Traffic & Transportation

Attention Mechanism-Based Convolutional Model for Collaborative Control of Traffic Signals at Multiple Intersections

Junlin ZHANG[1], Shiliang ZHANG[2], Jun QIAO[3], Jianping XU[4], Hao XIANG[5]

Original Scientific Paper
Submitted: 28 July 2025
Accepted: 10 Oct 2025
Published: 27 May 2026

[1] zhangjunlin@my.swjtu.edu.cn, School of Electrical Engineering, Southwest Jiaotong University, Chengdu, China; Sichuan CRRC SRIG Rail Transit Co., Ltd., Yibin, China
[2] Corresponding author, zhangshiliang@my.swjtu.edu.cn, School of Transportation and Logistics, Southwest Jiaotong University, Chengdu, China; Sichuan CRRC SRIG Rail Transit Co., Ltd., Yibin, China
[3] HAXUE007@163.com, New-type Rail Transit Design and Research Institute, China Railway Siyuan Survey and Design Group Co., Ltd., Suzhou, China
[4] jpxu-swjtu@163.com, School of Electrical Engineering, Southwest Jiaotong University, Chengdu, China
[5] xianghaoXXC@163.com, School of Electrical and Control Engineering, Xi'an University of Science and Technology, Xi'an, China; Sichuan CRRC SRIG Rail Transit Co., Ltd., Yibin, China

Publisher:
Faculty of Transport and Traffic Sciences, University of Zagreb

ABSTRACT

The necessity for collaborative traffic signal control optimisation is growing in importance as intelligent transportation technology continues to advance. The study started a related application investigation to successfully raise the degree of cooperative control of traffic lights at several crossings. Based on an in-depth analysis of the traffic scenarios at multiple intersections, the study clarified the key points for optimising the collaborative control of traffic signals. Subsequently, an attention convolution model was introduced to capture the dynamic information changes of key spatial regions and data in the traffic scene through a spatial and temporal attention mechanism. The average vehicle travel time under model control was tested to be 12.4% and 23.3% shorter than the other two methods, respectively. The ablation experiments showed that the model reduced the passing time by about 9.5% and 6.7%. Moreover, the attention is mainly focused on the nodes closer to the location, and the model is more spatially interpretable. The results illustrate that the designed collaborative control of traffic signals at multiple intersections method based on the convolutional model with attention mechanism can enhance the flexibility of traffic management on the basis of improving the efficiency of collaborative control. This offers compelling evidence in favour of building a more clever and effective cooperative traffic signal control system at several crossings.

KEYWORDS

collaborative control of traffic signals; multiple intersections; temporal attention mechanism; convolutional modelling; dynamic information.

1. INTRODUCTION

One of the biggest problems major cities worldwide are facing in the current period of rapid urbanisation is traffic congestion [1]. To maximise traffic flow, minimise vehicle delays and increase the capacity of metropolitan highways, effective traffic signal regulation is essential. Traditional traffic signal control methods are often based on fixed timing schemes or simple inductive adjustments, which are difficult to adapt to the increasingly complex and changing traffic conditions, and are unable to fully exploit the potential efficiency of the traffic system [2-4]. With the booming development of artificial intelligence technology, numerous innovative models and algorithms have emerged in the field of deep learning (DL), which provide new ideas and ways to solve the problem of collaborative control of traffic signals (CCTS) [5]. Among them, the convolutional model shows excellent performance in processing data with spatial structure and can effectively extract spatial features in traffic scenes. However, traffic data not only have spatial characteristics but also contain rich temporal dynamic information. The attention mechanism (AM), as a technical tool that can focus on key information, has achieved remarkable results in many fields in recent years. When working with complicated data, it can dynamically apply weights based on the significance of the data, allowing the model to focus more on the elements that are crucial for making decisions [6, 7]. Combining the AM with the convolutional model can give full play to the advantages of both, capture the spatio-temporal features in the traffic scene more accurately, and then realise more efficient CCTS at multiple intersections (MI) [8]. Hyperparameter optimisation is chosen as the main topic of this study because the selection of hyperparameters is directly related to the performance and application effect of machine learning models. In practical applications, machine learning models often face the problem of data diversity and complexity, and different combinations of hyperparameters can lead to very different results. Therefore, an in-depth study of the influence of hyperparameters can help us better understand the behaviour of models and improve the accuracy and stability of models. In addition, with the wide application of machine learning in various fields, how to efficiently select and adjust hyperparameters has become a practical problem to be solved urgently. The research on this topic has not only theoretical significance, but also practical application value. Through this study, we hope to provide systematic guidance for model optimisation and promote the further development of machine learning technology.

2. RELATED WORKS

As urban traffic congestion increases, it has emerged as a transportation research hotspot to achieve successful CCTS at MI. Numerous academics have conducted related research on CCTS. For efficient traffic control and management at smart city road crossings, Tomar et al. investigated the critical function of real-time traffic signal control technology. The study emphasised the usefulness of regional traffic signal networking by investigating synchronised traffic signal congestion and solutions. The results indicated that synchronised traffic signals were effective in controlling congestion [9]. Gholamhosseinian et al. conducted a comprehensive study of various intersection management methods for heterogeneous interconnected vehicles and proposed an autonomous intersection management approach. The study considered different types of vehicles and intersection objectives to evaluate the parameters of signalised and semi-autonomous intersections. The outcomes revealed that the proposed method was robust in managing intersections [10]. An adaptive traffic signal control system was presented by Miletić et al. to address the issue of everyday congestion in urban areas. The findings demonstrated that the suggested approach might serve as a technical foundation for congestion resolution [11]. A technique to create high-quality ban sign images for vision-based traffic sign identification and recognition issues was put out by Dewi et al. Through the analysis of neural network models with various backbone topologies and feature extractors, the study created synthetic sign images using deep convolutional generative adversarial networks. According to experimental findings, the suggested approach may enhance intersection concatenated traffic sign recognition performance [12]. Liu et al. addressed the issue of traffic control systems in smart cities and proposed a cloud-assisted IoT intelligent transportation system. The study monitored the vehicle flow and controlled the signals by installing IoT sensor-integrated cameras at each traffic signal corner. The results showed that the proposed system was stronger in automatic management [13]. However, most of these studies focused on solutions to specific problems and lacked an integrated and holistic perspective to consider different factors in CCTS.

However, convolutional modelling and AMs have been the subject of extensive research by numerous academics. Soydaner highlighted the important milestones in the progress of AMs in different tasks, looking for ideas to implement attention with neural networks. In this way, the study offered a roadmap for researchers to explore and inspire [14]. For automatic modulation recognition, Lin et al. developed a time-frequency AM based on convolutional neural networks (CNN) to determine whether time, frequency and channel information are more important for neural network modulation recognition. This study assessed the performance of the model and examined the efficacy of the suggested AM. The findings demonstrated that the suggested approach performed better than current AMs [15]. Ramaswamy et al. enhanced the effectiveness of aspect extraction and classification by combining implicit information from a long-term memory model with explicit knowledge from an external database in order to increase sentiment analysis accuracy. According to the evaluation results, the suggested model performed better in both opinion and aspect categorisation [16]. Kavianpour et al. used a DL model based on CNN bi-directional long and short-term memory and AMs to generate an earthquake prediction method. Zero-order hold preprocessing was utilised to predict the maximum magnitude and maximum number of earthquakes in the coming month with minimum error. According to the findings, the approach performs better than alternative approaches in terms of generalisation ability and performance [17]. Alirezazadeh et al. employed a convolutional block attention module to enhance neural network allocation in an effort to gather enough annotated image data to effectively train a DL model for plant disease recognition. To extract more discriminative characteristics and emphasise significant local regions, the study added lightweight attention modules to the neural network framework. According to experimental findings, the AM could successfully raise the pre-trained cellular neural networks' recognition accuracy [18].

At the same time, some scholars have conducted research on the application of deep learning methods in urban road traffic control. Babbar and Bedi proposed a deep learning-based approach to real-time traffic, accident and pothole detection for road traffic control. The method builds semantic relationships by collecting relevant data from Twitter, using multiple word embedding models, and processing them through natural language processing techniques of deep learning. After several pre-processing steps, the proposed classification model has an accuracy of 97% in real-time tweet classification, which is significantly better than existing methods and effectively improves the detection efficiency of road traffic events [19]. Aiming at road traffic control, Gupta and Verma proposed a low-altitude UAV image monitoring and surveillance method based on deep learning technology. The method utilises state-of-the-art object detection models to process aerial images from aerial datasets. To overcome the imbalance in the dataset, an additional 500 images were collected to supplement it. Experimental results show that the proposed method is superior to other models in a number of performance indicators, with an average accuracy of more than 88%, and the detection speed is more than 6 times higher than other models, demonstrating powerful real-time monitoring capabilities [20]. Aiming at road traffic control, Liu et al. proposed a highway traffic congestion detection and evaluation framework based on deep learning technology. The framework uses autoencoders to build a deep learning model, selects traffic congestion indicators to build an index model, and realises real-time detection by classifying and predicting highway traffic environment data. Based on the Shanghai expressway data of the China Traffic Administration Bureau, the experimental results show that the proposed framework can provide accurate and effective solutions for the next generation of expressway traffic management [21]. The Kamal H. team, addressing the issue of excessive vehicle fuel consumption in urban traffic signal control, utilised deep learning technology to conduct adaptive optimisation on the constructed digital twin of the traffic network. During the process, a reward function based on stranded vehicles was designed, and a deep learning model was used to simulate various traffic scenarios and adjust the signal strategy. The results proved that this deep reinforcement learning method successfully reduced fuel consumption [22]. The team led by Wang M. designed an adaptive model to address the real-time implementation challenges of large intersections in urban traffic signal control. During the process, a dual-network architecture is adopted, with two neural networks collaborating to adjust the signal phase and pedestrian clearance time in real time, and an invalid action masking method is applied to meet the spatial constraints of the refuge island. As a result, the experiment shows that the model effectively meets the capacity constraints, proving its practical value [23]. Kumar R.'s team developed an adaptive model using deep learning technology to address the problems of intelligent urban traffic congestion and vehicle classification in urban traffic signal control. During the process, road vehicles were innovatively classified by weight and assigned different weights. Based on experience playback and the training model of the target network, real-time control was achieved in combination with a dedicated short-range communication protocol. The results show that the deep learning model enables the traffic light signals to change adaptively, effectively alleviating congestion [24].

