IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 2 November 2023, accepted 29 December 2023, date of publication 4 January 2024, date of current version 26 April 2024.

Digital Object Identifier 10.1109/ACCESS.2024.3349690

RESEARCH ARTICLE

Spatial–Temporal Traffic Flow Prediction With Fusion Graph Convolution Network and Enhanced Gated Recurrent Units

CHUANG CAI[1], ZHIJIAN QU[1], LIQUN MA[1], LIANFEI YU[1], WENBO LIU[1], AND CHONGGUANG REN[1,2]

[1]School of Computer Science and Technology, Shandong University of Technology, Zibo 255000, China
[2]Jinan Inspur Data Technology Company Ltd., Jinan 250000, China

Corresponding author: Chongguang Ren (renchg@sdut.edu.cn)

This work was supported in part by the Youth Innovation Team Development Plan of Shandong Province Higher Education under Grant 2019KJN048.

ABSTRACT Accurately predicting traffic flow is paramount for the efficient operation of transportation systems. The key to enhancing prediction accuracy lies in effectively mining the intricate spatio-temporal correlations within traffic flow data. However, traditional traffic flow prediction methods that combine Graph Convolutional Network and Recurrent Neural Network have limitations in capturing comprehensive spatial correlation information and face challenges in modeling long-term temporal dependencies, consequently leading to suboptimal prediction performance. This study proposes a hybrid traffic flow prediction model based on fusion graph convolutional network and enhanced gate recurrent unit. Initially, a fusion graph structure is constructed based on adjacency graph and adaptive graph to better represent the correlations between nodes in the road network. Subsequently, the stacked fusion graph convolution module is utilized to capture multi-level spatial correlations and the enhanced gated recurrent unit is applied to extract multi-scale temporal correlations. In addition, the model integrates the extracted spatio-temporal features with the direct features through residual connection units, and utilizes the fused features for prediction, achieving superior predictive performance. The experimental results from four authentic datasets demonstrate that our proposed model outperforms state-of-the-art baseline models, showcasing an average enhancement of 3% in Mean Absolute Error(MAE), 3.3% in Root Mean Square Error(RMSE), and 2.7% in Mean Absolute Percentage Error(MAPE) across the four datasets.

INDEX TERMS Traffic flow prediction, spatio-temporal correlation, gated recurrent unit, graph convolution network, residual connect.

I. INTRODUCTION

With the rapid growth of traffic data and the continuous advancement of artificial intelligence technology, numerous cities are working on developing intelligent transportation systems to achieve efficient traffic management, accurate traffic resource allocation, and high-quality traffic services [1]. As a key component of intelligent transportation systems, traffic flow prediction plays an important role in realizing the efficient operation of transportation systems. The sensors deployed in the traffic system collect a large amount of traffic flow data. However, traffic data has a complex time correlation, spatial correlation, spatio-temporal correlation, and heterogeneity. Accurately capturing these correlations between traffic data is crucial to predicting traffic flow. Figure 1 shows the propagation of correlation information of central region A and its related regions in spatio-temporal dimensions.

In Figure 1, the nodes represent the region, and the undirected edges denote the correlation between regions. The different colors of the nodes and edges represent the various traffic conditions and the strength of the correlation

The associate editor coordinating the review of this manuscript and approving it for publication was Ángel F. García-Fernández.

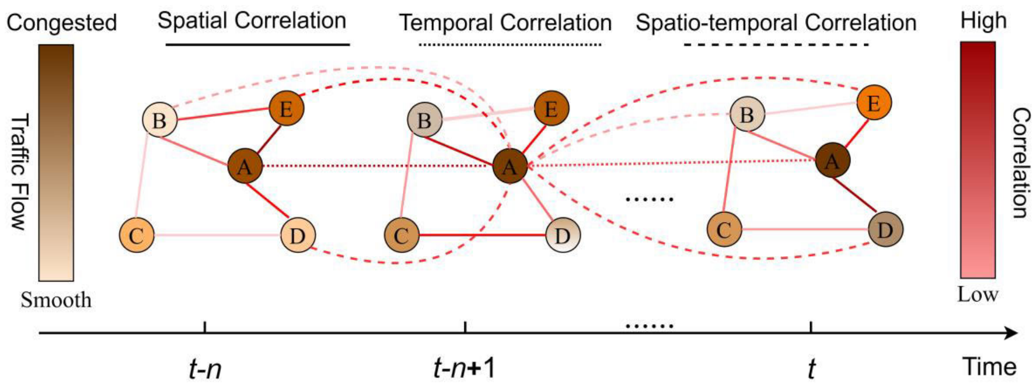


FIGURE 1. Propagation of correlation between different regions.

respectively. The dotted edges represents the temporal correlation between regions. Temporal correlation refers to the traffic states changing dynamically with time, and the traffic status of adjacent time steps are highly similar. Solid edges indicate spatial correlation, which means that traffic conditions between neighboring areas affect each other. It is usually expressed as the traffic flow of upstream roads affecting downstream roads. The dashed edges represent spatio-temporal correlation, indicating that the traffic conditions in a region are influenced by the related areas at different time steps.

Researchers have proposed a series of methods to better exploit the correlation between traffic data for accurate traffic flow prediction. In terms of exploring time correlation, recurrent neural networks (RNNs) and its various variants [2], [3], [4] have been widely applied. Through the RNNs model, the time dependency and sequential patterns in traffic flow data can be effectively captured. However, modeling solely based on temporal features has certain limitations. The methods that combine graph convolutional networks (GCNs) with recurrent neural networks have gradually become a research hotspot [5], [6]. Such methods typically represent the static road network structure as a topological graph and utilize GCNs to learn the spatial correlation between nodes. Simultaneously, RNNs are used to learn the temporal features of traffic flow data. This approach can integrate the spatial structure of the road network with the temporal relationship of traffic flow data, thereby capturing the spatio-temporal characteristics of traffic flow more comprehensively. However, due to the complexity of spatio-temporal correlation, these methods still face two challenges.

First, conventional RNN-based methods cannot capture long-term dependencies well in traffic flow data. RNNs suffer from problems such as gradient vanishing and memory decay in capturing long-term dependencies, making it difficult to effectively maintain and convey information between distant time steps. To address this, methonds [7], [8] introduced long short-term memory(LSTM) for traffic flow prediction, which utilizes a gating mechanism to selectively retain and transmit information, alleviating the problems of gradient vanishing and memory decay. Although traffic flow prediction based on LSTM can achieve good results, its model structure is complex and the training time spent is relatively long. FU et al. [9] employed a structurally simple gated recurrent unit(GRU) for traffic flow prediction and demonstrated that GRU can achieve faster and more accurate traffic flow prediction compared to LSTM. Subsequently, scholars [10], [11] have tried to utilize various variants of GRU for traffic flow prediction and achieved good results. However, it still cannot solve the problem that long-term dependencies are gradually faded out or covered by new information when they are transmitted through hidden state units with limited capacity.

Second, traditional GCN-based methods typically utilize predetermined graph structures established by distance metrics or other geographical connections. such methods are only able to capture the spatial correlation generated by roadway adjacency relationships, and are less capable of modeling other factors that may affect traffic flow. While methods such as AGCRN [12] and MTGNN [13] introduce adaptive graph structures to uncover potential spatial correlation within road network data, and ASTCN [14], En-GRN [15] try to capture different types of dependencies by combining adaptive graph structures with predefined graphs. However, these approaches either discard the fixed topological structure information of the road network and rely solely on adaptively generated graph structures from the model, or treat the adaptive graph as indirect spatial information, separate from the direct spatial information represented by the adjacency graph. This lack of effective fusion between the two types of spatial information leads to suboptimal prediction performance.

In view of the advantages and disadvantages of the above methods, we propose a traffic flow prediction model based on fusion graph convolution and enhanced gated recurrent units(FGCN-EGRU). In the spatial dimension, we utilize a fusion graph to assist the graph convolutional network in extracting spatial features. In the temporal dimension, we employ an enhanced gated recurrent unit for temporal feature extraction. The main contributions of our research are as follows.

A fusion graph structure is designed, which enhances the representation of node relationships by integrating the adaptive graph generated through the adaptive graph learning layer with the adjacency graph based on geographic location. A convolutional neural network based on the fusion graph is used to better explore the spatial correlations between traffic data.

An enhanced gated recurrent unit was designed, which utilizes attention scores and gating mechanisms to better preserve and propagate the correlation information in traffic flow data, addressing the limitation of recurrent neural network in capturing long-term dependencies of traffic flow data. A traffic prediction model was designed based on the above methods, which improves the accuracy of traffic flow prediction.

II. RELATED WORK

A. TIME SERIES INFORMATION FOR TRAFFIC FLOW PREDICTION

In the early days, the time series prediction method based on statistical modeling is very popular. This method is mainly aimed at the time correlation of traffic flow data and uses mathematical statistics to calculate the characteristics of historical data as the general characteristics to predict future data. Typical methods based on statistical modeling include the auto regressive moving average (ARMA) [16], vector auto regressive model (VAR) [17], Kalman filter [18], etc. Most of these prediction methods have simple structures and fast calculation speeds. While these methods are frequently employed in preliminary research phases, they exhibit limited adaptability to the nonlinear attributes of traffic flow. As the prediction time increases, the prediction error also rises sharply, which is only suitable for scenes with stable traffic conditions. Subsequently, machine learning methods have gradually become research hotspots, such as support vector machines [19], K-nearest neighbors [20], and Bayesian networks [21]. This type of method can model nonlinear factors in traffic data and extract more complex correlations. However, these methods require manual feature selection and often perform poorly in traffic prediction tasks with complex features.

