



ARTICLE

Spatiotemporal Prediction of Urban Traffics Based on Deep GNN

Ming Luo¹, Huili Dou² and Ning Zheng^{3,*}

¹School of Computer Science, Hangzhou Dianzi University, Hangzhou, 310018, China

²Zhejiang Institute of Communications, Hangzhou, 310012, China

³School of Cyberspace Security, Hangzhou Dianzi University, Hangzhou, 310018, China

*Corresponding Author: Ning Zheng. Email: nzheng@hdu.edu.cn

Received: 03 March 2023 Accepted: 13 September 2023 Published: 30 January 2024

ABSTRACT

Traffic prediction already plays a significant role in applications like traffic planning and urban management, but it is still difficult to capture the highly non-linear and complicated spatiotemporal correlations of traffic data. As well as to fulfil both long-term and short-term prediction objectives, a better representation of the temporal dependency and global spatial correlation of traffic data is needed. In order to do this, the Spatiotemporal Graph Neural Network (S-GNN) is proposed in this research as a method for traffic prediction. The S-GNN simultaneously accepts various traffic data as inputs and investigates the non-linear correlations between the variables. In terms of modelling, the road network is initially represented as a spatiotemporal directed graph, with the features of the samples at the time step being captured by a convolution module. In order to assign varying attention weights to various adjacent area nodes of the target node, the adjacent areas information of nodes in the road network is then aggregated using a graph network. The data is output using a fully connected layer at the end. The findings show that S-GNN can improve short- and long-term traffic prediction accuracy to a greater extent; in comparison to the control model, the RMSE of S-GNN is reduced by about 0.571 to 9.288 and the MAE (Mean Absolute Error) by about 0.314 to 7.678. The experimental results on two real datasets, Pe MSD7(M) and PEMS-BAY, also support this claim.

KEYWORDS

Urban traffic; traffic; temporal correlation; GNN; prediction

1 Introduction

The Internet of Things technology [1], which travels thousands of miles every day, has been applied to various traffic devices, generating a significant amount of traffic data. Intelligent transportation has emerged and demonstrated a vigorous development trend in response to the ongoing development of the social economy and the growing complexity of traffic road networks. How successfully this data is processed and used will have a significant impact on the growth of smart cities. Forecasting future traffic patterns in distinct urban areas is of vital importance in the study of intelligent transport systems. By analysing past traffic observations, traffic prediction predicts future metropolitan real-time traffic flows [2]. The public can enhance their travel experiences and safeguard their personal safety by managing traffic to avoid congested areas, preventing stampede situations and adopting safety precautions, minimising operational costs, and scheduling demand.



Prediction of traffic flows is necessary; prediction can help to anticipate risks, avoid congestion and prevent hazards. As a result, traffic prediction has become a hot research topic in recent years [3]. The key points to improve prediction accuracy are in the following areas: 1) A significant difficulty is how to effectively represent the spatial interactions, the temporal-spatial interactions, and the temporal interactions of crowd flows. Figs. 1 and 2 demonstrate how crowds create a complex web of interactions in both space and time. 2) How to effectively capture the potential similarity patterns between forecast data and historical traffic data (i.e., proximity data, periodic data and trend data) is another challenge [4]. The numbers in Fig. 2 show the cars that left from various locations and arrived at the destination at a specific time. For instance, yellow represents the vehicles that left from point A and blue represents the vehicles that arrived at point A. The vehicle traffic patterns are shown in Fig. 3, the x-axis represents time (seconds) and the y-axis represents observations (kb); closeness denotes previous data that is closer to the forecast, periodicity denotes the same time period each day, and trend denotes traffic data that exhibits a trend over a specific time period.

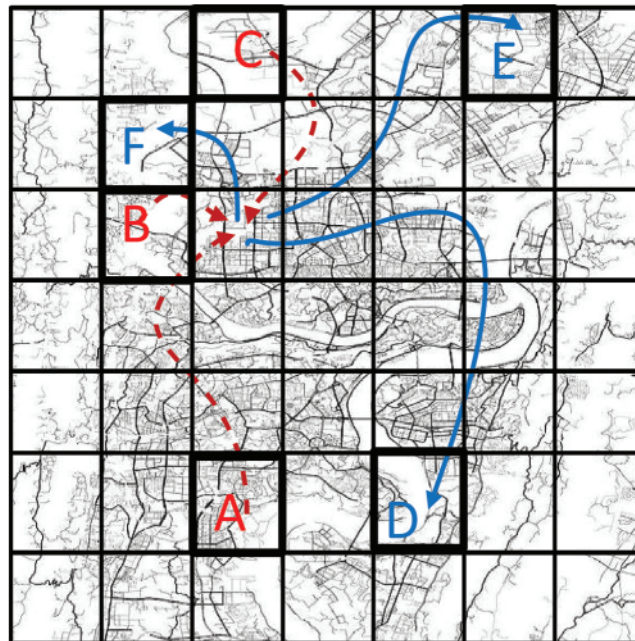


Figure 1: Spatial interaction of traffic data

Traditional traffic prediction techniques frequently only take into account a single information, such as a temporal or geographical feature; current deep learning techniques also take these features into account in combination, but they do not efficiently and completely capture them. To address the aforementioned issues, this research suggests S-GNN, a multi-attentional predictive RNN (Recurrent Neural Network) and a temporal sequence-dependent attentional convolutional long and short-term memory network [5]. CNN (Recurrent Neural Network) and predictive RNNs (Pred RNN) are first used to capture the spatiotemporal interaction features, and then a proximity attention module, a periodic attention module, and a trending attention module are used, respectively, to capture the proximity, periodicity, and trending of traffic data in the temporal sequence-dependent attention module.

time →

Regions	8am-9am	9am-10am	...	
A	90/270	110/208	...	
B	160/60	140/83	...	
C	130/69	100/89	...	
...	

Figure 2: Interaction of traffic data over time

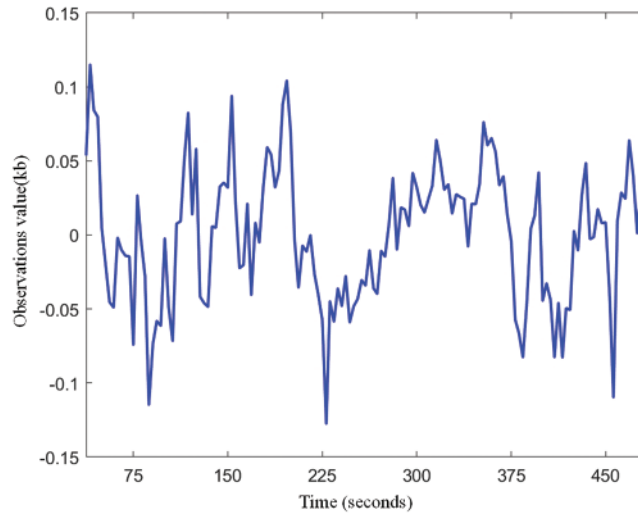


Figure 3: Vehicle traffic trends

The novel contributions of our S-GNN model can be highlighted as follows:

a. Comprehensive Representation: While various models incorporate time and space modules, our S-GNN aims to provide a more comprehensive representation by integrating multiple factors including flow, occupancy, and average speed simultaneously. Thus, it can capture complex relationships and nonlinear correlations among diverse traffic variables in a unified manner.

