optimization method for cost-driven planning process over IIoT," *J. Parallel Distrib. Comput.*, vol. 168, pp. 80–89, Oct. 2022.
- [25] F. Adelantado, M. Ammouriova, E. Herrera, A. A. Juan, S. S. Shinde, and D. Tarchi, "Internet of Vehicles and real-time optimization algorithms: Concepts for vehicle networking in smart cities," *Vehicles*, vol. 4, no. 4, pp. 1223–1245, Nov. 2022.
- [26] J. Qin and J. Liu, "Multi-access edge offloading based on physical layer security in C-V2X system," *IEEE Trans. Veh. Technol.*, vol. 71, no. 7, pp. 6912–6923, Jul. 2022.
- [27] H. Zhang, L. Z. Liu, H. Xie, Y. Jiang, J. Zhou, and Y. Wang, "Deep learning-based robot vision: High-end tools for smart manufacturing," *IEEE Instrum. Meas. Mag.*, vol. 25, no. 2, pp. 27–35, Apr. 2022.
- [28] M. Hosseinzadeh, A. Hemmati, and A. M. Rahmani, "Clustering for smart cities in the Internet of Things: A review," *Cluster Comput.*, vol. 25, no. 6, pp. 4097–4127, Dec. 2022.
- [29] S. Daftry, N. Abcouwer, T. D. Sesto, S. Venkatraman, J. Song, L. Igel, A. Byon, U. Rosolia, Y. Yue, and M. Ono, "MLNav: Learning to safely navigate on Martian terrains," *IEEE Robot. Autom. Lett.*, vol. 7, no. 2, pp. 5461–5468, Apr. 2022.
- [30] K. Yu, L. Lin, M. Alazab, L. Tan, and B. Gu, "Deep learning-based traffic safety solution for a mixture of autonomous and manual vehicles in a 5G-enabled intelligent transportation system," *IEEE Trans. Intell. Transp. Syst.*, vol. 22, no. 7, pp. 4337–4347, Jul. 2021.
- [31] X. Zhou, X. Xu, W. Liang, Z. Zeng, and Z. Yan, "Deep-learning-enhanced multitarget detection for end-edge-cloud surveillance in smart IoT," *IEEE Internet Things J.*, vol. 8, no. 16, pp. 12588–12596, Aug. 2021.
- [32] H. Song, S. Zhou, Z. Chang, Y. Su, X. Liu, and J. Yang, "Collaborative processing and data optimization of environmental perception technologies for autonomous vehicles," *Assem. Autom.*, vol. 41, no. 3, pp. 283–291, Jul. 2021.
- [33] F. Tang, B. Mao, N. Kato, and G. Gui, "Comprehensive survey on machine learning in vehicular network: Technology, applications and challenges," *IEEE Commun. Surveys Tuts.*, vol. 23, no. 3, pp. 2027–2057, 3rd Quart., 2021.
- [34] Y. Xianjia, J. P. Queralta, J. Heikkonen, and T. Westerlund, "Federated learning in robotic and autonomous systems," *Proc. Comput. Sci.*, vol. 191, pp. 135–142, 2021.
- [35] A. N. Wilson, A. Kumar, A. Jha, and L. R. Cenkaramaddi, "Embedded sensors, communication technologies, computing platforms and machine learning for UAVs: A review," *IEEE Sensors J.*, vol. 22, no. 3, pp. 1807–1826, Feb. 2022.
- [36] C. Liu, Y. Feng, D. Lin, L. Wu, and M. Guo, "Iot based laundry services: An application of big data analytics, intelligent logistics management, and machine learning techniques," *Int. J. Prod. Res.*, vol. 58, no. 17, pp. 5113–5131, Sep. 2020.
- [37] A. Boukerche, D. Zhong, and P. Sun, "A novel reinforcement learning-based cooperative traffic signal system through max-pressure control," *IEEE Trans. Veh. Technol.*, vol. 71, no. 2, pp. 1187–1198, Feb. 2022.
- [38] X. Kong, H. Gao, G. Shen, G. Duan, and S. K. Das, "FedVCP: A federated-learning-based cooperative positioning scheme for social Internet of Vehicles," *IEEE Trans. Computat. Social Syst.*, vol. 9, no. 1, pp. 197–206, Feb. 2022.
- [39] J. Yang, B. Guo, Z. Wang, and Y. Ma, "Hierarchical prediction based on network-representation-learning-enhanced clustering for bike-sharing system in smart city," *IEEE Internet Things J.*, vol. 8, no. 8, pp. 6416–6424, Apr. 2021.
- [40] A. G. Barto, R. S. Sutton, and C. W. Anderson, "Looking back on the actor-critic architecture," *IEEE Trans. Syst., Man, Cybern., Syst.*, vol. 51, no. 1, pp. 40–50, Jan. 2021.
- [41] S. Garg, K. Kaur, G. Kaddoum, P. Garigipati, and G. S. Auja, "Security in IoT-driven mobile edge computing: New paradigms, challenges, and opportunities," *IEEE Netw.*, vol. 35, no. 5, pp. 298–305, Sep. 2021.
- [42] H. Lu, P.-H. Ho, and M. Guizani, "Guest editorial: Special issue on Internet of Things for industrial security for smart cities," *IEEE Internet Things J.*, vol. 8, no. 8, pp. 6140–6142, Apr. 2021.



**LIYAN SHI** received the B.Eng. degree from the Zhengzhou University of Aeronautics, in 2004, and the M.Sc. degree from the Huazhong University of Science and Technology, in 2008. She is currently an Associate Professor with the Information Engineering and Artificial Intelligence School, The Open University of Henan. She has published five articles. Her current research interests include image processing, software engineering, and virtual reality technology.



**HAIRUI CHEN** received the B.Eng. degree in computer science and technology from Zhengzhou University, in 2002, the M.Sc. degree in computer application technology from the Beijing Institute of Technology, in 2005, and the Ph.D. degree from the Chengdu University of Technology, in 2015. She is currently an Associate Professor with the Zhongyuan University of Technology. Her current research interests include software engineering and virtual reality technology.

...