In recent years, traffic flow prediction based on deep learning methods has become a new trend. Recurrent neural networks have been more widely used in the field of traffic prediction due to their adeptness in time-series data. Lu [7] employed the variational modal decomposition to decompose unstable raw traffic flow data sequences into multiple stable sub-sequences, and utilized LSTM to capture the stable features of each sub-sequence for prediction, significantly enhancing predictive performance. Redhu and Kumar [8] proposed PSO-Bi-LSTM for short-term prediction, they used a nonlinear variable inertia weights improved particle swarm optimization algorithm to optimize the parameters of the Bi-LSTM, which effectively improves the speed of model training convergence and prediction accuracy. Sun et al. [10] proposed SSGRU, which utilizes stacked GRUs to enhance the model's ability to capture long-term dependencies. However, it still struggles to effectively address the issue of memory decay in GRU networks.

The advantages of these traffic flow prediction methods based on time series information lie in their simplicity, computational efficiency, and suitability for short-term forecasting. Nonetheless, this approach neglects spatial information within traffic flow, limiting its performance in complex scenarios and long-term predictions. Additionally, it struggles to handle data gaps, noise, and nonlinear variations. In summary, time series information-based prediction methods are suitable for straightforward real-time forecasting tasks. Yet, for more accurate predictions of intricate traffic conditions and long-term trends, it is necessary to consider more sophisticated approaches that incorporate spatial information and other factors.

B. SPATIO-TEMPORAL INFORMATION FOR TRAFFIC FLOW PREDICTION

In the traffic flow prediction problem, the spatial correlation between traffic flow data has an important influence on the prediction results. In the early days, many scholars tried to convert traffic network data into two-dimensional raster images, and then use convolutional neural networks to predict raster images. Ma et al. [22] transformed traffic network data into raster images, used a two-dimensional matrix to represent the spatio-temporal correlation of data, and then used two-dimensional convolution to predict. Zhang et al. [23] mapped traffic flow data to grid images, used convolutional neural networks to extract spatial features, and processed spatial feature sequences through long and short-term memory network layers to obtain time features. These method reduces the complexity of modeling and improves the computational efficiency of the model by dividing the traffic network into simplified grid cells, but it loses some of the detailed information in the original data, only shows good performance on data with a relatively simple road network structure, and only captures the static spatio-temporal correlation of the road network data.

Subsequently, graph convolutional networks began to be widely used in traffic flow prediction. Zhao et al. [6] proposed a method that combines graph convolutional networks with improved gated recurrent units (GRUs). They straightforwardly concatenate the improved GRU and GCN to achieve the extraction of spatio-temporal dependencies in the data. Yu et al. [24] proposed STGCN, which learns features of data through multiple spatio-temporal convolution modules that integrate graph convolution and gated temporal convolution, capturing complex spatio-temporal correlations of data using only a simple network structure. Li et al. [25] proposed DCRNN that captures the spatial correlations within traffic data through random walks using diffusion convolutions, and captures the spatio-temporal correlations of traffic data using a GRU embedded with diffusion convolutions. The method fully considers the spatial dependence and temporal characteristics of traffic data, which improves the accuracy of traffic flow prediction. These above methods primarily rely on a singular predefined adjacency graph structure, which is only able to capture the spatial correlation between neighboring locations and can show good performance when dealing with simple road network data.

In order to solve these problems and capture the spatio-temporal correlation between data more comprehensively. AGCRN [11] and MTGNN [12] adaptively generate graph structures by multiplying node embeddings, achieving performance similar to predefined adjacency matrices using only adaptive graphs. However, relying solely on a single adaptive graph structure fails to effectively capture the dynamic correlations in traffic data and overlooks the significant influencing factor of road network structure, which limits the prediction performance of the model. Chang et al. proposed En-GRN [15] to obtain indirect spatial information based on adaptive graph structure and direct spatial information based on the neighboring graphs, and fused these two types of information through the gating mechanism. This method synthesizes different types of spatial information and effectively integrates different information through a gating mechanism, which has a good prediction performance. Chen et al. [26] introduced GSTPRN, which constructed a position graph convolution module based on attention mechanisms, and expanded the receptive field of the convolution module through approximate personalized propagation, then integrated the position graph convolution and adaptive graph learning into a recurrent network for spatio-temporal correlation extraction. The two methods mentioned above make predictions by combining adaptive graphs with location information, enabling them to capture various types of dependencies in traffic data. However, they do not account for the dynamic changes in the correlation of traffic data, which impacts the potential for further improvement in model performance. Wenger et al. [27] proposed DDGCRN, which involves embedding matrices with dynamic signals extracted by multi-layer perceptrons to generate dynamic graph structures. In the model, the optimized graph convolutional networks are embedded within a GRU for spatio-temporal feature extraction, which effectively captures the dynamic spatiotemporal relationships in traffic flow. Xu and Liu [28] constructed a distance matrix using the Dijkstra algorithm and calculated dynamic weight matrices based on Pearson correlation coefficients, then combined adjacency matrices, dynamic weight matrices, and distance matrices into a hybrid adjacency matrix, enhanced the representation of road node relationships, then utilized graph convolutional GRU for spatio-temporal feature extraction, effectively capturing dynamic spatio-temporal dependencies between traffic data.

The aforementioned methods introduce different graph structures to capture different types of relationships and dependencies in traffic data, thus improving the accuracy and robustness of the traffic flow prediction model. However, these methods may rely too much on a particular graph structure and ignore other influencing factors, or adopt complex model architectures to extract information from multiple graph structures, or may not be sufficiently adapted and optimized in the way of fusing information from multiple graphs, thus having certain limitations.

In recent years, researchers have begun to introduce Transformer models into traffic flow prediction tasks and have achieved significant results. STTN [29] and Traffic Transformer [30] used the self-attention mechanism of the Transformer for spatio-temporal feature extraction from traffic data, improving the ability to model long-term dependency relationships. MGT [31] and STGHTN [32] took into account various types of spatial correlations, conducting spatial feature learning on multiple graphs, thereby enhancing the capability to capture spatial features.However, these models have relatively complex structures and require a substantial amount of training data and computational resources to fully unleash their powerful performance.

Inspired by these studies, we propose a traffic flow prediction model (FGCN-EGRU) to better handle traffic flow prediction tasks.

III. PRELIMINARIES

A. TRAFFIC TOPOLOGY

The structure of the actual traffic network is defined as a topology graph $G = (V, E, A)$, where $V = \{V_1, V_2, \ldots V_N\}$ is a set of nodes, representing N road sensor nodes in the traffic network; E is a set of edges, representing connectivity between nodes; $A \in R^{N\times N}$ is an adjacency matrix representing node adjacency relations, whose element $A_{i,j}$ represent connectivity between nodes i and j.

B. DATA REPRESENTATION

The traffic flow value of node i at moment t is expressed as x_t^i, and the traffic flow value of the whole road network at moment t is expressed as $X_t = \{x_t^1, x_t^2, \ldots x_t^N\} \in R^{N\times 1}$. The input traffic flow sequence of a prediction model is expressed as $\chi_T = \{X_1, X_2, \ldots X_T\} \in R^{T\times N\times 1}$, where T is the length of the input history traffic sequence. The output sequence of a traffic prediction model is $\chi_P = \{X_T{+}_1, X_T{+}_2, \ldots X_{T+P}\} \in R^{P\times N\times 1}$, where P is the length of the prediction sequence.

C. PROBLEM DEFINITION

The problem of traffic flow prediction can be modeled as learning a mapping relationship function $f(\cdot)$ from the historical input traffic sequence $\chi_T = \{X_1, X_2, \ldots X_T\} \in R^{T\times N\times 1}$ and the topology graph G to predict the future traffic sequence $\chi_P = \{X_T{+}_1, X_T{+}_2, \ldots X_{T+P}\} \in R^{P\times N\times 1}$. The prediction process can be expressed as formula (1).

$$\{X_T{+}_1, X_T{+}_2, \ldots X_{T+P}\} = f(\{X_1, X_2, \ldots X_T\}; G) \quad (1)$$

IV. METHODOLOGY

A. OVERVIEW

The structure of our proposed FGCN-EGRU is shown in Figure 2. Besides the input and output layers, the FGCN-EGRU model also includes fusion graph construction module, stacked graph convolution module, enhanced gated recurrent unit, and residual connection unit. The input layer converts the input data to a high-dimension space by linear to represent the complex spatial-temporal features. The fusion graph construction module generates a fusion graph by combining the adaptive graph and the adjacency graph. Subsequently, the augmented data and fusion graph are fed into a stacked graph convolutional module to extract multi-level spatial features. The enhanced gated recurrent unit utilizes the augmented data and the spatial features obtained from stacked graph convolutional to extract spatio-temporal features. The residual connection unit combines the directly extracted features from the original input data with the spatio-temporal features. Finally, the output layer utilizes the features fused by the residual connection unit to make predictions and obtain the final prediction results.