b. Enhanced Attention Mechanism: The proposed model employs a graph-based attention mechanism that goes beyond simple GAT. The attention mechanism in S-GNN adapts the attention weights to adjacent areas of target node, thus, it can effectively capture varying influences from different regions of the road network. This provides a finer-grained understanding of spatial correlations compared to traditional GAT.

c. Residual Connectivity and Model Complexity: We introduce residual connectivity within the proposed model, which contributes to better gradient flow during training, potentially leading to faster

convergence and improved prediction accuracy. Exploration of the optimal number of attention layers in our analysis indicates the importance of balancing model complexity and performance.

d. Extensive Evaluation: The proposed model is rigorously evaluated on real-world traffic datasets, PeMSD7(M) and PEMS-BAY, demonstrating its ability to achieve superior prediction accuracy compared to various baseline models. The results showcase its effectiveness not only in short-term but also in long-term traffic prediction scenarios.

The organization of the remaining paper is outlined as follows: [Section 2](#) presents an overview of related work in the field. In [Section 3](#), a comprehensive background understanding is provided. [Section 4](#) details the mathematical formulation of the proposed S-GNN model. The experimental results are presented in [Section 5](#). Finally, [Section 6](#) concludes the paper by summarizing the main findings and discussing potential future directions.

2 Related Works

The emergence and development of deep learning [6] have been a great improvement in capturing non-linear relationships between data, and therefore some scholars have also applied it to solve traffic flow prediction problems.

In order to handle the challenge of forecasting cab demand, the literature [7] presented a multi-view spatiotemporal network to understand the geographical links, temporal links, and regional similarity relationships that exist between data. According to the literature [8] suggested for capturing spatiotemporal aspects of global dynamics, GSTNet is a deep learning method made up of multilayer spatiotemporal blocks, each of which consists of a temporal module and a spatial module. Both the temporal and the spatial modules are used to record temporal and spatial characteristics, respectively. A deep learning framework (Dynamic GRCNN) was put forth in the literature [9] to address the issue of predicting urban traffic flow. Dynamic GRCNN models dynamic traffic station relationships as spatiotemporally correlated dynamic graph structures in order to capture the proximity, periodicity, and trends of traffic data based on historical traffic flows. A deep learning method for traffic prediction (MDGRU) was created in the literature [10] in response to the issue of travel delay, or the time needed for cars to change locations. This method develops a delay-based GRU (Gate Recurrent Unit) and builds the model by stacking. A two-stage deep STD (State Transform Diagram) was proposed in the literature [11] to predict urban traffic flow, which can simulate various regional perturbations and spatiotemporal propagation effects caused by various regional functions. Spatiotemporal perturbations are removed from historical traffic flows to enhance the learning of inherent traffic patterns in the prediction stage.

The literature [12] proposed a traffic flow prediction method that combines long and short-term memory networks, convolutional neural networks, using cellular neural networks to capture temporal correlations, and the model also takes into account external factors like weather in order to further improve the spatiotemporal characteristics of traffic flow data. In response to the complexity, nonlinear linkages, and unpredictability of traffic flow data, the literature [13] offered a novel spatiotemporal extended squeeze network (ST-ES-Net) framework, which designed numerous successful solutions. In the literature, a multi-task framework for the prediction of node and edge flows in spatiotemporal networks has been established [14]. The literature [15] proposed a deep learning embedding technique (DELA) for modelling fine-grained traffic information, route structure, and meteorological conditions.

In [16], Zhang et al. introduced the STAGCN model, which integrates a spatio-temporal attention mechanism into graph convolutional networks. This innovative approach demonstrated superior predictive accuracy compared to standard methods. In [17], Zhao et al. proposed MSTFGRN, a multi-spatio-temporal fusion graph recurrent network. By dynamically generating weighted adjacency matrices and incorporating global attention, MSTFGRN surpassed existing models in capturing complex dependencies. In [18], Lablack and Shen devised the STGM network, designed to tackle dynamic node interactions over time. Through a novel attention mechanism and efficient architecture, STGM achieved enhanced predictive performance while maintaining fewer parameters. In [19], Wang et al. introduced TransGAT, a dynamic graph attention network utilizing node embedding and TCN. This model excelled in various time intervals, exhibiting improved accuracy in traffic flow prediction. In [20], Li et al. developed SSTACON, a shared spatio-temporal attention convolution optimization network. SSTACON's unique design, encompassing attention layers and graph optimization, showcased superior results across different timeframes. In [21], Li et al. innovated CVSTGCN, a coordinated attention-based spatio-temporal graph convolutional network. By combining coordinate methods with graph convolutional networks, CVSTGCN demonstrated remarkable predictive power in both short and long-term traffic flow scenarios.

Although the aforementioned solutions have some impact on traffic flow, they ignore the spatiotemporal correlation, such as temporal features or spatial characteristics; current deep learning techniques also take these characteristics into account, but they do not effectively and fully capture them. However, these models suffer from various limitations including potential inability to capture abrupt traffic changes, and computational complexity, which could hinder real-time implementation.

3 Background Knowledge

3.1 The Traffic Prediction Problem

A traffic network can be defined as an graph $G = (V, E, A)$, where E represents the set of edges between nodes, reflecting the connection between links, and $A \in R^{N \times N}$ represents the adjacency matrix formed by Euclidean distance between sensors, The i and j elements $A = [i, j]$ represent the strength of the relationship between nodes v_i and v_j . If there is no close relationship between the two nodes, Then $A = [i, j] = 0$, each node records traffic characteristics, such as traffic value, speed, road occupancy, etc. In order not to lose generality, this study focuses on the prediction of speed, that is, the speed on the road section is regarded as the attribute of the network node [22]. Given the traffic conditions observed at the last T' hours, that is, the characteristic matrix $X^{(T')} = [x^{\tau-T'+1}, \dots, x^\tau] \in R^{N \times C}$, x_i is the traffic characteristics of each section at time i , and N is the number of sensors in the traffic network, That is, the number of nodes, C is the characteristic number. The traffic prediction problem can generally be expressed as the prediction of traffic information $(\bar{x}^{\tau+1}, \dots, \bar{x}^{\tau+T})$ at the next T hours by learning model F_θ on the premise of road network topology G and historical traffic data characteristic $X^{(T')}$, as shown in Eq. (1):

$$(\bar{x}^{\tau+1}, \dots, \bar{x}^{\tau+T}) = F_\theta ((x^{\tau-T'+1}, \dots, x^\tau); G) \quad (1)$$