In summary, although the above studies on AMs and convolutional models have achieved some success in their respective fields, most of them focus on single-task scenarios and lack application in complex and dynamic traffic scenarios. In view of this, the study proposes a CCTS at MI approach based on the convolutional model of AM, which combines spatial and temporal convolutional models. This can effectively capture the key information in the traffic data and thus improve the accuracy and efficiency of signal control. The core contribution of the research in urban traffic signal control mainly lies in achieving dynamic perception and adaptive focusing of key congestion nodes and peak hours in multi-intersection traffic flow through a convolutional model that integrates spatial and temporal dual attention mechanisms, solving the bottleneck problem that traditional fixed time series models are difficult to cope with the dynamic changes of complex road networks.

3. DESIGN OF AN AM-BASED METHOD FOR CCTS AT MI

3.1 Analysis and design of MI traffic signal cooperative control parameters

Traffic congestion has grown to be a significant issue affecting urban development and resident mobility due to the ongoing growth of the urban scale and the dramatic rise in the number of motor vehicles. Multiple intersection traffic signals, as a key regulator of traffic flow, have a direct relationship with the cooperative control effect on the operational efficiency of the entire urban transportation network. However, the traffic flow at different intersections presents a complex and variable situation, including significant differences in the traffic flow and pedestrian flow at different time periods, different dates and different seasons. Intersections themselves vary in topology, such as shape, number of lanes and spacing. This makes the trajectory of vehicles and the distribution of conflict points distinctive. At the same time, traffic delays, queue lengths, etc., always reflect the smoothness of traffic. The proportion of various types of vehicles, non-motorised vehicles and pedestrians in the traffic flow further complicates traffic coordination. By carefully analysing these factors, such as traffic flow, intersection structure, traffic delays, traffic composition, timing and other aspects, we can better understand the underlying principles of CCTS. As urban areas continue to grow and the number of motor vehicles increases sharply, traffic congestion has become a major issue affecting both urban development and residents' mobility. The parameter analysis of CCTS at MI is shown in *Figure 1*.

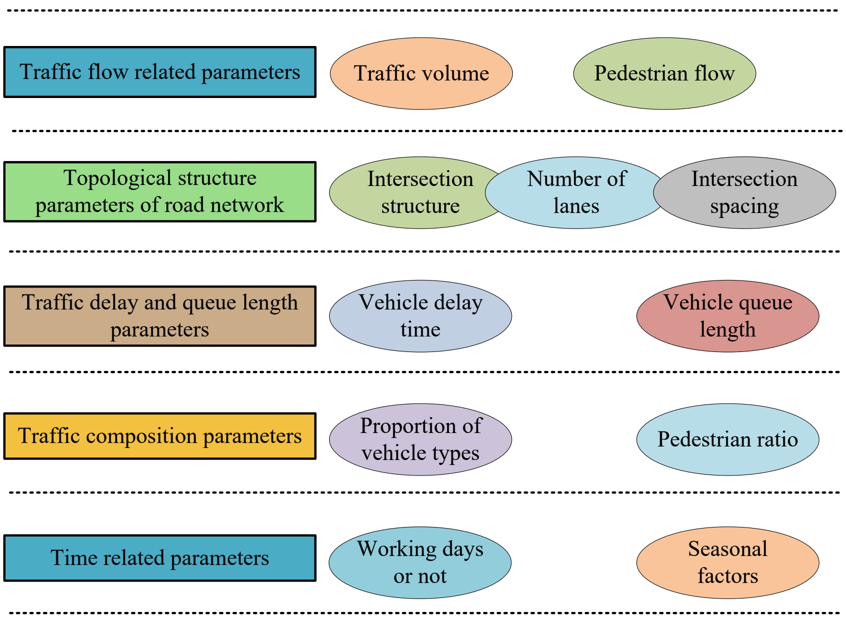


Figure 1 – Analysis of collaborative control parameters for traffic signals at MI

In *Figure 1*, traffic flow is one of the most critical parameters, including the flow of motor vehicles in different directions at each intersection. The study obtains real-time traffic flow data for each lane by setting up inductive coils, video monitoring and other devices at the intersection. Pedestrian flow has a significant impact on CCTS. When counting the pedestrian flow at each intersection, a reasonable pedestrian green light time is considered. This is to ensure that pedestrians can cross the road safely and calmly, but not too long to unduly affect the efficiency of motor vehicles. Different shapes of intersections and the number of lanes at each intersection will affect the trajectory of vehicles and the distribution of conflict points. For intersections with a large number of lanes, finer signal control schemes may be required to coordinate the sequence of vehicles in each lane. The distance between neighbouring intersections is also an important parameter. Parameter analysis revealed its key impact on the performance of collaborative control. In terms of intersection spacing, a closer spacing (such as <200 metres) requires a smaller phase difference (such as -10 to 30 seconds) to reduce vehicle starts and stops and significantly shorten passage time. In terms of traffic flow distribution, the imbalance of directional flow leads to the need for asymmetric green light allocation during peak hours, which will directly affect the queue length. In terms of the conflict matrix, high-severity conflict points require longer clearing times or specific phase sequences, which restrict the passage capacity. The study does not consider the flexible lane function in intersection control for the time being. The intersection passage schematic is shown in *Figure 2*.

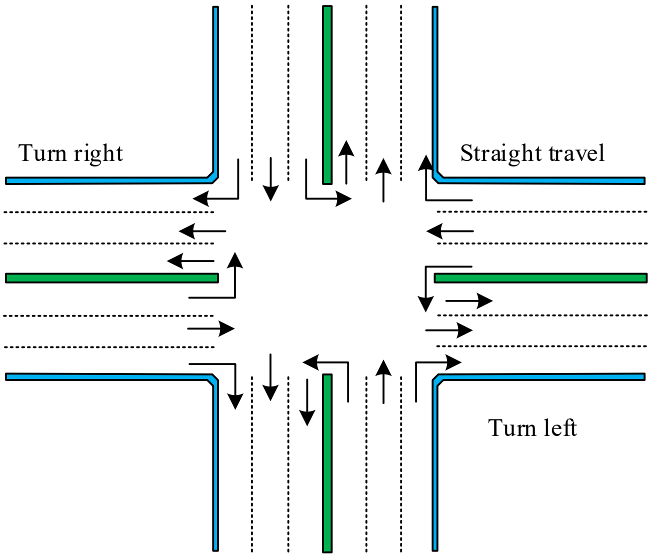


Figure 2 – Schematic diagram of intersection passage

In *Figure 2*, if the intersection spacing is close, the cooperative control of signals needs to be closer to avoid the situation of frequent stopping and starting of vehicles in a short period of time, which affects the smoothness of traffic. The phase difference of the signals can be reasonably set according to the intersection spacing, so that vehicles can pass through MI continuously at a more stable speed [25, 26]. In addition to the intersection topology parameters, traffic movement, waiting queue length parameters, traffic composition parameters and time-related parameters need to be analysed in a focused manner. The vehicle travel time is shown in *Equation 1*.

$$agt = \frac{1}{N_c}\sum_{i=1}^{N_c} t_{i,ed} - t_{i,st} \tag{1}$$

In *Equation 1*, agt represents the average travel time of the vehicle. N_c is the number of vehicles. $t_{i,ed}$ represents the arrival time of vehicle i to the region. $t_{i,st}$ represents the time when the vehicle leaves the region. The waiting queue length is shown in *Equation 2*.

$$queue = \sum_{i=1}^{N_e} q\,ueue_e \tag{2}$$

In *Equation 2*, $queue$ represents the waiting queue length. N_e represents the inbound sections in the road network. $queue_e$ represents the number of waiting vehicles. The vehicle movement conflict matrix in traffic movement is shown in *Figure 3*.