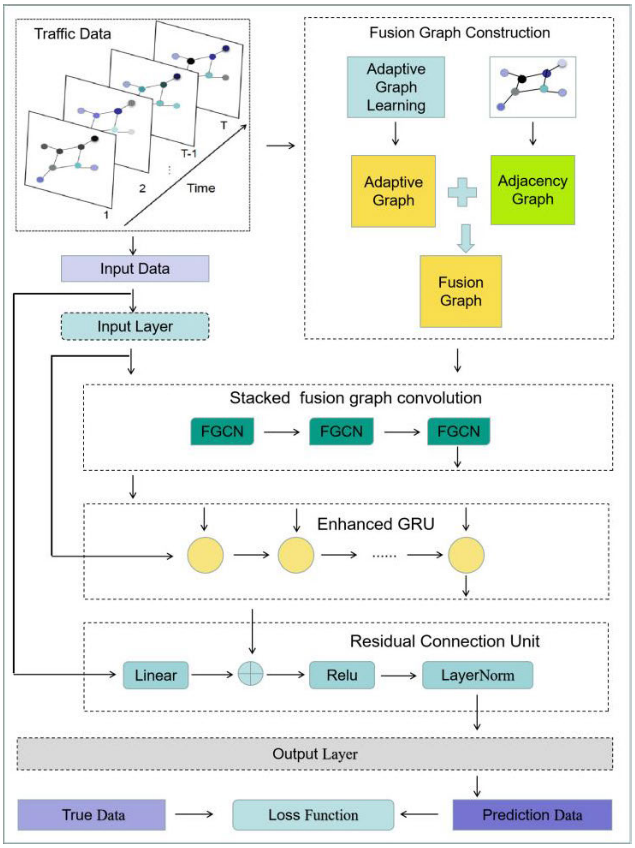


FIGURE 2. **The architecture of the FGCN-EGRU model.**

B. FUSION GRAPH CONVOLUTION

Traffic flow data contains abundant spatial correlation information, and effectively constructing the road network graph structure enables graph convolutional networks to better capture this spatial correlation information. In the traffic network, the relationships between nodes are influenced by various factors, a single adjacency graph structure is not sufficient to effectively represent the relationships between road nodes. Although some methods introduce adaptive graph to explore hidden relationships between nodes, they often treat the adaptive graph as indirect spatial information, fail to effectively integrate the predefined graph structure and the adaptive graph structure.

To enhance the representation of node relationships and better extract spatial correlations between traffic flow data, we propose a fusion graph convolutional network(FGCN). Firstly, we construct a fusion graph structure to enhance the representation of node relationships. Then, graph convolution operations are performed based on the fusion graph structure and Chebyshev approximation to better extract spatial correlations.

The adjacency graph $G_1 = (V, E^s, A_{adj})$ is constructed based on the road connection relationships which given the traffic dataset. The adjacency matrix A_{adj} of the adjacency graph G_1 is expressed as formula (2).

$$A_{adj} = \begin{cases} 1, & (v_i, v_j \in V) \,\&\, \left((v_i, v_j) \in E^s\right) \\ 0, & \text{otherwise} \end{cases} \tag{2}$$

The adaptive graph $G_2 = (V, E^{adp}, A_{cor})$ is generated by the adaptive graph learning layer in a data-driven manner. In the adaptive graph learning layer, two Learnable parameter matrices $P_1 \in R^{N\times d}$ and $P_2 \in R^{d\times N}$ are generated, where N is the number of road network nodes and d is the dimension of the parameter matrix. Then, P_1 and P_2 are multiplied to infer the spatial correlation between nodes, and the correlation matrix A_{cor} of adaptive graph G_2 is generated as formula (3).

$$A_{cor} = softmax\,(ReLU\,(P_1 \cdot P_2)) \tag{3}$$

In formula (3), the $ReLU$ activation function is used to eliminate the negative weight between nodes, and the $softmax$ activation function is used to normalize the matrix.

The fusion graph $G_3 = (V, E^f, A_{fus})$ is obtained by combining the adaptive graph and the adjacency graph. As shown in formula (4), the spatial matrix A_{fus} of the fusion graph is calculated by summing the adjacency A_{adj} and the correlation matrices A_{cor}.

$$A_{fus} = A_{adj} + A_{cor} \tag{4}$$

Guided by the loss function, the adaptive graph matrix is continuously updated during the training process, so that the dynamic correlation between nodes is mined and quantified from the data streams. By combining the adjacency graph and the adaptive graph, the fusion graph can extract the spatial dependency relationship from the time-varying dynamic traffic volume and the static road network structure. So, the fusion graph can better represent the dynamic spatial correlation of traffic status based on a given traffic network structure.

In order to fully leverage the rich representation of node relationships in the fused graph structure, we use a spectral graph convolution network [33] to extract the spatial correlation information of traffic data. The Laplacian matrix of the traffic network is defined as L=D - A_{fus}, where D is the degree matrix of A_{fus}. The Laplacian matrix can be decomposed into

$L = U\Lambda U^T$, Λ is a diagonal matrix composed of eigenvalues of L, and U is a Fourier basis composed of eigenvectors of L.

The spectral convolution operation is shown in equation (5)

$$gc(L, x) = \mathrm{g}_\theta(L)\ x = U\mathrm{g}_\theta(\Lambda)\, U^{\mathrm{T}}x \tag{5}$$

where, gc represents graph convolution operation, g_θ is convolution kernel, x is input data, and $g_\theta(\Lambda)$ is the spectral domain convolution kernel obtained by Laplacian matrix feature decomposition.

In traffic flow prediction tasks, the graph structure of realistic traffic road networks is relatively large, and the calculation of the convolution kernel in the spectral domain by Laplacian matrix decomposition has a great computational cost, which affects the training convergence of the model. Therefore, we use the Chebyshev polynomial approximation [34] of order K to solve this problem.

The recursive formula of chebyshev polynomial is $\mathrm{T}_0(x) = 1, \mathrm{T}_1(x) = x, \mathrm{T}_k(x) = 2xT_{k-1}(x) - T_{k-2}(x)$. Using Chebyshev polynomial approximation to replace the convolutional kernel as shown in equation (6).

$$\mathrm{g}_\theta(\Lambda) = \sum_{\mathrm{k=0}}^{\mathrm{K\text{-}1}} \theta_{\mathrm{k}}\mathrm{T}_{\mathrm{k}}\left(\tilde{\Lambda}\right)$$
$$\tilde{\Lambda} = \frac{2}{\lambda_{\max}}\Lambda - \mathrm{I_N} \tag{6}$$

where, θ_k is the polynomial coefficient, and $\lambda_{\max}$ is the largest eigenvalue. Using polynomial approximation expansion to solve this formula is equivalent to aggregation $0 \sim K-1$ of neighbor information around node by a convolution kernel.

The convolution operation after the Chebyshev approximation substitution is shown in equation (7).

$$gc(L, x) = U\sum_{\mathrm{k=0}}^{\mathrm{K\text{-}1}} \theta_{\mathrm{k}}\mathrm{T}_{\mathrm{k}}\left(\tilde{\Lambda}\right)U^{\mathrm{T}}x = \sum_{\mathrm{k=0}}^{\mathrm{K\text{-}1}} \theta_{\mathrm{k}}\mathrm{T}_{\mathrm{k}}\left(\tilde{L}\right)x \tag{7}$$

The fusion graph convolutional neural network directly performs convolution based on the Laplacian matrix, eliminating the need for feature decomposition and reducing computational complexity. Moreover, this graph convolutional networks based on fusion graphs can better utilize the relationships between nodes, thereby capturing spatial correlations between data more effectively.

C. STACKED FUSION GRAPH CONVOLUTION

The traffic network has a complex road connection relationship, and there is a wide range of spatial dependencies between road network data. When congestion occurs in some sections of the traffic network, it will not only affect the traffic conditions of nearby sections but also affect the traffic conditions in the distance. This indicates that there are short and long-range spatial dependencies in traffic flow data. How to better capture the multi-level spatial dependence relationship in road network data is important to improve the performance of traffic flow prediction. Since our fusion graph convolutional network can only extract $K-$ order neighbor information, it cannot well mine the long-range dependence of road network data, and as K gradually increases, the local feature extraction ability of the network will become weak.

To capture the multi-level spatial dependence, a stacked fusion graph convolution module is constructed. This module stacks three graph convolutional layers together, progressively passing and updating node features layer by layer, thereby gradually extracting higher-level spatial correlation information. The stacked graph convolution module can be represented as formula (8).

$$X_{mgc} = gc\left(\tilde{L}, gc\left(\tilde{L}, gc\left(\tilde{L}, x\right)\right)\right) \tag{8}$$

D. ENHANCED GATED RECURRENT UNIT

There are not only multi-level spatial correlations between traffic network data but also multi-scale temporal correlations. For example, the traffic state at a certain moment will not only affect the traffic state at the next neighbor moment but also affect the traffic state at multiple subsequent moments. Capturing the spatial correlation between traffic flow data through graph convolutional networks alone is not sufficient for accurate traffic flow prediction. Therefore, we introduce recurrent neural networks combined with graph convolution for spatio-temporal correlation extraction. However, traditional recurrent neural networks suffer from the issues of vanishing or exploding gradients, and when propagating information across time steps with a limited capacity of hidden state units, they tend to forget early relevant information. This leads to the network having difficulties in effectively capturing long-term temporal correlations between traffic flow data.

In order to better capture the correlation information among traffic flow data, we design an enhanced gated recurrent unit(EGRU). By introducing the gating mechanism, the flow of information is effectively controlled, thus mitigating the effects of gradient vanishing and gradient explosion. At the same time, an enhanced attention mechanism is used to focus and update the early correlation information, reducing its forgetting during the time-step propagation. The structure of the enhanced gated recurrent unit is shown in Figure 3.

The inputs of the enhanced gated recurrent unit are H_{t-1}, X_{mgc}, χ'_T, where, H_{t-1} is the hidden state at moment $t-1$, which holds the information of the spatio-temporal features extracted at moment t-1. X_{mgc} is the output of the stacked graph convolution, representing the spatial features extracted by the stacked graph convolution. χ'_T is the input data after dimension elevation, with the original temporal features of the data. In the enhanced gated recurrent unit, the reset gate r is calculated using X_{mgc}, allowing more relevant information from the previous state to be retained in the next state. The attention scores play a crucial role in combining X_{mgc} and H_{t-1}, leading to the creation of X_{st}, which is subsequently employed to calculate the update gate z. This update gate governs the degree to which the previous state influences the current state.