3.2 Modeling of Diagram Structure

When aggregating the information of nodes in the road network, the accessibility of adjacent nodes and the impact of upstream and downstream relationships should be considered [23]. Therefore, in this study, the road network is modeled graph structure, and the edge set is defined as E , and the vertex set is defined as V . The topological structure of the road network is $G = (V, E)$. Define the

adjacency matrix of the graph as A , then $G_t = (V, E, A)$ is the graph at the t time step, and V_t is its point set. The adjacency matrix A of the graph with node n is expressed as:

$$\mathbf{A} = \begin{bmatrix} 1 & a_{1,2} & \cdots & a_{1,n-1} & a_{1,n} \\ a_{2,1} & 1 & \cdots & a_{2,n-1} & a_{2,n} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ a_{n-1,1} & a_{n-1,2} & \cdots & 1 & a_{n-1,n} \\ a_{n,1} & a_{n,2} & \cdots & a_{n,n-1} & 1 \end{bmatrix} \in \mathbb{R}^{n \times n} \quad (2)$$

4 S-GNN Model

S-GNN model structure is depicted in Fig. 4. A number of traffic flow parameters are entered into the S-GNN model's input layer. To determine the difference between the anticipated value and the actual value and to continuously update the network's parameters, the data sequence is sent through the time module, space module, and output layer.

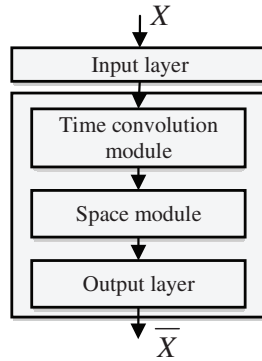


Figure 4: Structure of the S-GNN model

4.1 Multi Factor Input

Numerous factors, such as flow, occupancy, average speed, and others are taken into consideration while determining the traffic flow order. Each component has an impact on the others. It is necessary to investigate the nonlinear link between traffic forecast and traffic sequence further [24]. In order to identify the nonlinear link between the various traffic variables and make accurate predictions, this work models traffic, time occupancy, and average speed all at once.

4.2 Time Module

In the time dimension, the standard one-dimensional convolution is used to aggregate the characteristics of the time step. Compared with LSTM (Long Short-Term Memory) and GRU models, one-dimensional convolution does not rely on the previous output during training, does not need iteration, has less computation and low time complexity [25].

1D convolution operation is a fundamental mathematical operation which involves sliding a filter (also known as a kernel) along a 1D input sequence and computing the dot product between the filter and the corresponding portion of the input at each position. This operation allows the network to capture local patterns and features within the input sequence. The operation principle of one-dimensional convolution is shown in Fig. 5 (for more details see [26]).

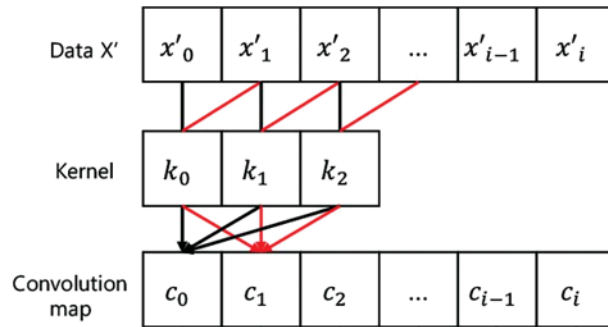


Figure 5: 1D convolution operation (rebuilt from [27])

Set the input data as X , the characteristic length of each vertex as C_i , and the time step as M [27]. The convolution operation with the size of one dimension is performed along the time dimension without padding, and the input X is mapped to the matrix Y , and the output of the time dimension is:

$$Y = Comv(X) \tag{3}$$

4.3 Space Module

GAT (Graph Attention Network) is used to process the directed graph in the spatial dimension, and the multi-head self-attention mechanism is utilised to create each attention subspace separately, lowering the likelihood that the model would become overfitted and increasing the model's capacity for spatial expression [28]. The GAT structure is shown in Fig. 6.

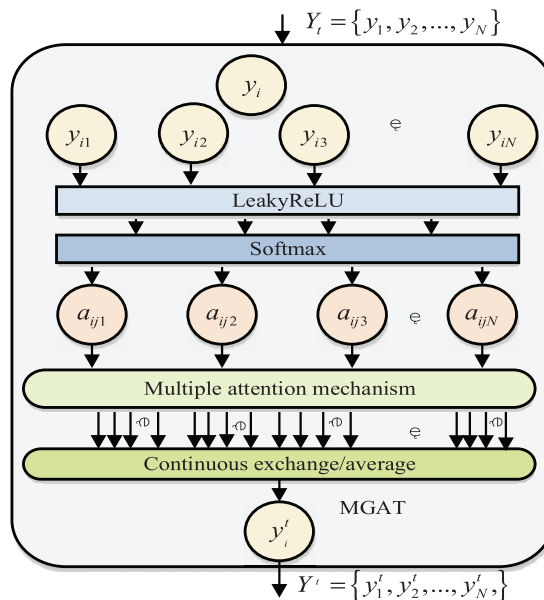


Figure 6: Structure of GAT model

Take Y module as the input of the space module:

$$Y = \{y_1, y_2, \dots, y_N\} \quad (4)$$

where y_N is node N feature vector after temporal module output. The spatial module output is:

$$Y' = \{y'_1, y'_2, \dots, y'_N\} \quad (5)$$

Let the weight matrix W that transforms the input features into output features, then the attention value of node j on node i is:

$$e_{ij} = a(Wy_i, W_j) \quad (6)$$

where $a(\cdot)$ is the function that calculates the correlation of the nodes.

Use the *LeakyReLU* function to activate it:

$$e_{ij} = \text{LeakyReLU}(a^T [Wy_i || Wy_j]) \quad (7)$$

where *LeakyReLU* (\cdot) is the activation function.

Find the attention values of all neighbors of node i . Use the *soft max* function to normalize the attention weights:

$$\alpha_{ij} = \text{soft max}(e_{ij}) = \frac{\exp[\text{LeakyReLU}(a^T [Wy_i || Wy_j])]}{\sum_{j \in N_i} \exp[\text{LeakyReLU}(a^T [Wy_i || Wy_j])]} \quad (8)$$

where *LeakyReLU* (\cdot) is the activation function; a is the vector to be learned; a^T is the transpose of vector \mathbf{a} ; and $||$ is the splicing operation.

With the graph attention layer feature extraction using weighted as output:

$$y'_i = \sigma \left(\sum_{j \in N_i} \alpha_{ij} \mathbf{W} y_j \right) \quad (9)$$

where $\sigma(\cdot)$ is the activation function; α_{ij} is the attention value of node i at node j .

If a multi-headed attention where K sets of attention coefficients are calculated independently and then averaged output is:

$$y'_i = \sigma \left(\frac{1}{K} \sum_{K=1}^K \sum_{j \in N_i} \alpha_{ij}^k \mathbf{W}^k y_j \right) \quad (10)$$

where α_{ij}^k is the k th set of attention values of node i on node j .