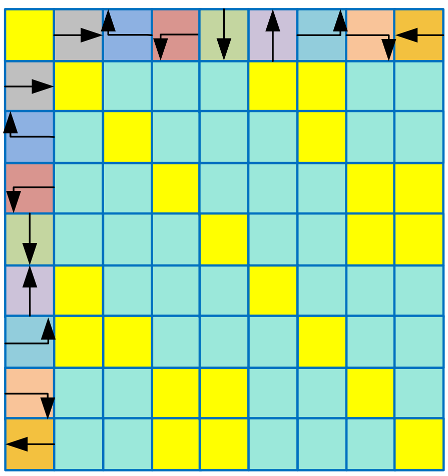


Figure 3 – Vehicle motion conflict matrix

Figure 3 shows that the vehicle motion conflict matrix is a tool used to quantify and analyse the level of potential conflict between different vehicle motions in a traffic scenario. The study is conducted by constructing a matrix structure that arranges combinations of various possible vehicle motion patterns. It also assesses the likelihood and severity of conflict for each pair of combinations, presenting the complex conflict relationships in traffic movements systematically and intuitively. The dimensions of the vehicle motion conflict matrix are usually determined by the major vehicle motion types involved in the traffic scenario. Common vehicle motion types include left turn, right turn and straight ahead [27, 28]. Each element in the matrix represents the information related to a conflict between a vehicle of one motion type and a vehicle of another motion type. By constructing and analysing the vehicle motion conflict matrix, the potential conflict points in traffic motion can be more clearly understood. This can provide valuable references for traffic planning, signal control and traffic facility layout. During the actual implementation of the model, real-time lane-level traffic flow data are obtained through induction coils and video surveillance. After standardisation processing, a time series is formed and input into the time convolution module of the model. The intersection topology is transformed into an adjacency matrix for modelling spatial dependencies in graph convolutional networks. For instance, adjacent intersections with a spacing of less than 200 metres are assigned higher weights in the adjacency matrix to drive the optimisation of phase differences. The vehicle motion conflict matrix is encoded as a three-dimensional tensor, and its element values represent the conflict intensity level, serving as the input feature of the spatial attention module.

3.2 Signal light cooperative control algorithm based on spatial AM and convolutional modelling

Following a comprehensive examination of the variables associated with the CCTS at MI, AMs and convolutional models are incorporated to achieve optimal and precise CCTS at MI. The spatial AM can focus on key spatial regions in the traffic scene and accurately capture important information, such as different intersections, different lanes, and spatial relationships between vehicles and pedestrians. It can quickly identify which areas are the source of traffic congestion, where there is a potential risk of conflict, etc. [29, 30]. Moreover, the convolutional model, with its powerful feature extraction capability, can deeply mine a large amount of traffic data and effectively extract the traffic patterns and laws hidden in the data. Whether it is the periodic change of traffic flow or the driving characteristics of different types of vehicles at intersections, etc., they can be accurately analysed [31]. The spatial attention matrix is shown in *Equation 3*.

$$S=Vs\cdot\sigma(GtWsGtT+bs) \tag{3}$$

In *Equation 3*, S represents the spatial attention matrix. $V_s, b_s \in \quad R^{N\times N}$ and $W_s \in R^{C_{r-1}\times C_{r-1}}$ represent parameters that can be learned by training. σ represents the activation function. $G_t^{(r-1)} \in R^{N\times C_{\tau-1}}$ represents the observation information of all nodes in t time. The node attention weights are shown in *Equation 4*.

$$S_{i,j}^{'} = \frac{exp(S_{i,j})}{\sum_{j=1}^{N} e\,xp(S_{i,j})} \tag{4}$$

In *Equation 4*, $S_{i,j}^{'}$ is the attention weight of two nodes, i and j. $S_{i,j}$ represents the degree of relevance of the two nodes. N represents the number of nodes. To realise the graph convolution (GC) based on the spectral domain, the study needs to represent the graph as a Laplace matrix form, as shown in *Equation 5*.

$$L = I_N - D^{-\frac{1}{2}}AD^{-\frac{1}{2}} \in R^{N\times N} \tag{5}$$

In *Equation 5*, L is the Laplace matrix. I_N is the unit matrix. D is the degree matrix. A represents the adjacency matrix. Split-cycle-offset optimisation technique (SCOOT) is introduced to be applied to the dynamic optimal control of urban road traffic signals. The process is shown in *Figure 4*.

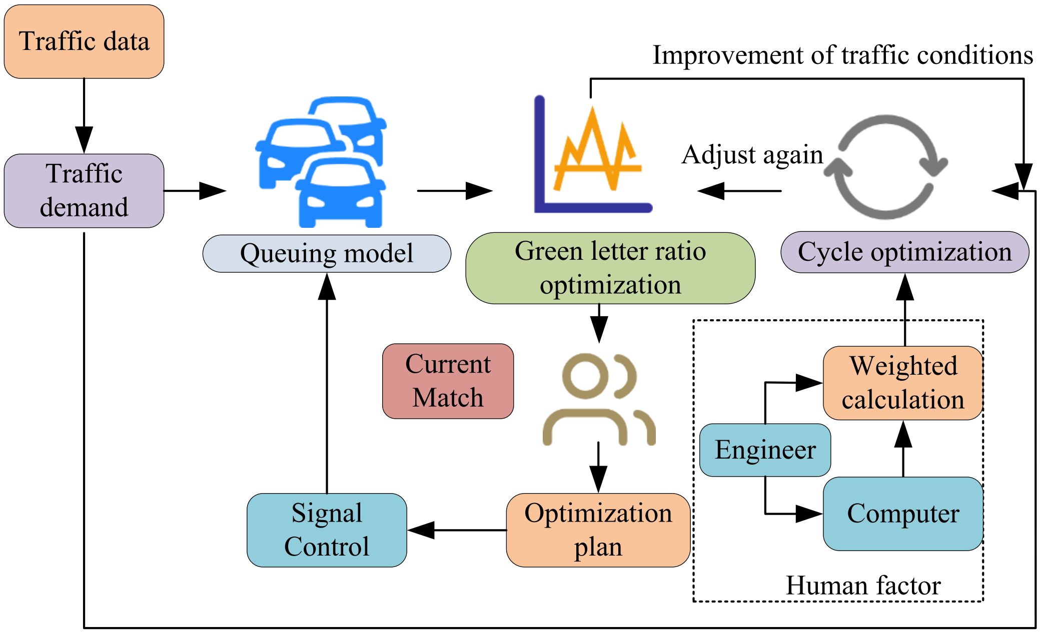


Figure 4 – SCOOT control system operation process

Figure 4 shows that the SCOOT system is capable of adjusting the signal control parameters in real time based on the latest conditions of the traffic flow. After adjustment, the changes in traffic flow continue to be monitored through the vehicle detection subsystem in an attempt to assess the effectiveness of the adjustment. If it is found that the adjusted traffic condition is not significantly improved, the system will analyse and adjust again. Afterwards, the control effect of traffic signals is continuously optimised [32, 33]. Traffic intersection signal control can be divided into two categories: single-point and multi-point, as shown in *Figure 5*.

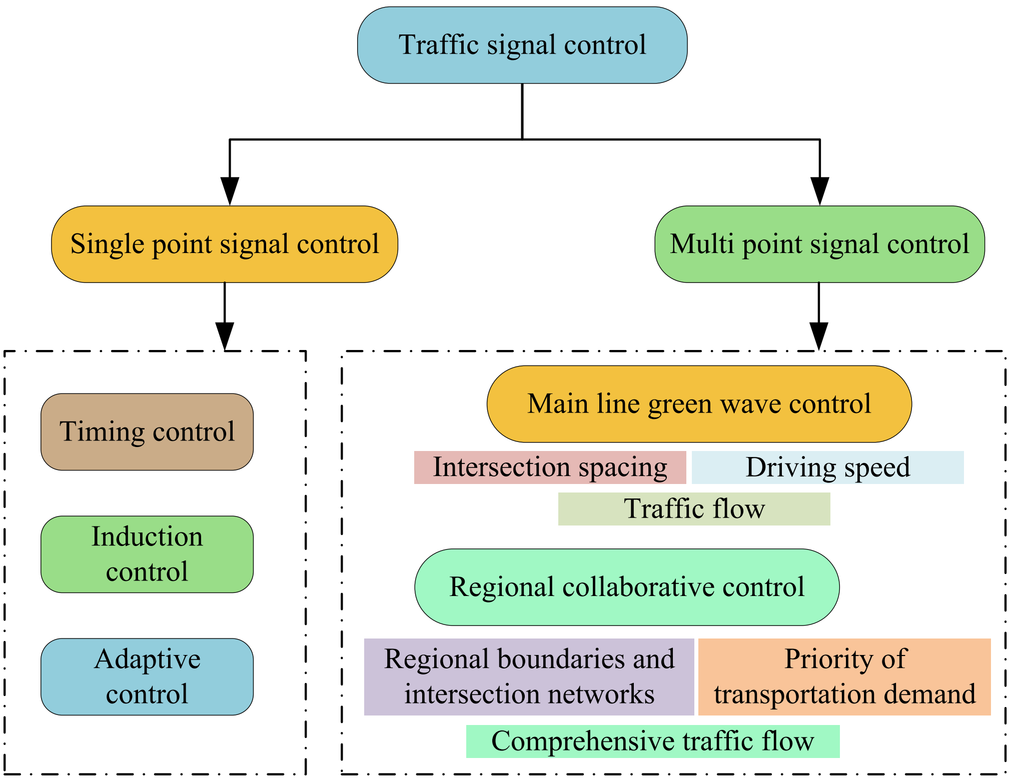


Figure 5 – Classification of traffic intersection signal control

In *Figure 5*, single-point control focuses on the traffic signal deployment of independent intersections, based on the intersection's own traffic flow characteristics to determine the timing scheme. Multi-point control focuses on the coordinated operation of MI in the region. It realises more efficient traffic diversion by considering the correlation between each intersection and the overall traffic situation. On this basis, the introduction of node degree can further deepen the understanding of signal control at traffic intersections. The node degree formula is shown in *Equation 6*.