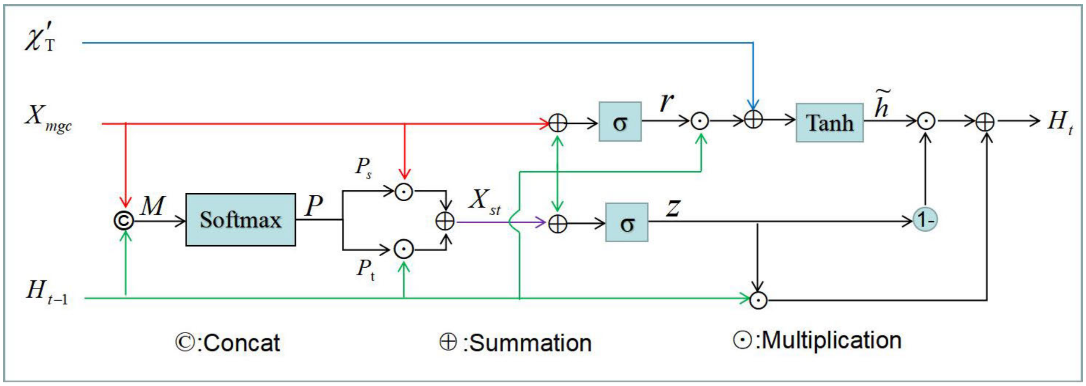


FIGURE 3. The architecture of enhanced gate recurrent unit.

The specific computation process for calculating the attention scores to merge X_{mgc} and H_{t-1} into X_{st}, which contains rich spatio-temporal features, is shown as follow:

Splicing the output X_{mgc} of the stacked graph convolution with the previous hidden state H_{t-1} to obtain M as formula (9).

$$M = X_{mgc}||H_{t-1} \tag{9}$$

The attention scores of X_{mgc} and H_{t-1} are calculated using the softmax function as follows:

$$P = softmax\ (M)$$
$$P_s, P_t = split\ (P) \tag{10}$$

where, *split* () is the tensor splitting operation, P_s and P_t are the attention scores matrices of X_{mgc} and H_{t-1}.

Combining X_{mgc} and H_{t-1} using the attention scores obtain X_{st} as formula (11).

$$X_{st} = P_s \cdot X_{mgc} + P_t \cdot H_{t-1} \tag{11}$$

The X_{st} integrates the temporal and spatial features in the road network data. Using it to compute the update gate allows more spatio-temporal features from the previous state to be copied into the new state, reducing the omission of relevant information during the gate update process.

The process of calculating the gate unit using X_{mgc} and X_{st}, and updating the hidden state with χ_T', is as follows.

$$r = \sigma\left(W_r \cdot X_{mgc} + U_r \cdot H_{t-1}\right)$$
$$z = \sigma\left(W_z \cdot X_{st} + U_z \cdot H_{t-1}\right)$$
$$\tilde{h} = \tanh\left(W_h \cdot \chi_{\mathrm{T}}' + r \cdot U_h H_{(t-1)}\right)$$
$$H_t = z * H_{(t-1)} + (1-z) * \tilde{h} \tag{12}$$

where, h is the candidate hidden unit, H_{t} is the current moment's hidden state, σ is the sigmode activation function, the W and U are learnable weight parameters.

The enhanced gated recurrent unit obtains spatio-temporal features through the original input, the output of the graph convolution, and the hidden units. It adopts a novel approach to compute the gate mechanism, effectively avoiding gradient issues and reducing the loss of correlation information during the time step propagation. Additionally, the network eliminates the bias terms in the update formula, reducing the number of parameters.

E. RESIDUAL CONNECTION UNIT

Ignoring the direct useful information in the original input data and relying solely on the spatio-temporal features extracted by the enhanced gated recurrent unit cannot achieve accurate predictions. Moreover, as the multiple different network stacks in the model, more and more activation functions are introduced, and the model begins to degrade and has problems such as gradient dispersion, resulting in poor prediction performance.

To solve the gradient vanishing and better utilize the features in the original data, we introduce a residual connection unit to fuse the origin input data with the extracted spatio-temporal features. As shown in Figure 1, the residual connection unit has two input branches, one is the original input data χ_T, and the other is the data processed by the gated recurrent unit H_t. The residual connections unit employs a linear layer to extract direct features from the original input data. These direct features are then added and fused with the spatio-temporal features. Subsequently, the fused features undergo *relu* activation function and *LayerNorm*. Utilizing these fused features enables more accurate prediction of future traffic flow data. Additionally, with the introduction of residual connections, gradients can propagate directly through cross-layer connections during backpropagation, avoiding the vanishing gradient issue.

The residual connection unit can be defined as formula (13).

$$X_{rc} = LayerNorm\left(relu\left(Linear\ (\chi_T) + relu\ (H_t)\right)\right) \tag{13}$$

We take the output of the residual connection unit X_{rc} as the final output and use the fully connected layer in the output layer to convert the output into the final prediction result χ_p. The calculation process is shown in equation (14).

$$\chi_p = X_{rc} W_o + b_o \tag{14}$$

where, W_o and b_o are weight matrices.

V. EXPERIMENT

A. DATASETS

We selected four benchmark datasets PEMS03, PEMS04, PEMS07, and PEMS08 in the field of traffic flow prediction for experiments. They are from the PEMS (Caltrans Performance Measurement System) traffic flow monitoring system. The system deployed 39,000 detectors in major areas of California and collected data every 30 seconds. The collected traffic characteristics include traffic flow, road occupancy, and traffic speed. The data set aggregates the original data every five minutes, containing 288 data points per day. The specific information of the dataset is shown in the following table.

TABLE 1. Details of four datasets.

Dataset	Nodes	Edges	Time range	Time steps
PEMS03	358	547	9/1/2018-11/30/2018	26208
PEMS04	307	340	1/1/2018-2/28/2018	16992
PEMS07	883	866	5/1/2017-8/31/2017	28224
PEMS08	170	295	7/1/2016-8/31/2016	17856

B. EVALUATION METRICS

The experiment selects three commonly used evaluation metrics in the research field, including Mean Absolute Error (MAE), Mean Absolute Percentage Error (MAPE), and Root Mean Square Error (RMSE). The specific calculation formulas are as follows.

$$MAE = \frac{1}{MN}\sum_{i=1}^{M}\sum_{j=1}^{N}\left|y_i^j - \hat{y}_i^j\right|$$
$$MAPE = \frac{1}{MN}\sum_{i=1}^{M}\sum_{j=1}^{N}\left|\frac{y_i^j - \hat{y}_i^j}{y_i^j}\right|$$
$$RMSE = \sqrt{\frac{1}{MN}\sum_{i=1}^{M}\sum_{j=1}^{N}\left(y_i^j - \hat{y}_i^j\right)^2} \tag{15}$$

In the above formula, y_i^j and $\hat{y}_i^j$ is the real value and the predicted value of node i at time j, M is the number of nodes in the road network, and N is the time step length of the sequence. The lower the value of the three evaluation indexes, the better the prediction performance of the model.

C. BASELINE METHODS

Compare the proposed model with the following representative models:

Chebynet [34]: An improved graph convolutional neural network, which uses Chebyshev polynomials to perform graph convolution operations, can effectively learn the representation of graphical data.

GRU [35]: An improved recurrent neural network, which uses an update gate and a reset gate to determine which information can eventually be used as the output of the gated recurrent unit.

STLGRU [36]: A spatio-temporal graph gated recurrent unit, which uses memory-enhanced attention and gating mechanisms to capture local and global spatio-temporal relationships of traffic data.

STSGCN [37]: A spatio-temporal network prediction model, which connects the spatial graphs of multiple neighbor time steps, and uses the graph convolution method to capture the local spatio-temporal correlation.

STGODE [38]: A spatio-temporal graph ordinary differential equation network, which captures dynamic spatio-temporal correlation through a tensor-based ordinary differential equation (ODE), and uses time-expanded convolution structure to capture time features.

STFGNN [39]: A spatio-temporal fusion graph neural network, which constructs a spatio-temporal graph by fusing multiple graphs and learns the spatio-temporal correlations using stacked gate convolution modules.

STDSGNN [40]: A dynamic fusion graph neural network, which integrates multiple spatial information through dynamic weighted fusion, and utilizes stacked injection structures to extract spatio-temporal features across multiple layers.

STGPCN(Kronecker) [41]: A spatio-temporal graph convolutional neural network, which combines graph product methods to construct graph structures and utilizes graph product convolutional networks to extract spatio-temporal features.

D. EXPERIMENT AND PARAMETER SETTINGS

The experiments were conducted in a Python environment using the Pytorch 1.12.1 framework. The CPU of the device used is Intel(R) Core(TM)i7-13700KF, and the GPU is NVIDIA GeForce RTX4090. The dataset was divided into a training set, validation set, and testing set in a ratio of 6:2:2. We train the model on a training set, optimize the model using a validation set, and finally use a testing set to evaluate the performance of the model. The data was normalized with zero-mean normalization. This study uses one hour of historical traffic flow data to predict the next hour of traffic flow data.

To achieve better model performance, we employed a manual tuning strategy to adjust the main parameters of the model on the validation set. The adjustments included: bach size = [8, 16, 32, 64], optimization = [Adam, Nadam, RMSProp, RAdam], learning rate = [0.1, 0.01, 0.001, 0.0001], the order of Chebyshev polynomials = [1, 2, 3, 4, 5],embedding dimension d = [5, 10, 15,20, 25], and elevated dimension F = [16, 32, 64, 128, 256]. After conducting experiments, we set the batch size to 32, chose NAdam as the model optimizer, set the learning rate to 0.001, and configured the order of Chebyshev coefficient to be 3. The embedding dimension and elevated

TABLE 2. The performance of the different models on the four datasets.