Then final prediction is:

$$\bar{y} = Y'w + b \quad (11)$$

The loss function is:

$$L(\bar{y}; W_\theta) = \|\bar{y}(y_{t-M+1}, \dots, y_t, W_\theta) - y'_{t+1}\|^2 \quad (12)$$

5 Experimental Results

5.1 Data Set Description and Setup

In the experiments, *PeMSD7 (M)* and *PEMS – BAY* datasets were used to evaluate the proposed model. *PeMSD7 (M)* Collected traffic information from 229 monitoring stations on the California Highway System during weekdays from May to June 2012. *PEMS – BAY* Contains traffic information collected from 326 sensors in the Bay Area, California, between January 01 and May 31, 2017. Traffic data information was summarized every 5 min and normalized using *Z – Score* as input. [Table 1](#) provides detailed statistical information on the dataset.

Table 1: Dataset statistical

Data	#Nods	#Edges	#Time steps
PeMSD7(M)	229	833	12673
PEMS-BAY	326	2370	52117

The topological information of the traffic network is represented by the graph adjacency matrix. For the *PEMS – BAY* dataset, the edge with the maximum weight between each pair of nodes is used to construct the undirected graph and represent the topological information of the traffic road network in this way; for the *PeMSD7 (M)* dataset, the weighted adjacency matrix W can be expressed as:

$$w_{ij} = \begin{cases} \exp\left(-\frac{d_{ij}^2}{\delta}\right), i \neq j \text{ and } \exp\left(-\frac{d_{ij}^2}{\delta}\right) \geq \varepsilon \\ 0, \text{ other} \end{cases} \quad (13)$$

where $\delta = 0.1$ and $\varepsilon = 0.5$, w_{ij} are the edge weights associated with d_{ij} (the distance between node i and node j), respectively, based on thresholds δ and ε that represent the distribution and sparsity of the weighted adjacency matrix W in the literature. By implementing a grid search strategy, the best parameters were located at the time of validation. Data from 12 observations (historical time step $\tau = 12$) were used on both datasets, i.e., a historical time window of 60 min was used to predict traffic conditions for the next 15, 30 and 60 min (time step $T = 3, 6, 9$). The temporal convolution layer's size of the expansion factor is set at 1, 2, 1, 2, 1 in order to cover the length of the input sequence. The spatiotemporal convolutional layer has 64, 16, 64 channels in total. The Adam optimizer, which has the benefits of straightforward implementation, effective calculation, low memory requirements, and suitability for situations involving huge amounts of data and parameters, is utilised for optimisation in this study [29]. The iteration period is 50, the batch size parameter is 50, the learning rate parameter is 0.001, and the weight decay factor of L2 regularization is 5×10^{-4} .

5.2 Assessment Metrics and Baselines

In this paper, three widely used evaluation metrics, Mean Absolute Error (MAE), Root Mean Square Error (RMSE) and Mean Absolute Percentage Error (MAPE) are used to measure the performance of the different methods. The baseline models compared are: 1) historical averaging model (HA); 2) ARIMA [12]; 3) FC-LSTM [14]; 4) STGCN [3]; 5) Graph Wave Net [7]; 6) Space-Time Transformation Network (STTN); 7) Graph Convolution Network (LSGCN) using gated line cells.

5.3 Comparison and Analysis of Results

Table 2 shows the results, and it is clear from the three-assessment metrics that the S-GNN performs admirably in both short- and long-term prediction [30]. Both datasets showed promising results for S-GNN, particularly in the long-term prediction task. The table presents performance metrics for various traffic flow prediction models on two datasets, PeMSD7(M) and PEMS-BAY, at different time intervals (15/30/60 min). The metrics include Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and Mean Absolute Percentage Error (MAPE%). The models range from traditional methods like HA and ARIMA to advanced neural network models like FC-LSTM, STGCN, Graph Wave Net, and more. The proposed S-GNN model consistently outperforms other models across both datasets and time intervals, demonstrating significantly lower MAE, RMSE, and MAPE%. Notably, the model achieves remarkable accuracy with MAE values as low as 1.90 for PeMSD7(M) and 1.23 for PEMS-BAY, showcasing its effectiveness in traffic flow prediction across different scenarios and prediction horizons.

Table 2: Performance of different methods on the PEMS-BAY and Pe MSD (M) datasets

Model	PeMSD7(M) (15/30/60 min)		
	MAE	RMSE	MAPE%
HA	4.02	7.21	10.62
ARIMA	5.56/6.25/6.76	9.01/9.14/9.73	12.93/13.95/16.93
FC-LSTM	3.58/3.95/4.36	6.21/7.04/7.94	8.618/9.56/10.62
STGCN	2.26/3.04/4.02	4.05/5.71/7.56	5.27/7.34/9.79
Graph Wave Net	2.15/2.81/3.43	4.02/5.49/6.73	4.94/6.90/8.58
LSGCN	2.23/2.97/3.82	3.99/5.48/7.10	5.15/7.19/9.61
STTN	2.15/2.71/3.32	4.05/5.38/6.52	5.06/6.89/7.99
STAGCN	2.19/6.79/3.29	4.26/5.26/6.92	5.30/6.62/7.35
MSTFGRN	2.27/2.86/4.17	4.08/5.30/6.87	5.26/6.46/7.35
STGM	2.35/2.80/4.15	4.11/4.92/6.70	4.64/6.30/7.38
TransGAT	2.17/2.69/3.64	3.95/4.80/6.54	4.55/6.26/7.24
SSTACON	2.05/2.87/3.50	3.87/4.87/6.22	4.50/5.88/7.15
CVSTGCN	1.98/2.61/3.41	3.67/4.76/6.05	4.38/5.78/6.98
Proposed S-GNN	1.90/2.43/3.10	3.37/4.38/5.47	4.165/5.33/6.89
Model	PEMS-BAY (15/30/60 min)		
	MAE	MAE	MAE
HA	2.89	5.60	6.81
ARIMA	1.63/2.24/3.39	3.31/4.77/6.51	3.51/5.41/8.30
FC-LSTM	2.06/2.21/2.38	4.20/4.56/4.97	4.81/5.21/5.71
STGCN	1.40/1.85/2.43	2.93/4.13/5.34	3.01/4.23/5.59
Graph Wave Net	1.31/1.68/1.96	2.74/3.68/4.64	2.75/3.71/4.53
STTN	1.37/1.68/1.96	2.88/3.80/4.51	2.90/3.79/4.59
STAGCN	1.87/2.05/2.67	3.24/3.85/4.28	3.22/4.42/4.30
MSTFGRN	1.96/1.96/2.70	3.16/3.72/4.12	3.10/4.26/4.25

(Continued)

Table 2 (continued)

Model	PeMSD7(M) (15/30/60 min)		
	MAE	RMSE	MAPE%
STGM	2.04/2.02/2.26	2.96/3.64/4.09	2.92/3.98/4.11
TransGAT	1.40/1.88/2.35	3.05/3.78/4.05	2.85/3.78/4.05
SSTACON	1.38/1.87/2.26	2.94/3.76/3.84	2.95/3.92/4.26
CVSTGCN	1.45/1.96/2.04	2.88/3.65/3.82	2.87/3.90/4.15
Proposed S-GNN	1.23/1.53/1.87	2.55/3.21/3.40	2.40/3.13/3.76