$$Dii{=}jAij \tag{6}$$

In *Equation 6*, D_{ii} represents the degree matrix of node i. A_{ij} represents the adjacency matrix between nodes i and j while satisfying *Equation 7*.

$$Lu = \lambda u \tag{7}$$

In *Equation 7*, λ represents the eigenvalue. u represents the eigenvectors. *Equation 8* is obtained by decomposing the eigenvalues of the Laplace matrix.

$$L = U\Lambda U^T \tag{8}$$

In *Equation 8*, $\Lambda = diag\{[\lambda_0, \dots, \lambda_{N-1}]\} \in R^N$ represents the diagonal matrix. U represents the combination of column vectors. The GC operation in the convolution operation is shown in *Equation 9.*

$$g_\theta * x = g_\theta(L)x = g_\theta(U\Lambda U^T)x = Ug_\theta(\Lambda)U^T x \tag{9}$$

In *Equation 9*, g_θ represents the convolution kernel. x is the graph signal of the traffic road network. $*$ is the GC operation. Chebyshev polynomials are then used to regulate the communication information between nodes. The GC formula is expressed as *Equation 10*.

$$g_\theta * x = g_\theta(\boldsymbol{L})x \approx \sum_{k=0}^{K-1} \theta_k \left(T_k\left(\frac{2}{\lambda_{max}}L - I_N\right) \odot \boldsymbol{S}'\right)x \tag{10}$$

In *Equation 10*, θ_k represents the coefficient vector. T_k represents Chebyshev polynomials. K is the number of layers of neighbouring nodes. λ_{max} represents the maximum eigenvalue of the Laplace matrix. $\odot$ represents the Hadamard product. $\boldsymbol{S}'$ represents the spatial attention matrix.

3.3 Signal light cooperative control algorithm based on temporal AM and convolutional modelling

Following an in-depth examination of the benefits of spatial AMs and convolutional models in discerning the spatial attributes of traffic scenes, this study delves into the temporal characteristics of traffic data. Then it puts forth a CCTS algorithm that is based on a temporal AM and a convolutional model. The study firstly expresses the historical node information weights through the AM as shown in *Equation 11*.

$$E = V_E \cdot \sigma(G^{(r-1)}W_E(G^{(r-1)})^T + b_E) \tag{11}$$

In *Equation 11*, E represents the temporal attention matrix. $V_E, b_E \in R^{T_{r-1}\times T_{r-1}}, W_E \in R^{C_r \times C_r}$ represents the parameters that can be trained. $G^{(r-1)}$ represents observation information. T represents the time series. The activation function is used to generate *Equation 12*.

$$E'_{ij} = \frac{exp(E_{ij})}{\sum_{j=1}^{T} e\,xp(E_{ij})} \tag{12}$$

In *Equation 12*, $E_{ij}^{'}$ represents the degree of correlation between the two information slices in the same time dimension. Then the historical information in the temporal attention matrix is fused to obtain *Equation 13*.

$$\hat{G} = G^{(r-1)} \times E^{'} = (g_\theta * G_1, g_\theta * G_2, \ldots . g_\theta * G_T)E^{'} \in R^{N \times C_r \times T} \tag{13}$$

In *Equation 13*, $\hat{G}$ is the collection of historical information. The capacity of the intersection road can be expressed in terms of flow rate as shown in *Figure 6*.

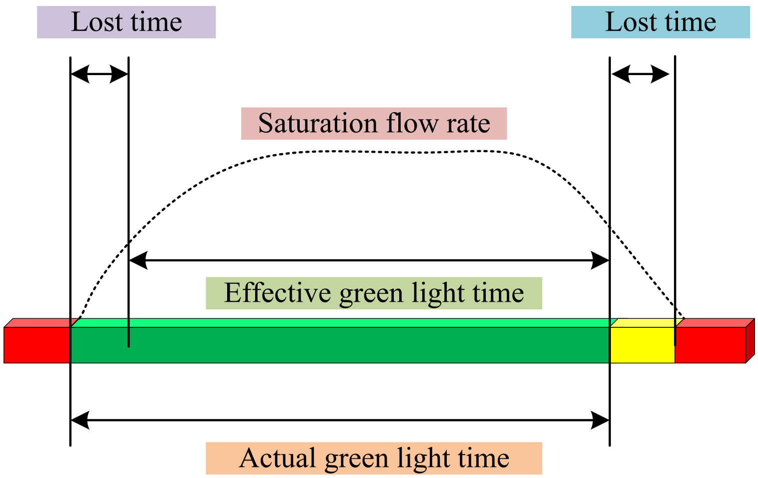


Figure 6 – Flow rate chart

In *Figure 6*, accurate calculation and assessment of intersection capacity are crucial for optimising traffic signal control in practical traffic engineering. The study can dynamically adjust the allocation of green light time according to the characteristics of traffic flow in different time periods. During peak hours, an appropriate increase in the green light time in the direction of higher saturation flow rate can effectively relieve traffic congestion. To improve traffic flow through intersections during peak hours, a more balanced green time allocation strategy can be adopted. In designing standard convolutional modules based on a temporal attention matrix, the time-series data must first be pre-processed. The temporal attention matrix is then computed, typically using methods such as the dot-product attention mechanism (AM), to determine the correlation weights between different time steps. These weights are subsequently applied to the original data to emphasise the most important time steps. Finally, a standard convolution operation is performed on the weighted data to define an appropriate convolution kernel for further processing. The standard convolution module is shown in *Equation 14*.

$$\hat{V}_h = ReLU\left(\Phi * \left(ReLU\left(g_\theta * \hat{G}\right)\right)\right) \tag{14}$$

In *Equation 14*, $\hat{V}_h$ represents the standard convolution module. $ReLU$ represents the activation function. Φ represents the standard convolution kernel in the time dimension. $*$ represents the standard convolution operation. The standard convolutional module is then fed into the decision module, and the policy is updated by minimising the loss function and iterating continuously, as shown in *Equation 15*.

$$L(\theta_n) = \frac{1}{T}\sum_{t=1}^{t=T}\sum_{i=1}^{I=N}\left(r_i^t + \gamma \max_{a^{'}} Q(o_i^{t^{'}}; a_i^{t^{'}}, \theta_{n-1}) - Q(o_i^t; a_i^t, \theta_n)\right)^2 \tag{15}$$

In *Equation 15*, $L(\theta_n)$ represents the update strategy. T represents the time interval. $Q(o_i^t; a_i^t, \theta_n)$ represents an independent learning algorithm. n is the quantity of iterations. N is the quantity of intersections in the traffic network. The structure of the finally obtained attention-based convolutional model coordination control algorithm is shown in *Figure 7*.

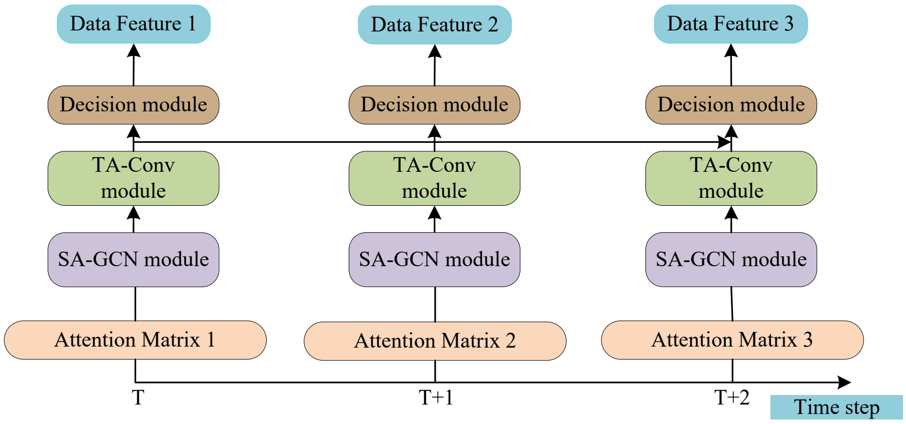


Figure 7 – Attention-based convolutional model coordination control algorithm structure

Figure 7 shows that the algorithm model consists of three modules. Among them, the decision module, as the core command centre of the whole model, assumes the important responsibility of synthesising information from various parties and making key decisions. It receives all kinds of traffic-related data features extracted and analysed from the convolutional model based on temporal and spatial AMs, and deeply integrates and weighs this information through the built-in intelligent algorithm. The convolution model based on temporal AM mainly focuses on the feature capture and analysis of traffic data in the time dimension. It is able to keenly perceive the fluctuation and change rules of traffic flow in different time periods, and through the in-depth learning of historical traffic data, the model can clearly identify the cyclical change characteristics of various parameters. The convolution model based on the spatial AM accurately identifies key spatial nodes by deeply mining the spatial layout and the spatial relationship between the elements in the traffic scene. It can reduce the traffic problems caused by irrational spatial layout or the inability to grasp the spatial relationship. In terms of time parameters, the research takes peak hour volume (PHV) as the core time dimension indicator, which directly reflects the load characteristics of the road network during critical periods and always reflects the degree of traffic smoothness. Historical traffic data integrates different time periods. In particular, the peak-hour traffic volume is input as an independent time series feature into the time convolution module, as well as the traffic patterns of dates and seasons to capture dynamic spatiotemporal features.