	PEMS03			PEMS04			PEMS07			PEMS08		
Methods	**MAE**	**RMSE**	**MAPE(%)**	**MAE**	**RMSE**	**MAPE(%)**	**MAE**	**RMSE**	**MAPE(%)**	**MAE**	**RMSE**	**MAPE(%)**
Chebynet	36.47	54.51	52.09	52.65	70.84	43.38	60.4	79.26	37.19	50.71	68.5	45.08
GRU	20.47	34.67	20.63	26.46	41.12	18.23	29.38	44.98	12.96	21.37	33.24	13.70
STSGCN	17.48	29.21	16.78	21.19	33.65	13.9	24.26	39.03	10.21	17.13	26.86	10.96
STLGRU	16.75	27.04	15.88	21.05	33.44	13.88	23.06	37.07	9.99	16.83	26.35	10.74
STGPCN (Kronecker)	17.11	28.99	16.48	20.96	33.35	13.78	24.02	38.77	10.08	16.41	25.60	10.43
STDSGNN	16.12	25.59	16.15	20.67	32.40	13.83	22.91	34.95	10.06	16.73	25.59	10.84
STGODE	16.5	27.84	16.69	20.84	32.82	13.77	22.99	37.54	10.14	16.81	25.97	10.62
STFGNN	16.77	28.34	16.3	**19.83**	31.88	**13.02**	22.07	35.8	**9.21**	16.64	26.22	10.6
FGCN-EGRU	**15.51**	**26.39**	**14.59**	19.97	**31.82**	13.28	**22.02**	**34.97**	9.31	**16.05**	**25.26**	**10.24**

dimension were determined as 10 and 64, with details of the choices visible in the experimental section.

E. PREDICTION PERFORMANCE COMPARISON

On four datasets, FGCN-EGRU and eight baseline models are used to predict traffic flow data for the next hour. The prediction performance of different models is shown in Table 2. Compared with other methods, FGCN-EGRU achieves the best prediction performance on PEMS03 and PEMS08 datasets. The evaluation index MAE and MAPE on the PEMS04 dataset are slightly poorer than the model STFGNN, and only the MAPE on the PEMS07 dataset is slightly higher than the model STFGNN. Overall, our model shows relatively good prediction performance in four datasets.

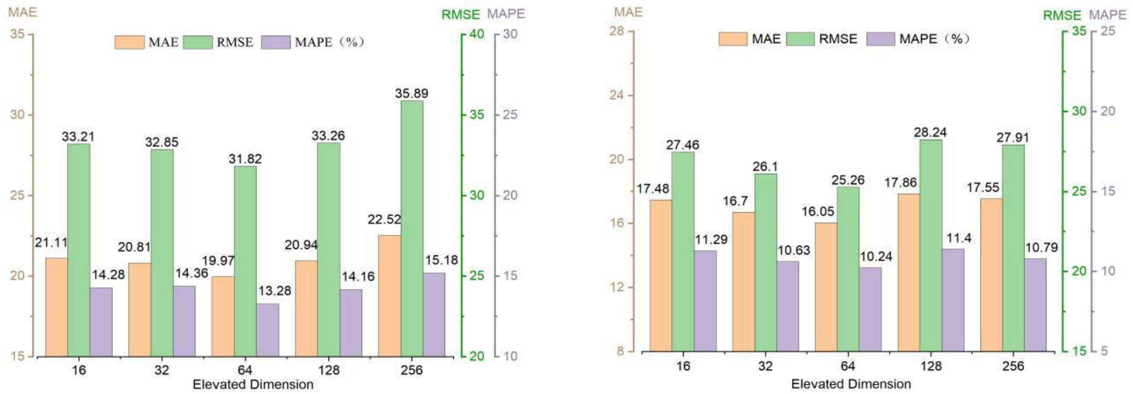
The experimental results are analyzed from four aspects: time correlation, spatial correlation, spatio-temporal correlation, and heterogeneity of traffic flow. The model GRU, which only considers temporal correlation, has weak prediction performance. GRU can use gating mechanisms to capture temporal correlation in traffic flow data and combine relevant historical information with current traffic flow data for prediction. However, since this model does not exploit spatial correlation in traffic flow data, its prediction performance is less than ideal. The model Chebynet, which only considers spatial correlation, has poor prediction performance. Chebynet combines spatial topology with traffic flow data for convolution and can extract some spatial correlation in traffic flow data for prediction. However, because it does not fully extract spatial correlation in data and ignores temporal correlation, it performs poorly in traffic flow prediction tasks. Models based on spatio-temporal correlation, such as STLGRU, STSGCN, STDSGNN, and STGPCN (Kronecker), have better prediction performance than Chebynet and GRU models. This is because the change in traffic flow is influenced by both time and space dimensions, and these methods use multiple convolutions or combine gating mechanisms to exploit the complex spatio-temporal correlation in traffic flow data for prediction, achieving good results. STDSGNN and STGPCN (Kronecker) perform convolution based on multi-graph information, which allows them to better capture spatial correlations in the data, resulting in superior predictive performance compared to STLGRU and STSGCN. Models considering heterogeneity, such as STGODE and STFGNN, have better prediction performance than STDSGNN and STGPCN (Kronecker) because the spatio-temporal correlation in traffic flow data is dynamic. These two models use some modules to exploit the dynamic spatio-temporal correlation in traffic flow data for prediction and achieve good results. FGCN-EGRU has better prediction performance than STGODE and STFGNN, mainly because the stacked fusion graph convolutional layers and enhanced gated recurrent unit in our model can capture spatio-temporal correlation in data effectively.

F. ABLATION STUDY

In order to optimize the model structure and verify the validity of the model components. Experiments were conducted on the PEMS04 and PEMS08 datasets to explore the effects of different hyper parameters, different components, and different graph matrices on the model prediction performance.

1) EFFECT OF HYPER PARAMETER

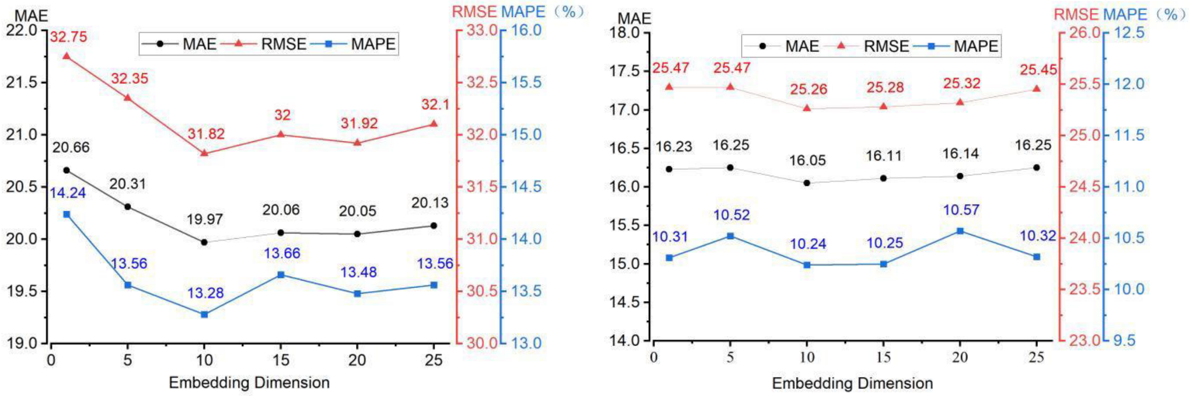
The dimension F of the input layer elevated data has a significant impact on the predictive performance of the model. Either too high or too low a dimension is



(a) On the PEMS04 dataset

(b) On the PEMS08 dataset.

FIGURE 4. MAE,RMSE and MAPE of different elevated dimensions on two datasets.



(a)On the PEMS04 dataset

(b) On the PEMS08 dataset

FIGURE 5. MAE,RMSE and MAPE of different embedding dimension on two datasets.

detrimental to the performance of model. In order to find the optimal elevated dimension, the elevated dimension F is set to [16,32,64,128,256] for the experiments respectively, and the experimental results are shown in Figure 4.

From the evaluation index of prediction performance, it can be seen that with the increasing dimension, the prediction performance of the model first increases and then decreases. When the dimension is 64, the model performance is relatively good. The reason for this phenomenon is that the higher dimension can enhance the expression ability of the data and improve the fitting ability of the model. However, too high a dimension will cause the model to learn abnormal sample data characteristics, resulting in overfitting.

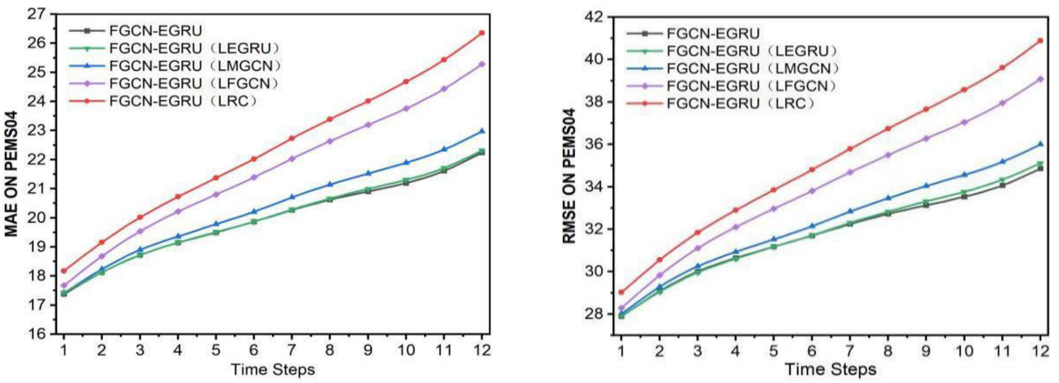
The dimension d of the parameter matrix is of great significance for exploring the spatio-temporal correlation between road nodes. Too large or too small a matrix dimension will affect the quality of the generated correlation matrix. The embedding dimension d is set to [1,5,10,15,20,25] for the experiments respectively, and the experimental results are shown in Figure 5.