The experimental results show that traditional traffic prediction statistical models, like the HA and ARIMA models, do not have satisfactory prediction results because they cannot handle complex spatiotemporal data, while deep learning-based spatiotemporal prediction methods, like FC-LSTM, STGCN, Graph Wave Net, STTN, and S-GNN, tend to achieve better prediction results than traditional prediction statistical models. STGCN captures both temporal and spatial dependence jointly through gated convolution and graph convolution-based approaches, which is significantly better than FC-LSTM, but it relies on graph structure, ignores global spatial features, and cannot determine the significance of inter-node relationships. FC-LSTM uses LSTM to capture the temporal dependence of traffic data, but it is still insufficient in capturing spatial dependence [31]. In comparison to STGCN, Graph Wave Net's bi-directional diffusion graph convolution takes into account the influence of directionality, improving prediction results; however, STTN performs better in long-term prediction than Graph Wave Net, but still has issues with short-term prediction, and the overall prediction results are still slightly below the model proposed. The Pe MSD(M) dataset is smaller relative to the PEMS-BAY dataset, i.e., smaller in both time and space, resulting in a much larger standard deviation of its acquisition speed than the PEMS-BAY dataset, and consequently the capture of spatiotemporal correlations of traffic data on the Pe MSD(M) dataset is relatively more difficult [32]. This is why the majority of baseline models and the models in this study perform better on the PEMS-BAY dataset.

MAE, RMSE and MAPE values are plotted for both short- and long-term forecasts on the Pe MSD(M) dataset, as shown in Fig. 7. In general, as the forecast time grows, the corresponding forecast difficulty and forecast error also increases. From the Fig. 7, it can be seen that the performance of traditional prediction models is poor and deep learning-based models can achieve better results in spatiotemporal prediction, and the prediction performance of the network model designed in this paper is higher than that of other deep learning methods. In order to more intuitively show our model's prediction of traffic flow, the prediction effect is shown in Fig. 8.

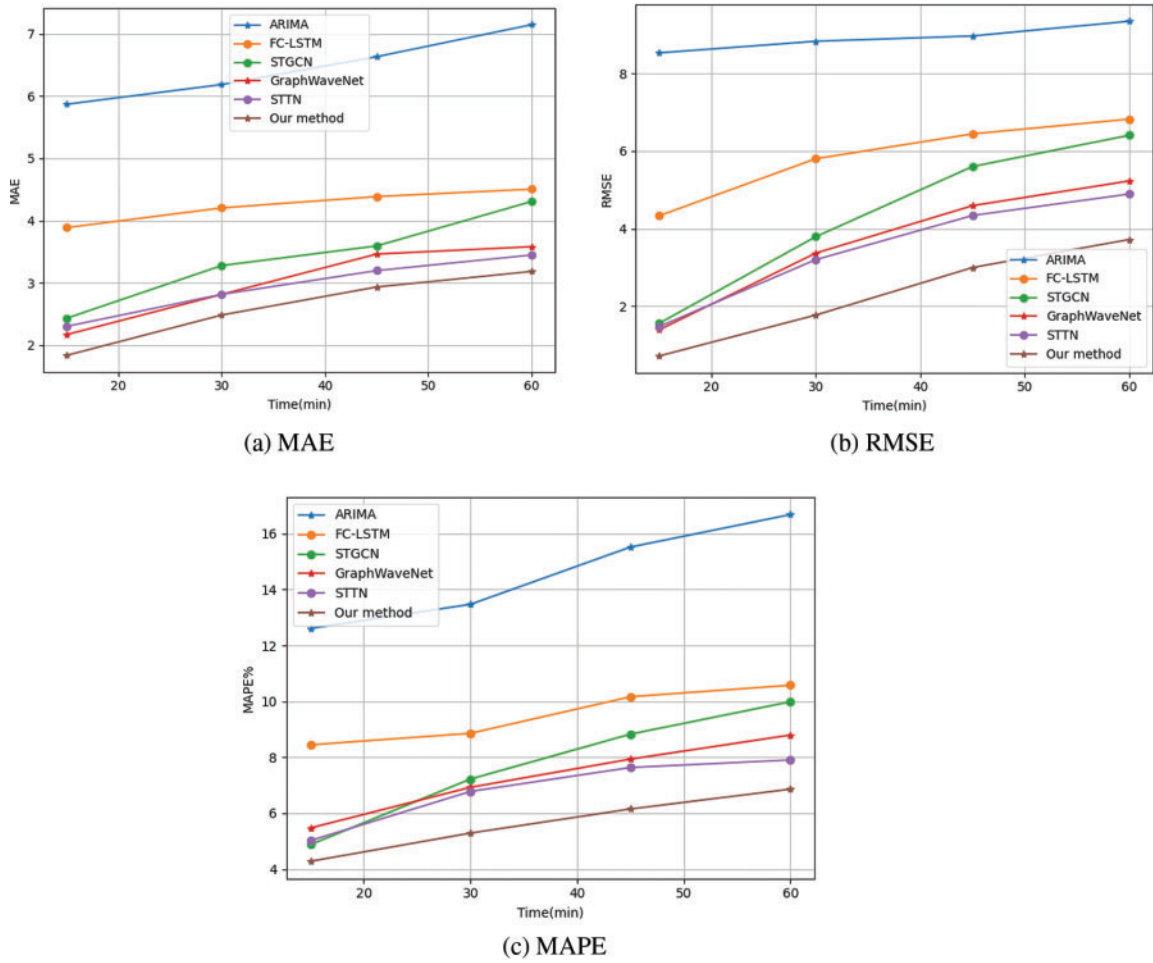


Figure 7: Results of the PeMSD7(M) prediction

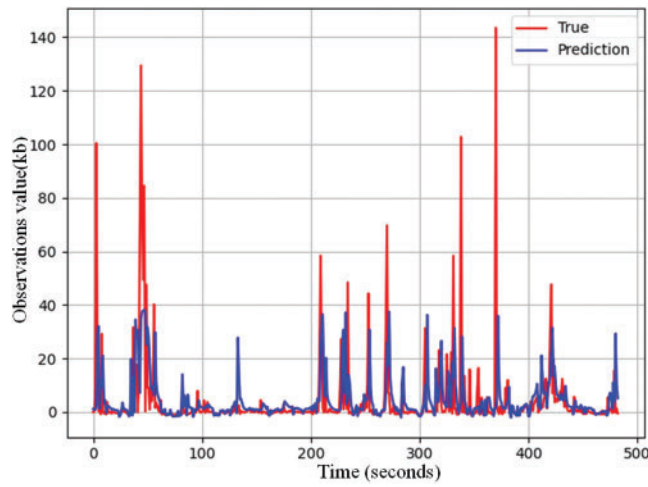


Figure 8: Our technique for predicting traffic

5.4 Network Design Analysis

In order to assess the effectiveness of the attention mechanism described in this research, the proposed model S-GNN, the model S-GNN-1 without the attention mechanism, and the STGCN model with the same iteration process were compared on the Pe MSD(M) dataset and the PEMS-BAY dataset. Figs. 9 and 10 show the outcomes, respectively. The outcomes show that S-GNN can reduce prediction error and improve the model's performance.