Since the traffic scene is mainly affected by traffic flow, intersection topology, vehicle distribution and other spatial factors, it is sufficient to use CNN for spatial feature extraction. These spatial factors can be efficiently captured by CNNS without the need for additional time-based methods. The collaborative control model proposed in this paper realises the dynamic optimisation of multi-intersection signal control through the end-to-end convolutional neural network architecture, instead of using the spatial feature extraction capability of CNN as an auxiliary tool of traditional control algorithms. Specifically, the model takes the spatio-temporal data of traffic flow (including vehicle motion conflict matrix, nodal degree matrix and historical flow series) as input, and focuses on key areas and time segments respectively through a spatio-temporal dual attention mechanism. The spatial convolutional module uses the graph convolutional network to model the intersection topological relationship, and extracts the spatial features such as the conflict intensity between lanes and the correlation degree of adjacent intersections. The time convolution module captures periodic and burst fluctuation patterns of traffic through one-dimensional extended convolution. After the two types of features are weighted by attention, the signal phase duration and green band parameters are generated by the fully connected decision module, and the closed-loop optimisation is completed by the real-time feedback mechanism of the SCOOT system.

Hyperparameter selection is an important part of machine learning model optimisation. Hyperparameters are parameters that need to be specified manually during model training, rather than parameters that are automatically learned through training data. The selection of different hyperparameters can significantly affect the performance of the model. For example, in a neural network, the learning rate determines the amplitude of each updated weight, and too high a learning rate may cause instability in the model, while too low a learning rate may make the training process too slow. In addition, regularisation parameters can control the complexity of the model; proper regularisation can prevent overfitting, but too strong regularisation may lead to underfitting. Therefore, when selecting hyperparameters, systematic tuning is studied through techniques such as cross-validation to find the best combination of parameters, thereby improving the accuracy and generalisation ability of the model.

4. SIMULATION EXPERIMENTS ON COOPERATIVE CONTROL OF SIGNALS BASED ON AM BASED CONVOLUTIONAL MODEL

4.1 Comparative experiment of cooperative control algorithms for traffic signals at MI

To analyse the performance of the CCTS at the MI algorithm based on the convolutional model with AM, the study abbreviates the convolutional model based on AM as A-C. Traffic movements are simulated using the CityFlow simulation simulator. The traffic network diagram in the CityFlow simulator is shown in *Figure 8*.

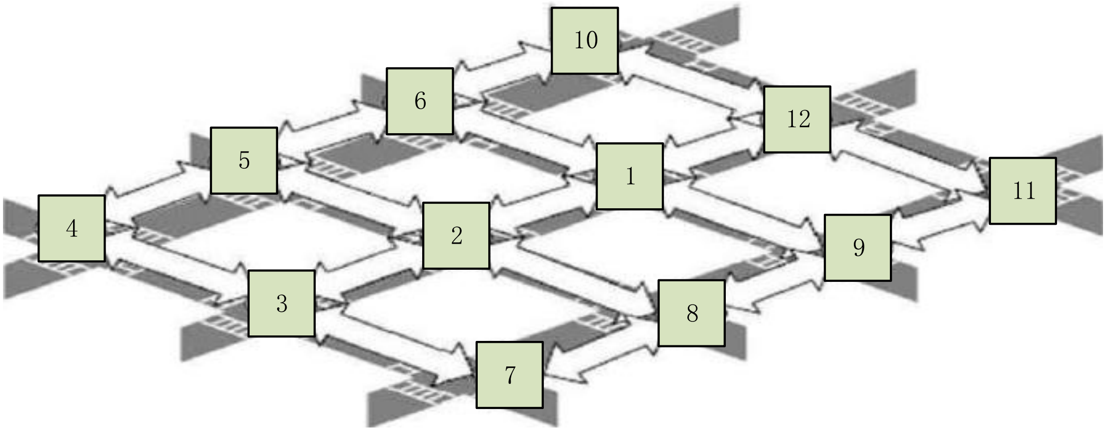


Figure 8 – CityFlow simulator traffic network

The route is a traffic scenario from place D of a city to various other places. The number of turning intersections from place D to various places is set as shown in *Table 1*.

Table 1 – The number of turning intersections from D to various places

Destination	Index	Number of through movements	Number of left turns	Number of right turns	Total number of intersections
E	Number of intersections	231	162	143	536
E	Rate	43%	30%	27%	100%
F	Number of intersections	2	2	1	5
F	Rate	40%	40%	25%	100%
G	Number of intersections	4	1	1	6
G	Rate	66.6%	16.6%	16.6%	100%
Average proportion		49.9%	28.9%	22.8%	100%

The training rounds are set to 100, the time convolution kernel is 1*3, the time slice length is 5, and the learning rate is 0.01. In order to support the training and verification of the convolution model based on an attention mechanism in the cooperative control of traffic signals at multi-intersections, samples were obtained through multi-source data acquisition and simulation scenario construction in the research experiment. Data collection was carried out on the actual urban road network corresponding to the simulated routes from point D to Points E, F and G, including the intersections and connections involved in these roads. The CityFlow simulation platform builds a road network based on the morning rush hour commuting corridors from point D to areas E, F and G in a certain city. It contains a total of 547 signal intersections, with a simulation duration of 3,600 seconds and a time granularity of 1 second. The road network geometry is calibrated based on the measured coil data from 7:00 to 9:00 in May 2024. The saturation flow rate of the entrance lanes at each intersection is set at 1700 to 1800 pcu/h. The OD distribution of traffic demand is generated in the proportion of 49.9% for straight driving, 28.9% for left turning and 22.8% for right turning. The arrival of vehicles follows the Poisson distribution. The initial queue length is taken from the measured average of the morning rush hour.

The speed limit on the section is 60 km/h. The signal control scheme adopts a unified 120-second four-phase timing control. First, the northbound direction is allowed to pass for 36 seconds, followed by the southbound direction for 35 seconds, the eastbound direction for 31 seconds, and the westbound direction for 30 seconds. The pedestrian phase is fixed for 25 seconds. The green signal ratio in each direction is dynamically fine-tuned based on the measured flow weight to ensure that the simulation scene is consistent with the real morning rush hour traffic conditions. The traffic flow in this area shows a multi-centre aggregation feature, mainly distributed in the core business district, major traffic arteries and radial roads connecting various functional areas. The traffic demand during the morning and evening rush hours is significantly higher than that during the off-peak hours, resulting in a distinct tidal phenomenon. Local congestion often occurs at some key nodes and around public facilities such as schools and hospitals during specific periods. Historical traffic data integrates traffic patterns of different time periods, dates and seasons to capture dynamic spatio-temporal characteristics. The Max Pressure traffic control method and Colight control method are compared with the A-C model control method. Vehicle travel time at MI under different control methods is evaluated. The Max Pressure control method is a classic theory-driven approach. Its core lies in balancing the pressure of the road network and is widely adopted in both basic research and practical applications. The Colight control method represents the recent multi-intersection collaborative signal control framework based on reinforcement learning, optimising the collaborative effect through a distributed decision-making mechanism. Comparing these two methods can reflect the advantages and disadvantages of the research methods. *Figure 9* displays the outcomes of the experiment.

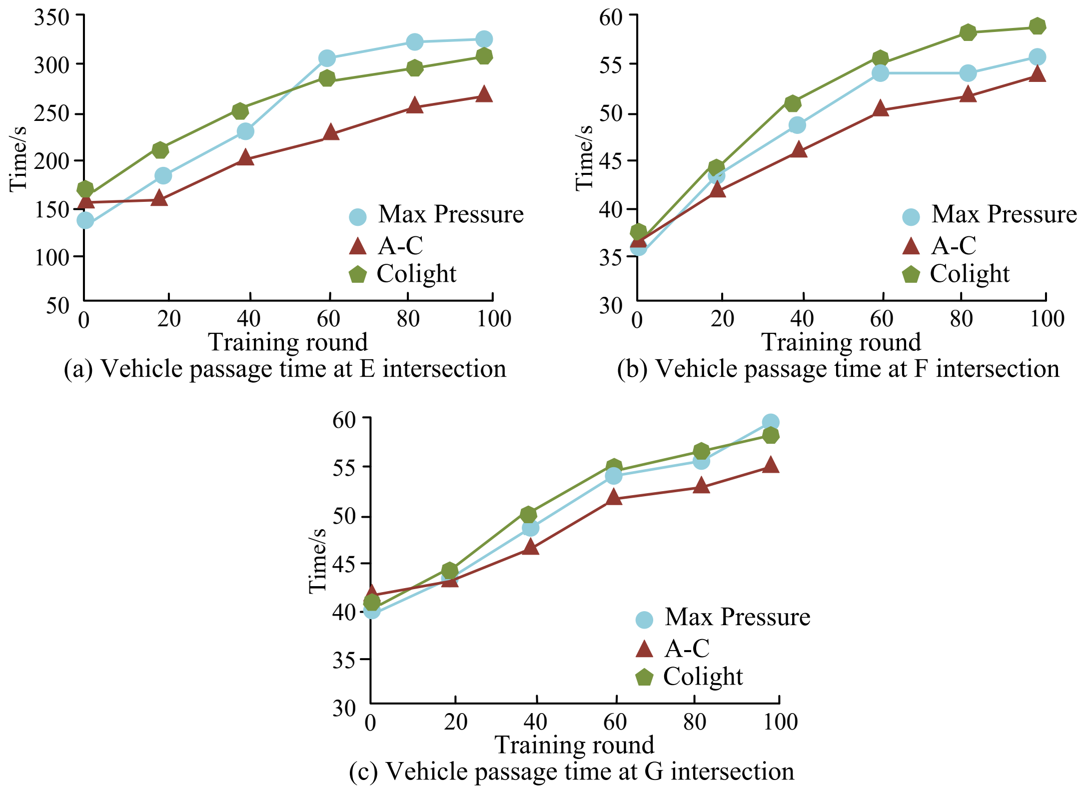
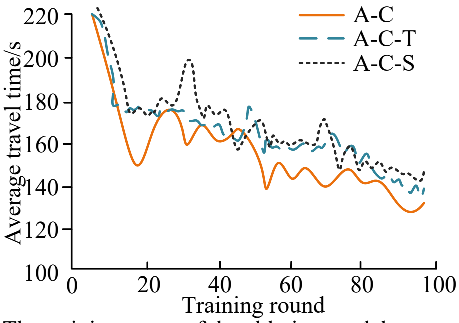
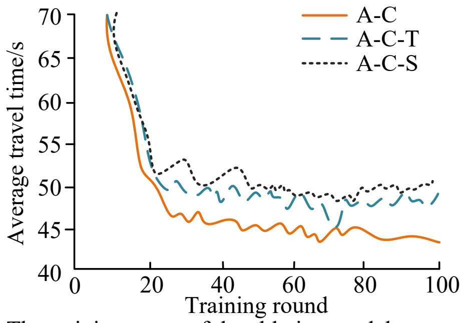