As the embedding dimension increases, the prediction performance of the model first increases and then decreases. When the dimension is 10, the model performance is relatively good. This is because the small embedding dimension will limit the representation of node information so that the model cannot infer the accurate spatial correlation from the limited node information. As the embedding dimension increases, the node information is enriched, and the correlation matrix derived by the model is more accurate. The prediction performance is better, but as the embedding dimension continues to increase, the number of model parameters also increases rapidly, model optimization becomes difficult, and the prediction performance is weakened.

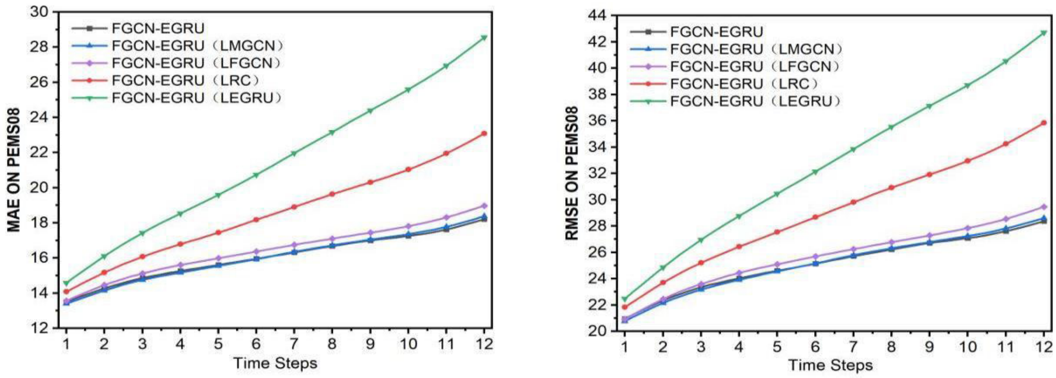
According to the above two sets of experimental results, the number of elevated dimensions of the model is set to 64, and the embedding dimension is set to 10.

2) EFFECT OF DIFFERENT COMPONENTS

FGCN-EGRU consists of four main units: fusion graph convolution neural network, stacked graph convolution module, enhanced gated recurrent unit, and residual connection unit. To verify the validity of each unit of the model, we designed four variant models for ablation experiments on PEMS04 and PEMS08.



(a)The predicted results of different model variants on PEMS04



(b)The predicted results of different model variants on PEMS08

FIGURE 6. Performance of different model variants at different time steps.

FGCN-EGRU(LRC): removes the residual connection unit from the model.

FGCN-EGRU(LFGCN): uses a traditional graph convolution neural in the model instead of the fusion convolution neural to capture spatial features.

FGCN-EGRU(LMGCN): removes the stacked graph convolution layers and uses only a single layer of graph convolution for spatial feature extraction.

FGCN-EGRU(LEGRU): uses conventional gated recurrent units in the model instead of enhanced gated recurrent units to capture temporal features.

The experimental results are shown in Fig. 6. The FGCN-EGRU model achieves the best prediction performance on both datasets, indicating that the model framework can effectively integrate different units. The FGCN-EGRU(LRC) model has the worst prediction performance as a whole, because the residual connection unit can solve the problem of model optimization caused by multi-layer network stacking, after removing the residual connection unit, this problem will seriously affect the prediction performance of the model. The prediction performance of the FGCN-EGRU(LMGCN) model is slightly lower than that of the FGCN-EGRU, which reflects that the stacked graph convolutional network can capture multi-level spatial correlation and improve the prediction performance.

The FGCN-EGRU(LEGRU) model has a very poor prediction performance on the PEMS08 dataset, because the number of road network nodes in the PEMS08 dataset is relatively small, and the spatial correlation between data is weak. At this time, the model mainly relies on mining the time correlation between data streams for prediction. After the lack of enhanced gated recurrent units, the model cannot capture the time correlation between data well, and the prediction performance is poor. FGCN-EGRU(LEGRU) has a similar prediction performance to the FGCN-EGRU model on the PEMS04 dataset, because the number of road network nodes in the PEMS04 datasets is relatively large, and the spatial correlation between data is strong. Using the stacked fusion graph convolution block in the model to capture the rich spatial correlation between data can achieve good prediction accuracy. The prediction performance of the FGCN-EGRU(LFGCN) model on the PEMS04 data set is poor, and the prediction performance on the PEMS08 data set is slightly lower than that of the FGCN-EGRU, because the model does not capture the complex spatial correlation in the PEMS04 dataset well using traditional graph convolutional networks, the prediction performance is weak. The spatial correlation of the PEMS08 data set is relatively simple, using the enhanced gated recurrent unit to obtain time features can achieve good prediction.

TABLE 3. Prediction results of different adjacency matrices on two datasets.

Datasets	Metrics	I	A_{adj}	A_{cor}	A_{fus}
PEMS04	MAE	22.97	21.92	21.04	19.97
	RMSE	36.45	34.55	33.20	31.82
	MAPE（%）	15.45	14.63	14.00	13.28
PEMS08	MAE	17.76	16.83	16.05	16.05
	RMSE	28.16	26.28	25.37	25.26
	MAPE（%）	11.14	10.56	10.18	10.24

3) EFFECT OF DIFFERENT GRAPH MATRIX

This part explores the influence of four different graphs on the prediction performance. Table 3 shows the prediction evaluation index scores of the model using unit matrix I, adjacency matrix A_{adj}, correlation matrix A_{cor}, and fusion matrix A_{fus} at different time steps. When using unit matrix I, the model does not obtain the spatial correlation information of traffic flow data, but FGCN-EGRU still has good prediction performance on two data sets, which shows the effectiveness of our proposed framework. When using the adjacency matrix A_{adj}, the model obtains the spatial correlation generated by the spatial connection relationship, and the prediction performance has been improved. The prediction performance of the model using the correlation matrix A_{cor} is better than that using the adjacency matrix A_{adj}, because the node correlation information contained in the A_{cor} is more comprehensive than A_{adj}. The fusion matrix A_{fus} combines the adjacency matrix A_{adj} and the correlation matrix A_{cor}, which contains abundant spatial correlation information and has better prediction performance in the model.

4) COMPUTATIONAL COMPLEXITY ANALYSIS

To further assess FGCN-EGRU, we compared the model and its different variants in terms of parameter count, training time, and model performance on the PEMS08 dataset (as shown in Table 4).

GCN-GRU: The base model for FGCN-EGRU, sharing the same model structure as FGCN-EGRU. FGCN-EGRU improves upon GCN-GRU by replacing the GCN with FGCN and replacing GRU with EGRU, achieving optimal model performance.

FGCN-GRU: In FGCN-GRU, FGCN is introduced to replace GCN for spatial feature extraction while retaining the GRU for temporal feature extraction.

GCN-EGRU: In GCN-EGRU, EGRU is introduced to replace GRU for temporal feature extraction while retaining the GCN for spatial feature extraction.

GCN-GRU employs traditional graph convolutional networks and GRU for spatiotemporal feature extraction, with minimal parameter count and training time, but lower predictive performance. With the introduction of FGCN, FGCN-GRU showed some increases in parameter count and training time, as FGCN combines adaptive graph structures with adjacency graph structures for spatial feature extraction, thereby adding to parameter count and training time. Upon introducing EGRU, GCN-EGRU exhibited a significant increase in parameter count, primarily because EGRU utilizes attention scores to combine spatial and temporal information for update gate computation. The introduction of attention mechanisms led to an increase in parameter count and training time. FGCN-EGRU, with the incorporation of FGCN and EGRU, experienced some increases in model parameter count and training time, but displayed significant improvements in predictive performance. Considering the balance between computational cost and performance enhancement, the computational complexity of FGCN-EGRU remains reasonable.

TABLE 4. Comparison of parameters, training time, and model performance of different models.