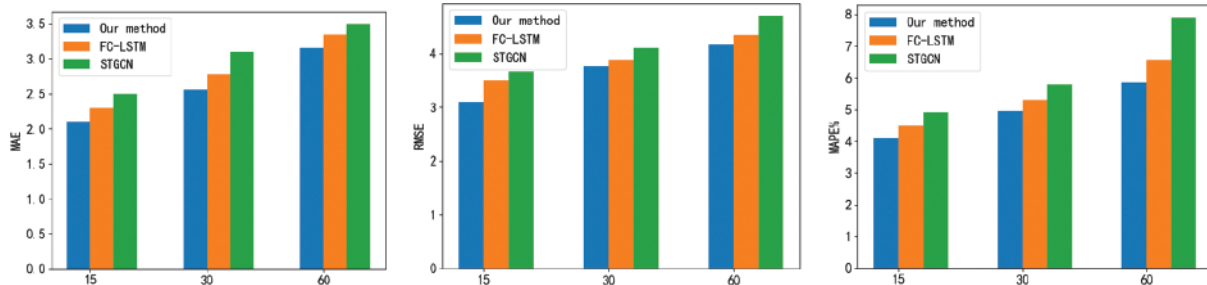


Figure 9: PeMSD7(M) dataset

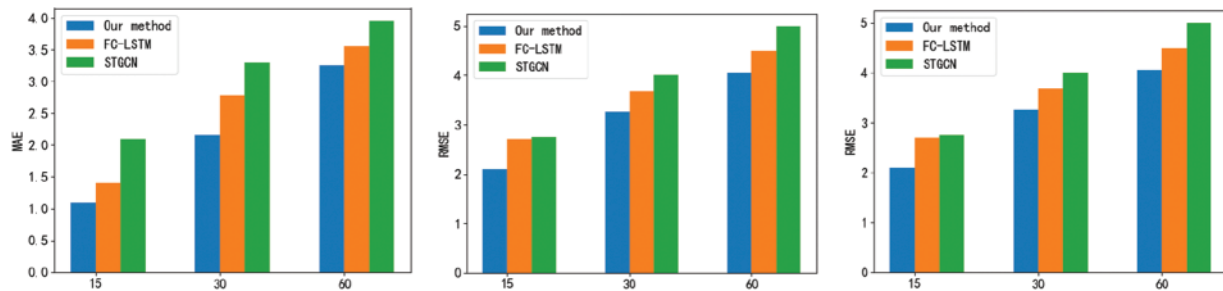


Figure 10: PEMS-BAY dataset

5.5 Convergence Speed Analysis

This section aimed to assess the convergence rates of three distinct models: STGCN, FC-LSTM (after residuals removal), and S-GNN, employing the PeMSD7(M) dataset for evaluation. By plotting training loss curves against the number of iterations, the study examined the convergence behavior for 15, 30, and 60-min forecasts. The obtained results, presented in Fig. 11, exhibited intriguing patterns.

S-GNN model consistently demonstrated the most rapid convergence to the lowest loss values across all time horizons. This effect was particularly pronounced in the context of long-term predictions. The outcome underscores S-GNN's exceptional fitting capability for the specific test dataset, accentuating its superiority over both STGCN and FC-LSTM model [33].

These findings offer valuable insights into the dynamics of model convergence, shedding light on S-GNN's effectiveness in adapting to the intricacies of the PeMSD7(M) dataset. In essence, S-GNN emerges as a compelling solution, boasting remarkable convergence rates and highlighting its potential for accurate and efficient traffic flow prediction, especially in prolonged forecasting scenarios.

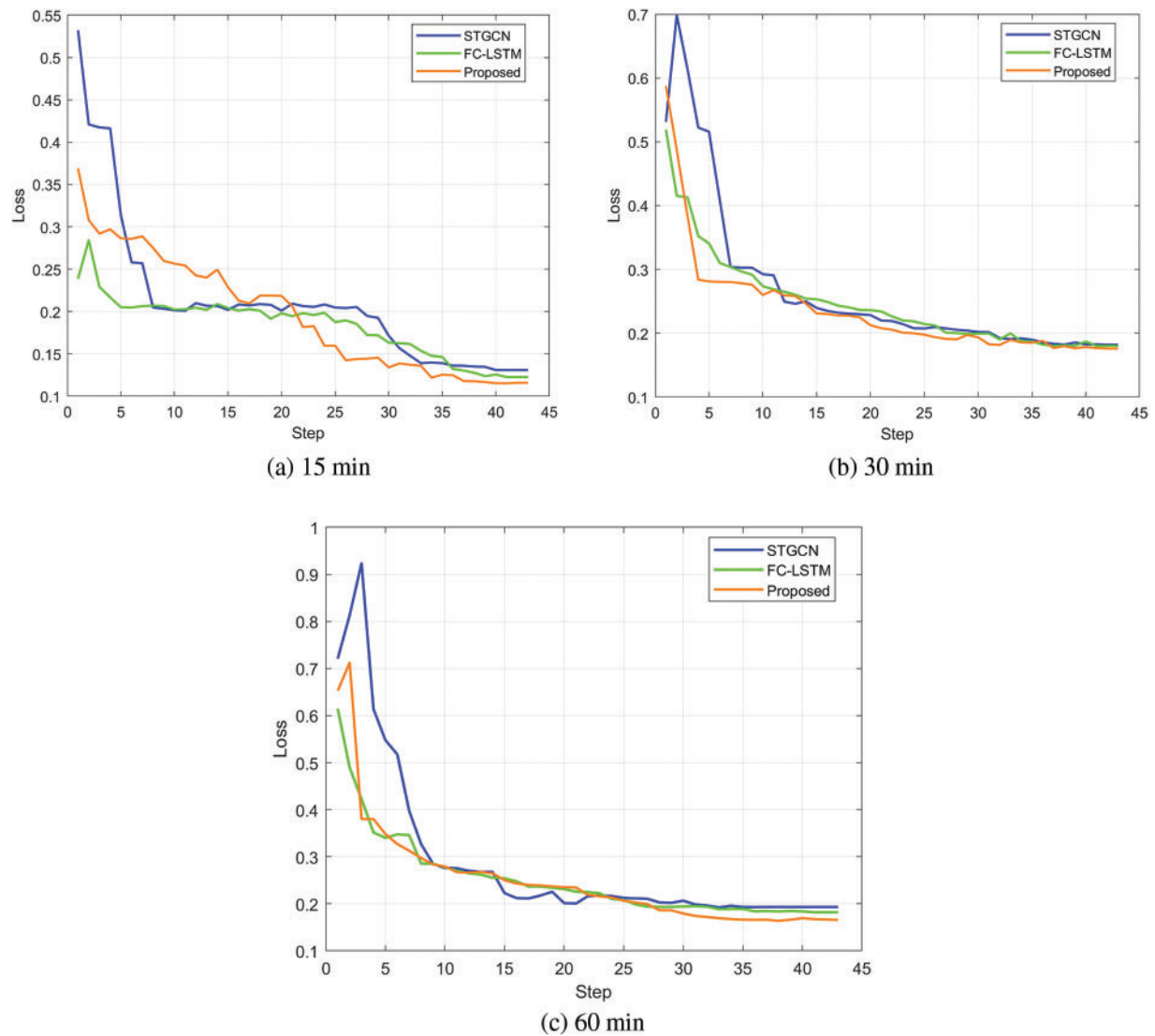


Figure 11: Convergence rate analysis

5.6 Analysis of Attentional Layers

The Pe MSD(M) dataset was used for the graph attention layer validation trials, and the outcomes are displayed in [Table 3](#). According to the experimental findings, the training time and prediction error rise as the number of attention layers grows, therefore the prediction performance is higher when the number of attention layers is 1.