Figure 9 – Vehicle passage time at MI under different control methods

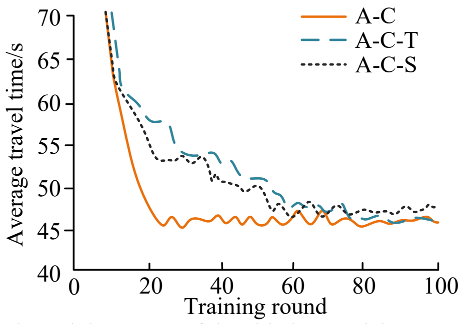
In *Figure 9*, the A-C algorithm demonstrates excellent performance in controlling vehicle travel time. *Figure 9a* shows that when the number of training rounds is 100, the A-C algorithm performs well in terms of vehicle travel time on section E, reaching a minimum of about 240 s. This fully demonstrates its powerful ability in reducing the travel time of vehicles at the E intersection. The vehicle travel time of the Max Pressure and Colight methods is significantly higher under the corresponding number of wheels, reaching about 300 s and 290 s, respectively, an increase of 25% and 27% compared to the A-C algorithm. In *Figure 9b*, the A-C algorithm also shows significantly better performance in terms of vehicle travel time on section F, with a vehicle travel time of 53 s at the end of training. The final vehicle travel time for the Max Pressure and Colight methods is approximately 55 s and 58 s, respectively. The A-C algorithm reduced the average vehicle travel time on section F by 12.4% and 23.3% respectively, compared to the other two methods. In *Figure 9c*, the A-C algorithm also shows relatively lower and smoother performance in terms of vehicle travel time on the G section, ultimately reaching 52 s. At this point, the final vehicle travel time for the Max Pressure and Colight methods is 56 s and 55 s, respectively. The A-C algorithm reduces travel time by approximately 6.2% and 5.1% compared to these two algorithms. The results of the study show that the A-C algorithm shows significant advantages in vehicle travel time control at several intersections, which can provide a more efficient and reliable scheme for traffic control and effectively reduce the vehicle travel time. Through the ablation experiment, the average travel time of vehicles under the control of the A-C algorithm, A-C-T algorithm without the time module, and A-C-S algorithm without the spatial module are compared, as shown in *Figure 10*. The A-C-T algorithm, without the time module, has removed the module in its original design that specifically handles temporal dynamics, while retaining the spatial feature processing module as well as modules A and C. Correspondingly, the A-C-S algorithm, without spatial module removes the module specifically for handling spatial dependencies, while retaining the temporal feature processing module as well as modules A and C.



(a) The training curve of the ablation model on section E



(b) The training curve of the ablation model on section F



(c) The training curve of the ablation model on section G

Figure 10 – The average travel time of vehicles in the ablation experiment

In *Figure 10*, the A-C algorithm clearly performs better in controlling the vehicle passing time in the ablation experiment. In *Figure 10a*, when the number of training rounds is 100, the A-C algorithm performs better on the vehicle passing time of the E section, reaching a minimum of about 140 s. Moreover, its vehicle passing time decreases steeply when the number of training rounds is less, showing excellent control. The vehicle passing time of the A-C-T and A-C-S methods is significantly higher at the corresponding number of rounds, reaching about 150 s and 160 s, respectively. This is an increase of 12.2% and 14.3%, respectively, compared to the A-C algorithm. In *Figure 10b*, the A-C algorithm also performs significantly better in terms of vehicle passing time in section F, with a vehicle passing time of 40 at the end of training. The final vehicle passing times for the A-C-T and A-C-S methods are about 45 s and 50 s, respectively. The average vehicle passing time of the A-C algorithm in section F is shorter than that of the other two methods by 11.1% and 20%, respectively. In *Figure 10c*, the A-C algorithm also performs relatively lower and smoother in terms of vehicle passing time on the G roadway section, finally reaching 45. At this time, the final vehicle passing time of the A-C-T and A-C-S methods is 48 and 49, respectively. The A-C algorithm reduces the passing time by about 9.5% and 6.7% compared to these two algorithms. The results of the study show that the A-C algorithm can divert the traffic flow more efficiently and reduce the vehicle's retention time on the roadway section, which in turn reduces the traffic congestion. The average speeds of cars using various control schemes are then contrasted. The results are shown in *Figure 11*.

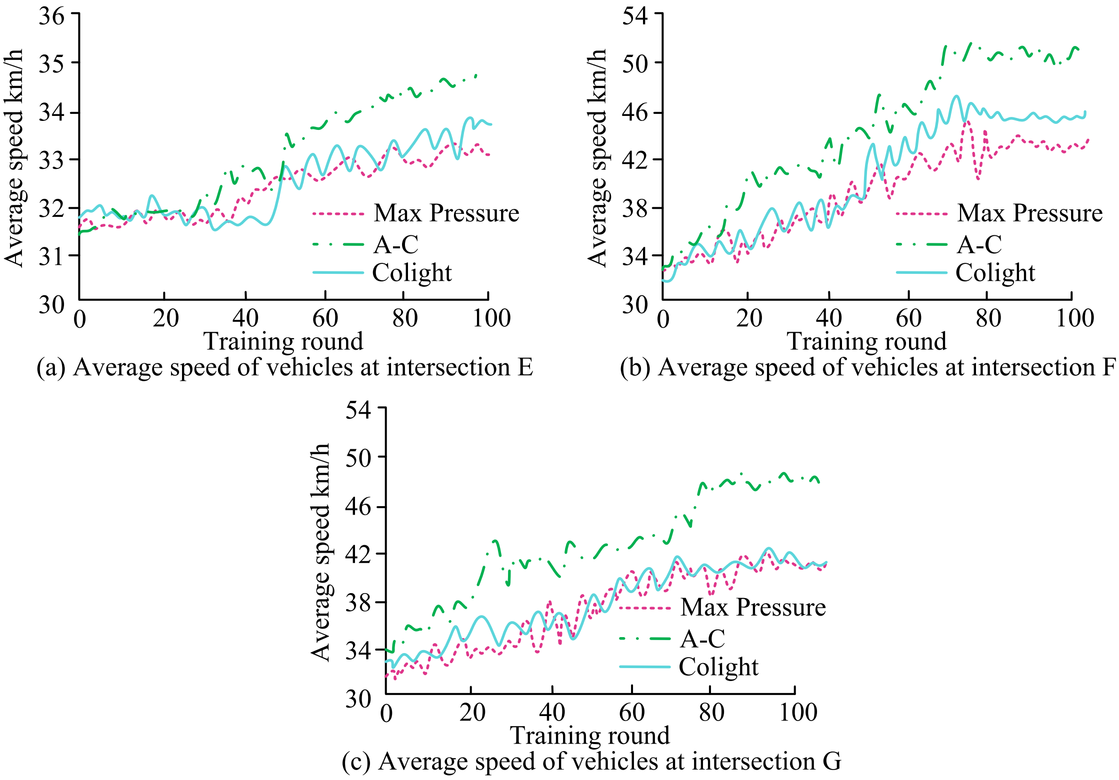


Figure 11 – Average speed of vehicles under different control methods

Additionally, *Figure 11* displays variations in the average speed performance of vehicles under various control schemes. In *Figure 11a*, on the E section, the average speed of vehicles under Max Pressure control slowly increases with the increase of training rounds, and the final average speed of vehicles is 32 km/h. The upward trend of Colight is similar to the Max Pressure method, with a final speed of 33 km/h. The average speed curve position of the vehicle in the A-C algorithm model is significantly higher than the other two, and the final speed is 34 km/h. The average speed increased by 15.2% and 11.4% respectively, compared to Max Pressure and Colight. In *Figure 11b*, on the F section, the average speed of vehicles under Max Pressure control finally reached 42 km/h, while Colight reached 45 km/h. The A-C algorithm has an average vehicle speed 21.3% and 16.7% higher than these two methods, respectively. In *Figure 11c*, on the G section, the average speed of vehicles controlled by the A-C algorithm is still the highest, reaching 47 km/h, which is 23.2% and 22.1% higher than Max Pressure and Colight, respectively. The results of the study show the effectiveness and reliability of the A-C algorithm in traffic control. In real-world traffic scenarios, higher average vehicle speeds mean smoother traffic flow, and vehicles are able to pass through the roadway more quickly, reducing travel time. This is essential for lessening traffic congestion. Regardless of the road segment, the A-C algorithm is able to optimise traffic signal control so that vehicles can travel at a better speed.