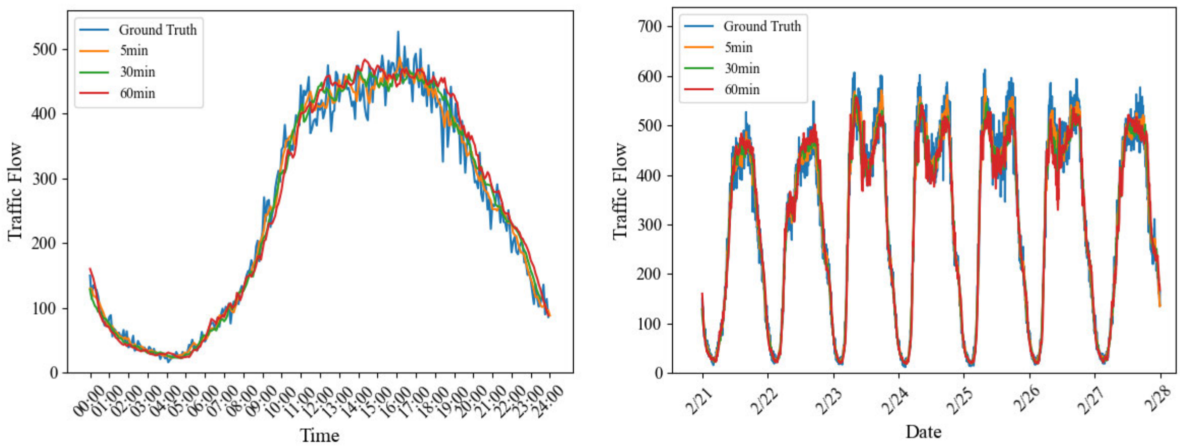
Model	Number of Parameters	Training Time(s/100epoch)	Metric（MAE）
GCN-GRU	104685	2503.37	21.52
FGCN-GRU	108085	2805.79	21.47
GCN-EGRU	130016	3486.22	16.46
FGCN-EGRU	133416	3930.86	16.05

G. VISUALIZATION

1) PREDICTION RESULTS VISUALIZATION

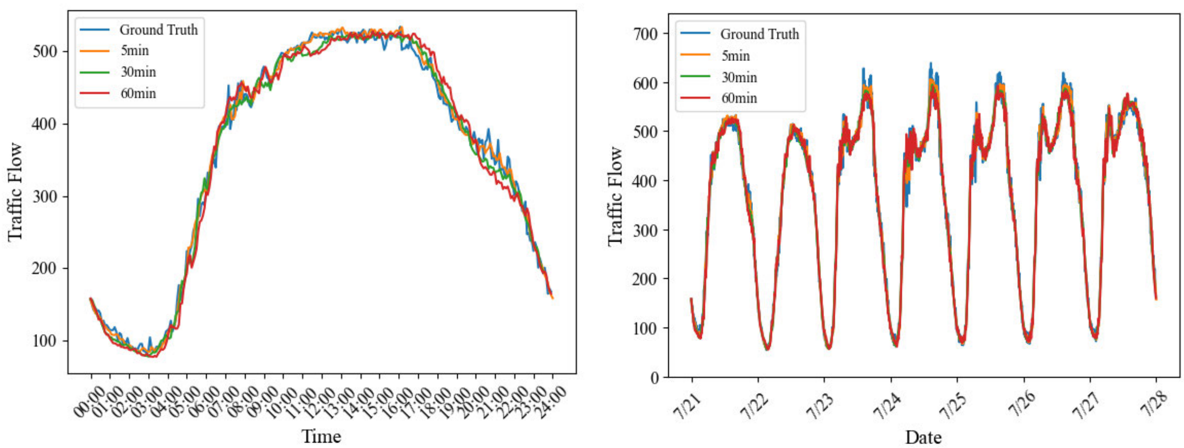
To better showcase the model's prediction performance, we visualized the traffic flow values of Section 123 from the PeMS04 dataset and Section 80 from the PeMS08 dataset. The visualization included real traffic flow values of the road and the model's traffic flow predictions for 5 minutes, 30 minutes, and 60 minutes. For these two roads, we visually display the traffic flow of one day and one week respectively, and the results are shown in Figure 7 and Figure 8. It can be seen from the figure that the traffic flow data curves predicted in different time domains are very close to the real data curves, which indicates that the model can capture the traffic flow

IEEE Access



(a) One-day traffic flow at road 123 (b) One-week traffic flow at road 123

FIGURE 7. **Visualization of traffic flow on the PEMS04 dataset.**



(a) One-day traffic flow at road 80 (b) One-week traffic flow at road 80

FIGURE 8. **Visualization of traffic flow on the PEMS08 dataset.**

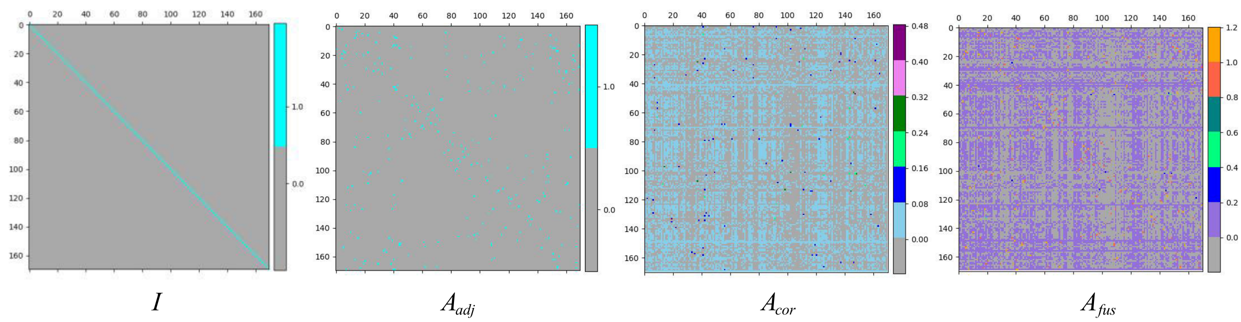


FIGURE 9. **Heatmap of four kinds of adjacency matrix on PEMS08.**

characteristics under the real road network and make accurate predictions.

2) DIFFERENT GRAPH MATRIX VISUALIZATION

To better validate the effectiveness of our proposed fusion graph. Figure 9 and Figure 10 visualize the unit matrix I, adjacency matrix A_{adj}, correlation matrix A_{cor}, and fusion matrix A_{fus} on the PEMS04 and PEMS08 datasets. Hotspots of different colors represent the correlation weights between different nodes. It can be seen from the graph that the non-zero elements of the I on the two datasets are concentrated on the diagonal, which ignores the correlation between

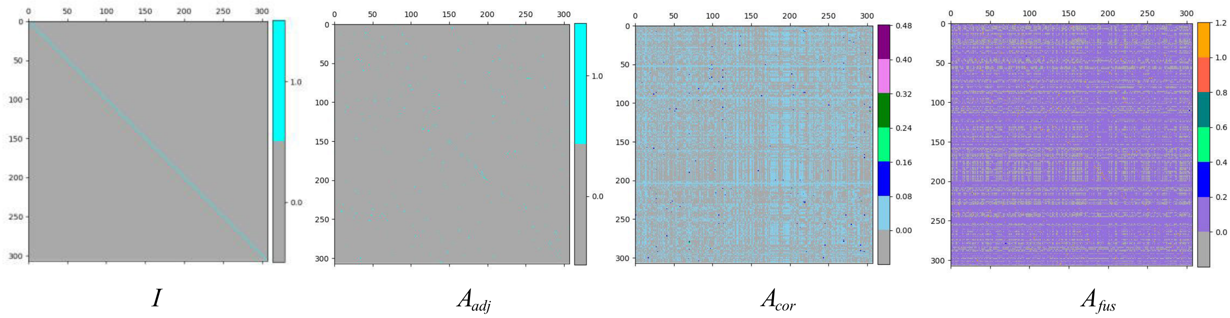


FIGURE 10. Heatmap of four kinds of adjacency matrix on PEMS04.

different nodes. The A_{adj} is sparse, and only a small amount of data is non-zero, which means that the matrix only captures the node relationships from the spatial adjacency. The A_{cor} contains more non-zero elements than the A_{adj}, indicating that the correlation matrix obtained by the data-driven method contains richer node relationship information. The A_{fus} is dense than the A_{cor}, and the overall element value is greater than the correlation matrix, which reflects that the fusion correlation matrix has more perfect spatial node relationships.

VI. CONCLUSION

In this paper, a new traffic flow prediction model based on fusion graph convolution and enhanced gated recurrent units is proposed. The model uses the fusion graph convolutional network combined with enhanced gated recurrent unit to extract temporal and spatial features, and fuses the extracted spatio-temporal features with the direct features through the residual connect unit for flow prediction. We conducted ablation experiments on partial hyper parameters, different adjacency matrices, and some model variants to select the optimal model structure. The results of rich experiments on four real data show that the prediction performance of the FGCN-EGRU model is better than other comparison models.

We visualized the prediction results to demonstrate the accuracy of the predictions. And the different graph matrices are visualized too, which shows that the fusion graph proposed by us has better spatial correlation information.

In future work, we will consider constructing an abundant graph structure, designing a better multi-graph fusion method, and combining weather information, holidays, and other external factors to further improve the accuracy of the model prediction.

REFERENCES

[1] M. R. Jabbarpour, H. Zarrabi, R. H. Khokhar, S. Shamshirband, and K.-K.-R. Choo, "Applications of computational intelligence in vehicle traffic congestion problem: A survey," *Soft Comput.*, vol. 22, no. 7, pp. 2299–2320, Apr. 2018.

[2] K. Cho, C. Gulcehre, D. Bahdanau, F. Bougares, and H. Schwenk, "Learning phrase representations using RNN encoder–decoder for statistical machine translation," 2014, *arXiv:1406.1078*.

[3] D. Kang, Y. Lv, and Y.-Y. Chen, "Short-term traffic flow prediction with LSTM recurrent neural network," in *Proc. IEEE 20th Int. Conf. Intell. Transp. Syst. (ITSC)*, Oct. 2017, pp. 1–6.

[4] J. Zheng and M. Huang, "Traffic flow forecast through time series analysis based on deep learning," *IEEE Access*, vol. 8, pp. 82562–82570, 2020.

[5] L. Zhao, Y. Song, C. Zhang, Y. Liu, P. Wang, T. Lin, M. Deng, and H. Li, "T-GCN: A temporal graph convolutional network for traffic prediction," *IEEE Trans. Intell. Transp. Syst.*, vol. 21, no. 9, pp. 3848–3858, Sep. 2020.

[6] Y. Zhao, X. Han, and X. Xu, "Traffic flow prediction model based on the combination of improved gated recurrent unit and graph convolutional network," *Frontiers Bioeng. Biotechnol.*, vol. 10, Feb. 2022, Art. no. 804454.