5.7 Computational Complexity Analysis

On the PEMS-BAY dataset, the number of parameters, training time, and testing time of STGCN, STTN, and proposed S-GNN were further compared. In [Table 4](#), the experimental findings are displayed. The S-GNN suggested in this paper has faster training and testing times than STTN since STTN stacks numerous spatiotemporal converters to speed up calculation. In contrast to STGCN,

our training and testing periods are a little bit longer. This is as a result of the addition of an attention module, which increases the network’s classification accuracy. Additionally, due to the increase in parameters brought on by the usage of the attention module and residual connectivity, the number of parameters in our network is higher than that of STGCN and STTN.

Table 3: Effect of different attention layers on the prediction performance of the Pe MSD(M) dataset

Attention layer	MAE (15/30/60 min)	RMSE (15/30/60 min)	MAPE (15/30/60 min)	Training time (epoch)
1	1.88/2.41/3.08	3.35/4.36/5.45	4.13/5.31/6.87	14 s
2	1.90/2.40/3.41	3.56/4.67/5.85	4.34/5.58/7.26	23 s
3	1.89/2.84/3.53	3.69/4.76/5.98	4.56/5.80/7.47	36 s

Table 4: Computational complexity of the PEMS-BAY dataset

Mode	#Parameters	Training time (epoch)	Test time
STGCN	582784	56 s	50 s
STTN	483421	268 s	258 s
Proposed S-GNN	659388	82 s	77 s

6 Conclusion

This paper introduced an efficient approach for addressing the challenges of accurate traffic prediction by proposing the Spatiotemporal Graph Neural Network (S-GNN) model. The complex and non-linear spatiotemporal correlations in traffic data posed significant hurdles to conventional prediction methods. S-GNN model was designed to capture both short-term and long-term traffic patterns to achieve better performance compared to existing statistical models and state-of-the-art deep learning models. The model’s architecture contained temporal and spatial modules that effectively processed multiple traffic data inputs. The temporal module utilized one-dimensional convolutions, allowing for efficient time-dependent feature aggregation. Meanwhile, the spatial module employed the Graph Attention Network (GAT) to capture intricate global spatial relationships within the road network. Extensive experiments were conducted on real-world datasets, PeMSD7(M) and PEMS-BAY, to validate the efficacy of the proposed S-GNN model. The model demonstrated superior predictive accuracy, especially in long-term prediction tasks, effectively surpassing various baseline models. The results highlighted S-GNN’s capability to handle the intricacies of spatiotemporal traffic data and its potential contribution to traffic planning, management, and urban development. Furthermore, the research analyzed critical aspects of the proposed S-GNN model, encompassing the influence of attentional layers and computational complexity. The findings revealed that a single attention layer yielded optimal results, and S-GNN’s computational efficiency was on par with relevant state-of-the-art models. When contrasted with the control model, S-GNN demonstrates a noteworthy reduction in RMSE by approximately 0.571 to 9.288, and in MAE (Mean Absolute Error) by around 0.314 to 7.678. This assertion is reinforced by the outcomes of experiments conducted on two authentic datasets, PeMSD7(M) and PEMS-BAY.

Future work in this area will be directed towards further optimizing the network architecture, streamlining computational complexity, and enhancing the overall performance of traffic forecasting models. The goal is to refine S-GNN by fine-tuning its parameters, exploring novel attention mechanisms, and investigating advanced techniques to efficiently handle larger datasets. Additionally, efforts will be concentrated on reducing computational complexity without compromising predictive accuracy, thereby making the model more practical for real-time applications. Additionally, the future work could explore alternative architectural choices by incorporating more complex temporal modeling techniques, to address the specific challenges associated with very long-term traffic prediction.

Acknowledgement: Thanks to our tutors and researchers for their assistance and guidance.

Funding Statement: This thesis was supported by Science and Technology Plan Project of Zhejiang Provincial Department of Transportation “Research and System Development of Highway Asset Digitalization Technology in Use Based on High-Precision Map” (Project Number: 202203); in part by Science and Technology Plan Project of Zhejiang Provincial Department of Transportation: Research and Demonstration Application of Key Technologies for Precise Sensing of Expressway Thrown Objects (No. 202204).

Author Contributions: Study conception and design: Ming Luo, Huili Dou; data collection: Ning Zheng; analysis and interpretation of results: Ming Luo, Huili Dou, Ning Zheng; draft manuscript preparation: Huili Dou, Ning Zheng. All authors reviewed the results and approved the final version of the manuscript.

Availability of Data and Materials: The experimental data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest: The authors declare that they have no conflicts of interest to report regarding the present study.

References

- [1] D. Pavlyuk, “Spatiotemporal cross-validation of urban traffic forecasting models,” *Transportation Research Procedia*, vol. 52, pp. 179–186, 2021.
- [2] D. Wang, J. Zhu, Y. Yin, J. Ignatius, X. Wei *et al.*, “Dynamic travel time prediction with spatiotemporal features: Using a GNN-based deep learning method,” *Annals of Operations Research*, vol. 139, pp. 1–21, 2023.
- [3] J. Tang and J. Zeng, “Spatiotemporal gated graph attention network for urban traffic flow prediction based on license plate recognition data,” *Computer-Aided Civil and Infrastructure Engineering*, vol. 37, no. 1, pp. 3–23, 2022.
- [4] J. Bai, J. Zhu, Y. Song, L. Zhao, Z. Hou *et al.*, “A3T-GCN: Attention temporal graph convolutional network for traffic forecasting,” *ISPRS International Journal of Geo-Information*, vol. 10, no. 7, pp. 485, 2021.
- [5] Z. Cao, L. Lu, C. Chen and X. Chen, “Modeling and simulating urban traffic flow mixed with regular and connected vehicles,” *IEEE Access*, vol. 9, pp. 10392–10399, 2021.
- [6] Y. Djenouri, A. Belhadi, J. C. W. Lin, D. Djenouri and A. Cano, “A survey on urban traffic anomalies detection algorithms,” *IEEE Access*, vol. 7, pp. 12192–12205, 2019.
- [7] H. Zou, K. Cao and C. Jiang, “Spatiotemporal visual analysis for urban traffic characters based on video surveillance camera data,” *ISPRS International Journal of Geo-Information*, vol. 10, no. 3, pp. 177, 2021.