4.2 Performance experiment of cooperative control algorithm for traffic signals at MI

To further analyse the performance of the CCTS at the MI algorithm based on the convolutional model of the AM, the study continues to use the datasets of MI G, E and F. The study is conducted using the data sets of the G, E and F intersections. The temporal performance, module validity, spatial dependence and green wave analysis of the A-C algorithm model are analysed. The training time of the Max Pressure traffic control method, Colight control method, A-C-T algorithm and A-C-S algorithm from the ablation experiments is compared with the training time of the A-C algorithm for E and F intersection datasets, respectively. *Figure 12* displays the findings of the time performance analysis.

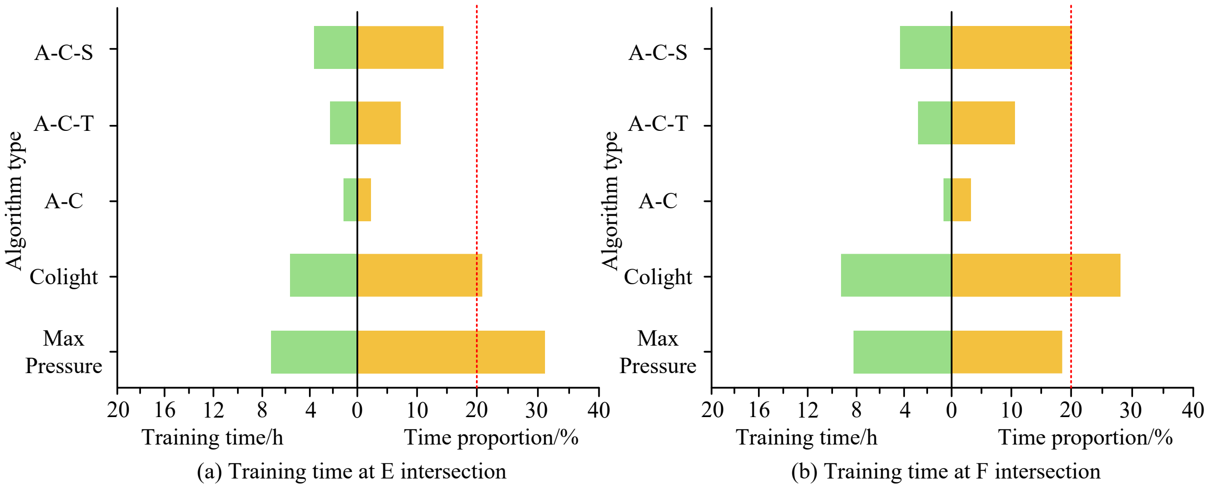


Figure 12 – Time performance analysis results

In *Figure 12*, each algorithm's training time performance varies significantly depending on the road stretch. In *Figure 12a*, on the E section, Max Pressure has the longest training time, reaching nearly 8 hours, while accounting for over 20%. The next is Colight, with training time accounting for over 20%. The training time of A-C-S is slightly longer than A-C-T. A-C has the shortest training time, only 2 hours. In *Figure 12b*, on the F section, Colight has the longest training duration, exceeding Max Pressure and accounting for over 20% of the total. Next is A-C-S, which accounts for just 20% of the training time. Furthermore, there are Max Pressure and A-C-T. The results of the study show that the A-C algorithm has an obvious advantage in terms of training time. In the tests on different road sections, the A-C algorithm is always able to complete the training in the shortest time. It indicates that the algorithm has higher efficiency in practical applications and can provide a faster and more accurate solution for traffic signal control. Next, the impact of module count on average journey time is examined. The results are shown in *Figure 13*.

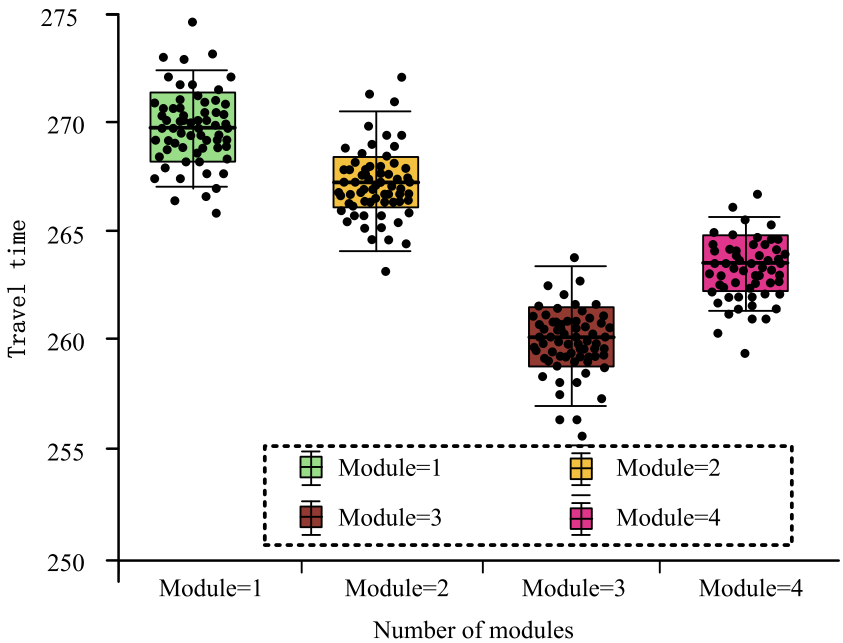


Figure 13 – The impact of the number of testing modules on the average travel time

In *Figure 13*, as the number of stacked modules varies, the average travel time of the car also shows a linear change. When there is one module, the travel time is mostly between 265 and 275 s, which is relatively long. When the number of modules is 2, the travel time is between 263 and 273 s, which is slightly lower compared to those with fewer modules. When the number of modules is 3, the travel time is between 255 and 263 s, and the time decreases again. Moreover, when the number of modules increases again, the travel time actually increases. The research results indicate that there is an optimal number of modules that minimises the average travel time of the car, which is 3 modules. Then we analyse the spatial dependencies of the model. In the road network, 11 nodes are chosen, and the influence coefficients of each node on the same node are computed independently. *Table 2* displays the results of the computation.

Table 2 – Influence coefficient of different nodes on the same node

Node name	Node 1	Node 2	Node 3	Node 4	Node 5	Node 6	Node 7	Node 8	Node 9	Node 10
Influence coefficient	0.194%	0.173%	0.124%	0.118%	0.102%	0.095%	0.054%	0.041%	0.033%	0.012%
	0.205%	0.184%	0.114%	0.108%	0.101%	0.088%	0.062%	0.051%	0.027%	0.010%
Average impact coefficient	0.1995%	0.1785%	0.119%	0.113%	0.1015%	0.0915%	0.058%	0.046%	0.03%	0.011%

Table 2 shows that the attention is mainly focused on the nodes that are closer to the location, and therefore the model is more spatially interpretable. Finally, the green wave analysis is performed for 4 nodes as shown in *Figure 14*.

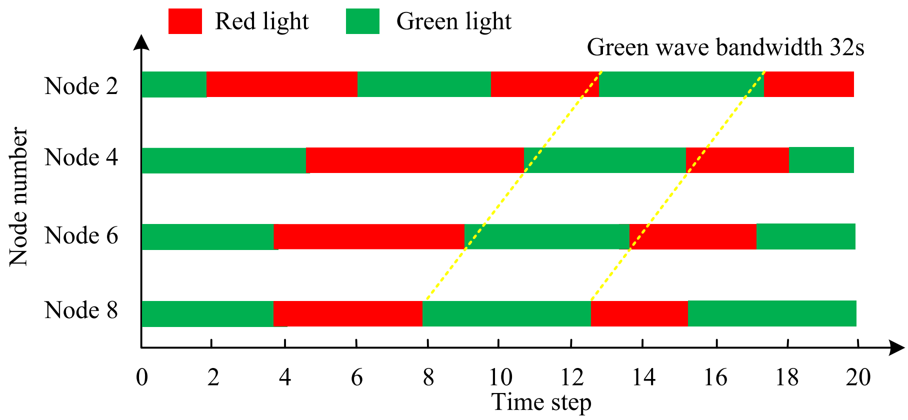


Figure 14 – Green wave analysis of 4 nodes

In *Figure 14*, the common green wave period of the four nodes is 114 s in 20 time steps. The green wave bandwidth is 32 s when the average travelling speed of the green wave is 55 km/h. It shows that the vehicles passing in the last few seconds have a high probability of being able to catch up with the green wave of the next intersection, which proves once again the effectiveness of the A-C model in coordinating and controlling signals.