[7] J. Lu, "An efficient and intelligent traffic flow prediction method based on LSTM and variational modal decomposition," *Meas., Sensors*, vol. 28, Aug. 2023, Art. no. 100843.

[8] P. Redhu and K. Kumar, "Short-term traffic flow prediction based on optimized deep learning neural network: PSO-Bi-LSTM," *Phys. A, Stat. Mech. Appl.*, vol. 625, Sep. 2023, Art. no. 129001.

[9] R. Fu, Z. Zhang, and L. Li, "Using LSTM and GRU neural network methods for traffic flow prediction," in *Proc. 31st Youth Academic Annu. Conf. Chin. Assoc. Autom. (YAC)*, Nov. 2016, pp. 324–328.

[10] P. Sun, A. Boukerche, and Y. Tao, "SSGRU: A novel hybrid stacked GRU-based traffic volume prediction approach in a road network," *Comput. Commun.*, vol. 160, pp. 502–511, Jul. 2020.

[11] Q. Li, R. Cheng, and H. Ge, "Short-term vehicle speed prediction based on BiLSTM-GRU model considering driver heterogeneity," *Phys. A, Stat. Mech. Appl.*, vol. 610, Jan. 2023, Art. no. 128410.

[12] L. Bai, L. Yao, C. Li, X. Wang, and C. Wang, "Adaptive graph convolutional recurrent network for traffic forecasting," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 33, 2020, pp. 17804–17815.

[13] Z. Wu, S. Pan, G. Long, J. Jiang, X. Chang, and C. Zhang, "Connecting the dots: Multivariate time series forecasting with graph neural networks," in *Proc. 26th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, California, CA, USA, Aug. 2020, pp. 753–763.

[14] L. Zhao, Y, Zhang, and Z. Zhang, "Adaptive spatial–temporal convolution network for traffic forecasting," in *Proc. KSEM*, 2022, pp. 287–299.

[15] L. Chang, C. Ma, K. Sun, Z. Qu, and C. Ren, "Enhanced road information representation in graph recurrent network for traffic speed prediction," *IET Intell. Transp. Syst.*, vol. 17, no. 7, pp. 1434–1453, Feb. 2023.

[16] J. Klepsch, C. Klüppelberg, and T. Wei, "Prediction of functional ARMA processes with an application to traffic data," *Econometrics Statist.*, vol. 1, pp. 128–149, Jan. 2017.

[17] Z. Z. Wang, A. Safikhani, Z. Y. Zhu, and D. S. Matteson, "Regularized estimation in high-dimensional vector auto-regressive models using spatio-temporal information," 2020, *arXiv:2012.10030*.

[18] X.-S. Trinh, D. Ngoduy, M. Keyvan-Ekbatani, and B. Robertson, "Incremental unscented Kalman filter for real-time traffic estimation on motorways using multi-source data," *Transportmetrica A, Transp. Sci.*, vol. 18, no. 3, pp. 1127–1153, Dec. 2022.

[19] T. Devi, K. Alice, and N. Deepa, "Traffic management in smart cities using support vector machine for predicting the accuracy during peak traffic conditions," *Mater. Today, Proc.*, vol. 62, pp. 4980–4984, Jan. 2022.

[20] G. Lin, A. Lin, and D. Gu, "Using support vector regression and K-nearest neighbors for short-term traffic flow prediction based on maximal information coefficient," *Inf. Sci.*, vol. 608, pp. 517–531, Aug. 2022.

[21] J. L, Kong, X. M. Fan, X. B. Jin, and M. Zuo, "Traffic flow prediction via variational Bayesian inference-based encoder–decoder framework," 2022, *arXiv:2212.07194*.

[22] X. Ma, Z. Dai, Z. He, J. Ma, Y. Wang, and Y. Wang, "Learning traffic as images: A deep convolutional neural network for large-scale transportation network speed prediction," *Sensors*, vol. 17, no. 4, p. 818, Apr. 2017.

[23] J, B. Zhang, Y. Zheng, and D. K. Qi, "Deep spatio-temporal residual networks for citywide crowd flows prediction," in *Proc. AAAI*, vol. 31, no. 1. San Francisco, CA, USA, Feb. 2017, pp. 1655–1661.

[24] B. Yu, H. T. Yin, and Z. X. Zhu, "Spatio-temporal graph convolutional networks: A deep learning framework for traffic forecasting," 2017, *arXiv:1709.04875*.

[25] Y. G. Li, R. Yu, C. Shahabi, and Y. Liu, "Diffusion convolutional recurrent neural network: Data-driven traffic forecasting," 2017, *arXiv:1707.01926*.

[26] Y. Chen, K. Li, C. K. Yeo, and K. Li, "Traffic forecasting with graph spatial–temporal position recurrent network," *Neural Netw.*, vol. 162, pp. 340–349, May 2023.

[27] W. Weng, J. Fan, H. Wu, Y. Hu, H. Tian, F. Zhu, and J. Wu, "A decomposition dynamic graph convolutional recurrent network for traffic forecasting," *Pattern Recognit.*, vol. 142, Oct. 2023, Art. no. 109670.

[28] M. Xu and H. Liu, "A flexible deep learning-aware framework for travel time prediction considering traffic event," *Eng. Appl. Artif. Intell.*, vol. 106, Nov. 2021, Art. no. 104491.

[29] M. X. Xu, W. R. Dai, C. M. Liu, and X. Gao, "Spatial–temporal transformer networks for traffic flow forecasting," 2020, *arXiv:2001.02908*.

[30] H. Yan, X. Ma, and Z. Pu, "Learning dynamic and hierarchical traffic spatiotemporal features with transformer," *IEEE Trans. Intell. Transp. Syst.*, vol. 23, no. 11, pp. 22386–22399, Nov. 2022.

[31] X. Ye, S. Fang, F. Sun, C. Zhang, and S. Xiang, "Meta graph transformer: A novel framework for Spatial–Temporal traffic prediction," *Neurocomputing*, vol. 491, pp. 544–563, Jun. 2022.

[32] J. Liu, Y. Kang, H. Li, H. Wang, and X. Yang, "STGHTN: Spatial–temporal gated hybrid transformer network for traffic flow forecasting," *Int. J. Speech Technol.*, vol. 53, no. 10, pp. 12472–12488, May 2023.

[33] T. N. Kipf and M. Welling, "Semi-supervised classification with graph convolutional networks," 2016, *arXiv:1609.02907*.

[34] M. Defferrard, X. Bresson, and P. Vandergheynst, "Convolutional neural networks on graphs with fast localized spectral filtering," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 29, 2016, pp. 1–9.

[35] J. Chung, C. Gulcehre, K. Cho, and Y. Bengio, "Empirical evaluation of gated recur-rent neural networks on sequence modeling," 2014, *arXiv:1412.3555*.

[36] K. K. Bhaumik, F. F. Niloy, S. Mahmud, and S. Woo, "STLGRU: Spatio-temporal lightweight graph GRU for traffic flow prediction," 2022, *arXiv:2212.04548*.

[37] C. Song, Y. Lin, S. Guo, and H. Wan, "Spatial–temporal synchronous graph convolutional networks: A new framework for spatial–temporal network data forecasting," in *Proc. 34th AAAI Conf. Artif. Intell.*, vol. 34, no. 1. New York, NY, USA: AAAI Press, Apr. 2020, pp. 914–921.

[38] Z. Fang, Q. Long, G. Song, and K. Xie, "Spatial–temporal graph ODE networks for traffic flow forecasting," in *Proc. 27th ACM SIGKDD Conf. Knowl. Discovery Data Mining*, Aug. 2021, pp. 364–373.

[39] M. Li and Z. Zhu, "Spatial–temporal fusion graph neural networks for traffic flow forecasting," in *Proc. AAAI Conf. Artif. Intell.*, vol. 35, no. 5, May 2021, pp. 4189–4196.

[40] R. Zhang, F. Xie, R. Sun, L. Huang, X. Liu, and J. Shi, "Spatial–temporal dynamic semantic graph neural network," *Neural Comput. Appl.*, vol. 34, no. 19, pp. 16655–16668, Oct. 2022.

[41] Z. Tan, Y. Zhu, and B. Liu, "Learning spatial–temporal feature with graph product," *Signal Process*, vol. 210, Sep. 2023, Art. no. 109062.

CHUANG CAI received the bachelor's degree from the Shandong University of Technology, in 2021, where he is currently pursuing the master's degree. His research interests include cloud computing, big data analysis, and traffic flow prediction.

ZHIJIAN QU received the Ph.D. degree in engineering from the Beijing University of Posts and Telecommunications, in 2011. He is currently an Associate Professor. His research interests include data analysis algorithms and application technologies, intelligent computing, and evolutionary algorithms and their applications.

LIQUN MA was born in Zibo, Shandong. She is currently pursuing the master's degree with the Shandong University of Technology. Her research interests include artificial intelligence and intelligent data processing.

LIANFEI YU is currently pursuing the master's degree with the Shandong University of Technology, with a major research focus on cloud computing, big data analysis, and traffic flow prediction.

WENBO LIU received the bachelor's degree from the Shandong University of Technology, in 2020, where he is currently pursuing the master's degree. His research interests include cloud computing, big data analysis, and network traffic prediction.

CHONGGUANG REN received the Ph.D. degree in computer science and technology from the Nanjing University of Science and Technology, in 2013. He currently serves as the Director of the Data Intelligence Research Center at Shandong University of Technology and the Leader of the Outstanding Young Innovation Team at Shandong Higher Education Institutions. His research interests include artificial intelligence, intelligent robots, and intelligent data processing.