- [8] S. Fang, Q. Zhang, G. Meng, S. Xiang and C. Pan, "GSTNet: Global spatial-temporal network for traffic flow prediction," in *Proc. of 28th Int. Joint Conf. on Artificial Intelligence (IJCAI'19)*, AAAI Press, pp. 2286–2293, 2019.
- [9] F. Zhou, L. Li, K. Zhang and G. Trajcevski, "Urban flow prediction with spatial-temporal neural odes," *Transportation Research Part C: Emerging Technologies*, vol. 124, no. 12, pp. 102912, 2021.
- [10] N. A. Gilbert, J. L. Stenglein, T. R. Van Deelen, P. A. Townsend and B. Zuckerberg, "Behavioral flexibility facilitates the use of spatial and temporal refugia during variable winter weather," *Behavioral Ecology*, vol. 33, no. 2, pp. 446–454, 2022.
- [11] C. Ye, Y. Ding, P. Wang and Z. Lin, "A data-driven bottom-up approach for spatial and temporal electric load forecasting," *IEEE Transactions on Power Systems*, vol. 34, no. 3, pp. 1966–1979, 2019.
- [12] L. S. Darby, C. J. Senff, R. J. Alvarez II, R. M. Banta, L. Bianco *et al.*, "Spatial and temporal variability of ozone along the colorado front range occurring over 2 days with contrasting wind flow," *Elementa Science of the Anthropocene*, vol. 9, no. 1, pp. 00146, 2021.
- [13] M. E. Cyr, A. G. Etchin, B. J. Guthrie and J. C. Benneyan, "Access to specialty healthcare in urban versus rural US populations: A systematic literature review," *BMC Health Services Research*, vol. 19, no. 1, pp. 1–17, 2019.
- [14] Y. Yang, A. Heppenstall, A. Turner and A. Comber, "Using graph structural information about flows to enhance short-term demand prediction in bike-sharing systems," *Computers Environment and Urban Systems*, vol. 83, pp. 101521, 2020.
- [15] H. Yuan, X. N. Zhu, Z. Hu and C. H. Zhang, "Deep multi-view residual attention network for crowd flows prediction," *Neurocomputing*, vol. 404, pp. 198–212, 2020.
- [16] H. Zhang, L. Chen, J. Cao, X. Zhang, S. Kan *et al.*, "Traffic flow forecasting of graph convolutional network based on spatio-temporal attention mechanism," *International Journal of Automotive Technology*, vol. 24, no. 4, pp. 1013–1023, 2023.
- [17] W. Zhao, S. Zhang, B. Zhou and B. Wang, "Multi-spatio-temporal fusion graph recurrent network for traffic forecasting," *Engineering Applications of Artificial Intelligence*, vol. 124, pp. 106615, 2023.
- [18] M. Lablack and Y. Shen, "Spatio-temporal graph mixformer for traffic forecasting," *Expert Systems with Applications*, vol. 228, pp. 120281, 2023.
- [19] T. Wang, S. Ni, T. Qin and D. Cao, "TransGAT: A dynamic graph attention residual networks for traffic flow forecasting," *Sustainable Computing, Informatics and Systems*, vol. 36, pp. 100779, 2022.
- [20] P. Li, C. Ke, H. Tu, H. Zhang and X. Zhang, "Shared spatio-temporal attention convolution optimization network for traffic prediction," *Journal of Information Processing Systems*, vol. 19, no. 1, pp. 130–138, 2023.
- [21] M. Li, M. Li, B. Liu, J. Liu, Z. Liu *et al.*, "Spatio-temporal traffic flow prediction based on coordinated attention," *Sustainability*, vol. 14, no. 12, pp. 7394, 2022.
- [22] V. Osipov, V. Nikiforov, N. Zhukova and D. Miloserdov, "Urban traffic flows forecasting by recurrent neural networks with spiral structures of layers," *Neural Computing and Applications*, vol. 32, no. 18, pp. 14885–14897, 2020.
- [23] Z. Chen, B. Zhao, Y. Wang, Z. Duan and X. Zhao, "Multitask learning and GCN-based taxi demand prediction for a traffic road network," *Sensors*, vol. 20, no. 13, pp. 3776, 2020.
- [24] B. Shen, X. Xu, J. Li, A. Plaza and Q. Huang, "Unfolding spatial-temporal patterns of taxi trip based on an improved network kernel density estimation," *International Journal of Geo-Information*, vol. 9, no. 11, pp. 683, 2020.
- [25] Y. Huang, K. Shi, H. Zong, T. Zhou and J. Shen, "Exploring spatial and temporal connection patterns among the districts in chongqing based on highway passenger flow," *Remote Sensing*, vol. 12, no. 1, pp. 27, 2019.
- [26] S. Nam, H. Park, C. Seo and D. Choi, "Forged signature distinction using convolutional neural network for feature extraction," *Applied Sciences*, vol. 8, no. 2, pp. 153, 2018.
- [27] P. Wang, T. Zhang, Y. Zheng and T. Hu, "A multi-view bidirectional spatiotemporal graph network for urban traffic flow imputation," *International Journal of Geographical Information Science*, vol. 36, no. 6, pp. 1231–1257, 2022.

- [28] L. Kang, G. Hu, H. Huang, W. Lu and L. Liu, "Urban traffic travel time short-term prediction model based on spatiotemporal feature extraction," *Journal of Advanced Transportation*, vol. 2020, no. 332, pp. 1–16, 2020.
- [29] N. Lysyak, "Methodical approaches to the analysis of the state of urbanized systems," *Socio-Economic Problems of the Modern Period of Ukraine*, vol. 4, no. 144, pp. 18–26, 2020.
- [30] N. Serok, O. Levy, S. Havlin and E. Blumenfeld-Lieberthal, "Unveiling the inter-relations between the urban streets network and its dynamic traffic flows: Planning implication," *Environment and Planning B*, vol. 46, no. 7, pp. 1362–1376, 2019.
- [31] H. Zhao, J. Zhao, X. Y. Xu, S. H. Wu and Q. Zhu, "Spatial distribution, sources and risk assessment of polycyclic aromatic hydrocarbons in a rapid urbanization city: Shenzhen," *Journal of Ecology and Rural Environment*, vol. 35, no. 1, pp. 38–45, 2019 (In Chinese).
- [32] Z. Xiao, X. Fu, L. Zhang and R. S. M. Goh, "Traffic pattern mining and forecasting technologies in maritime traffic service networks: A comprehensive survey," *IEEE Transactions on Intelligent Transportation Systems*, vol. 21, no. 5, pp. 1796–1825, 2020.
- [33] X. Luo, C. Zhang and L. Bai, "A fixed clustering protocol based on random relay strategy for EHWSN," *Digital Communications and Networks*, vol. 9, no. 1, pp. 90–100, 2023.