5. CONCLUSION

Aiming at the problem of CCTS at MI, the study firstly analysed the layout structure of the multi-intersection traffic scene. It detailed the topology of each intersection, lane settings and inter-intersection connectivity, and analysed the business process of traffic signal control. To optimise the accuracy of traffic signal control, the study introduced the A-C model and used its spatial AM to accurately focus on the key areas in the traffic scene. Meanwhile, the temporal AM was used to effectively capture the dynamic change characteristics of traffic data in the time dimension, so as to improve the accuracy of traffic signal timing and phase adjustment. The outcomes revealed that the A-C algorithm reduced the vehicle travel time by 25% and 27% compared to the other two control algorithms, respectively. The average vehicle travel time in section F was reduced by 11.1% and 20% compared to the other two methods, respectively. Meanwhile, the average speed increased by 15.2% and 11.4% over Max Pressure and Colight, respectively. The results indicated that the control method proposed in the study could effectively improve the accuracy of traffic signal control and the efficiency of traffic flow operation. However, the study did not fully consider the impact of special circumstances on the control of traffic signals. Consequently, a subsequent in-depth investigation will be carried out for these special circumstances. The objective is to achieve a more perfect and significant optimisation effect of CCTS.

FUNDINGS

The research is supported by: China Railway Fourth Survey and Design Institute Group Co., Ltd. Research and Application of Key Technologies for Semi rigid and Semi flexible Pavement Structure of Rubber Wheel Track in Urban Rail Transit (No. 2021K080); China Railway Fourth Survey and Design Institute Group Co., Ltd. Research on the Rapid Detection Method and Key Technologies of Road Ruts in Intelligent Rail Express System (No. KY2023063S); China Railway Fourth Survey and Design Institute Group Co., Ltd. An electronic guided rubber wheel system based on infusion type semi flexible material for anti rutting pavement structure and construction method (No. 2022C10); New-type Rail Transit Design and Research Institute of China Railway Siyuan Survey and Design Group C., Ltd. Yibin ART T1 and T4 Line Integration Operation Project (No. 2024KC001).

REFERENCE

[1] Guo MH, et al. Attention mechanisms in computer vision: A survey. *Comput Vis Media*. 2022;8(3):331-368. DOI: 10.1007/s41095-022-0271-y.

[2] Pei Y, et al. FCN-attention: A deep learning UWB NLOS/LOS classification algorithm using fully convolution neural network with self-attention mechanism. *Geo-Spatial Inf Sci*. 2024;27(4):1162-1181. DOI: 10.1080/10095020.2023.2178334.

[3] Lu E, Hu X. Image super-resolution via channel attention and spatial attention. *Appl Intell*. 2022;52(2):2260-2268. DOI: 10.1007/s10489-021-02464-6.

[4] Bhosle K, Musande V. Evaluation of deep learning CNN model for recognition of devanagari digit. *Artif Intell Appl*. 2023;1(2):114-118. DOI: 10.47852/bonviewAIA3202441.

[5] Preethi P, Mamatha HR. Region-based convolutional neural network for segmenting text in epigraphical images. *Artif Intell Appl*. 2023;1(2):119-127. DOI: 10.47852/bonviewAIA2202293.

[6] Zhao Q, et al. HyperAttentionDTI: Improving drug-protein interaction prediction by sequence-based deep learning with attention mechanism. *Bioinformatics*. 2022;38(3):655-662. DOI: 10.1093/bioinformatics/btab715.

[7] Cheng Z, et al. IIFDTI: predicting drug-target interactions through interactive and independent features based on attention mechanism. *Bioinformatics*. 2022;38(17):4153-4161. DOI: 10.1093/bioinformatics/btac485.

[8] Muruganantham P, Balakrishnan SM. Attention aware deep learning model for wireless capsule endoscopy lesion classification and localization. *J Med Biol Eng*. 2022;42(2):157-168. DOI: 10.1007/s40846-022-00686-8.

[9] Tomar I, Sreedevi I, Pandey N. State-of-art review of traffic light synchronization for intelligent vehicles: current status, challenges, and emerging trends. *Electronic*. 2022;11(3):465. DOI: 10.3390/electronics11030465.

[10] Gholamhosseinian A, Seitz J. Acomprehensive survey on cooperative intersection management for heterogeneous connected vehicles. *IEEE Access*. 2022;10:7937-7972. DOI: 10.1109/ACCESS.2022.3142450.

[11] Miletić M, et al. A review of reinforcement learning applications in adaptive traffic signal control. *IET Intell Transp Syst*. 2022;16(10):1269-1285. DOI: 10.1049/itr2.12208.

[12] Dewi C, et al. Synthetic data generation using DCGAN for improved traffic sign recognition. *Neural Comput Appl*. 2022;34(24):21465-21480. DOI: 10.1007/s00521-021-05982-z.

[13] LiuC, Ke L. Cloud assisted internet of things intelligent transportation system and the traffic control system in the smart city. *J Control Decis*. 2023;10(2):174-187. DOI: 10.1080/23307706.2021.2024460.

[14] Soydaner D. Attention mechanism in neural networks: where it comes and where it goes. *Neural Comput Appl*. 2022;34(16):13371-13385. DOI: 10.1007/s00521-022-07366-3.

[15] Lin S, Zeng Y, Gong Y. Learning of time-frequency attention mechanism for automatic modulation recognition. *IEEE Wirel Commun Lett*. 2022;11(4):707-711. DOI: 10.1109/LWC.2022.3140828.

[16] Ramaswamy SL, Chinnappan J. RecogNet-LSTM + CNN: a hybrid network with attention mechanism for aspect categorization and sentiment classification. *J Intell Inform Syst*. 2022;58(2):379-404. DOI: 10.1007/s10844-021-00692-3.

[17] Kavianpour P, et al. A CNN-BiLSTM model with attention mechanism for earthquake prediction. *J Supercomput*. 2023;79(17):19194-19226. DOI: 10.1007/s11227-023-05369-y.

[18] Alirezazadeh P, Schirrmann M, Stolzenburg F. Improving deep learning-based plant disease classification with attention mechanism. *GesundePflanzen*. 2023;75(1):49-59. DOI: 10.1007/s10343-022-00796-y.

[19] Babbar S, Bedi J. Real-time traffic, accident, and potholes detection by deep learning techniques: a modern approach for traffic management. *Neural Comput Appl*. 2023;35(26):19465-19479. DOI: 10.1007/s00521-023-08767-8.

[20] Gupta H, Verma OP. Monitoring and surveillance of urban road traffic using low altitude drone images: a deep learning approach. *Multimed Tools Appl*. 2022;81(14):19683-19703. DOI: 10.1007/s11042-021-11146-x.

[21] Kamal H, et al. Digital-twin-based deep reinforcement learning approach for adaptive traffic signal control. *IEEE Internet Things J*. 2024;11(12):21946-21953. DOI: 10.1109/JIOT.2024.3377600.

[22] Wang M, et al. UniTSA: A universal reinforcement learning framework for V2X traffic signal control. *IEEE Trans Veh Technol*. 2024;73(10):14354-14369. DOI: 10.1109/JIOT.2025.3586315.

[23] Kumar R, Sharma NVK, Chaurasiya VK. Adaptive traffic light control using deep reinforcement learning technique. *Multimed Tools Appl*. 2024;83(5):13851-13872. DOI: 10.1007/s11042-023-16112-3.

[24] Li W, et al. LSTM-TCN: dissolved oxygen prediction in aquaculture, based on combined model of long short-term memory network and temporal convolutional network. *Environ Sci Pollut Res*. 2022;29(26):39545-39556. DOI: 10.1007/s11356-022-18954-8.

[25] Fan J, et al. Parallel spatio-temporal attention-based TCN for multivariate time series prediction. *Neural Comput Appl*. 2023;35(18):13109-13118. DOI: 10.1007/s00521-021-05958-z.

[26] Huan H, et al. A text classification method based on a convolutional and bidirectional long short-term memory model. *Connect Sci*. 2022;34(1):2108-2124. DOI: 10.1080/09540091.2022.2098926.

[27] Zhang D, et al. An efficient lightweight convolutional neural network for industrial surface defect detection. *Artif Intell Rev*. 2023;56(9):10651-10677. DOI: 10.1007/s10462-023-10438-y.

[28] Gou C, Zhou Y, Li D. Driver attention prediction based on convolution and transformers. *J Supercomput*. 2022;78(6):8268-8284. DOI: 10.1007/s11227-021-04151-2.

[29] Gupta D, Sharma A. A two-stage attention augmented fully convolutional network-based dynamic video summarization. *Multimedia Syst*. 2023;29(6):3685-3701. DOI: 10.1007/s00530-023-01154-2.

[30] Zhang H, et al. Attention-based feature fusion generative adversarial network for yarn-dyed fabric defect detection. *Text Res J*. 2023;93(5-6):1178-1195. DOI: 10.1177/00405175221129654.

[31] Hu X, et al. Attention-based conv-LSTM and Bi-LSTM networks for large-scale traffic speed prediction. *J Supercomput*. 2022;78(10):12686-12709. DOI: 10.1007/s11227-022-04386-7.

[32] Li Z, et al. Fusion attention mechanism bidirectional LSTM for short-term traffic flow prediction. *J Intell Transp Syst*. 2024;28(4):511-524. DOI: 10.1080/15472450.2022.2142049.

[33] Zohourianshahzadi Z, Kalita JK. Neural attention for image captioning: review of outstanding methods. *Artif Intell Rev*. 2022;55(5):3833-3862. DOI: 10.1007/s10462-021-10092-2.