IEEE Access
Multidisciplinary : Rapid Review : Open Access Journal

Received 24 June 2023, accepted 24 July 2023, date of publication 31 July 2023, date of current version 14 August 2023.

Digital Object Identifier 10.1109/ACCESS.2023.3300232

RESEARCH ARTICLE

Spatio-Temporal Residual Graph Convolutional Network for Short-Term Traffic Flow Prediction

QINGYONG ZHANG, MEIFANG TAN, CHANGWU LI, HUIWEN XIA, WANFENG CHANG, AND MINGLONG LI

School of Automation, Wuhan University of Technology, Wuhan 430070, China

Corresponding author: Huiwen Xia (wendyxia@whut.edu.cn)

This work was supported by the Natural Science Foundation of Hubei Province of China under Grant 2019CFB571.

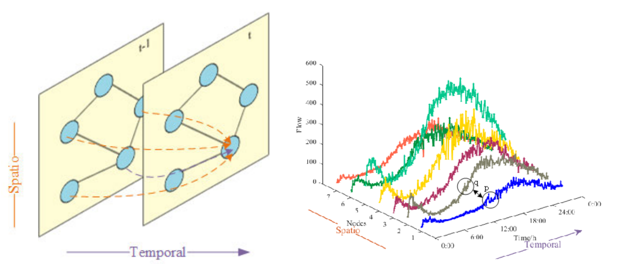
ABSTRACT Accurate spatio-temporal traffic flow prediction is a significant research direction in the intelligent transport system. Current prediction methods have limitations in spatio-temporal feature extraction, and the prediction results have poor performance. In this paper, a short-term traffic flow prediction model based on a Spatio-Temporal Residual Graph Convolutional Network (STRGCN) is proposed to solve the problem of poor accuracy in extracting the spatial and temporal correlation in the short-term traffic flow prediction task. Firstly, a Deep Full Residual Graph Convolutional Network (DFRGCN) module is used to learn the spatial correlation. Secondly, a Bidirectional Gated Recurrent Unit based on the Attention mechanism (ABi-GRU) is used to accurately obtain the temporal dependence of traffic flow data. Finally, the experimental results show that the STRGCN model achieves better prediction performance and stability on three publicly available datasets compared to the baseline methods.

INDEX TERMS Attention mechanism, bidirectional gated recurrent unit, graph convolutional networks, short-term traffic flow prediction, spatio-temporal traffic flow prediction.

I. INTRODUCTION

As people's living standards improve and the pace of urban modernization accelerates, the burden on traffic systems becomes increasingly severe. Intelligent Transport System (ITS) was proposed to alleviate the tremendous pressure on traffic system management by using traffic flow information to formulate appropriate traffic control strategies. The spatio-temporal traffic flow prediction provides ITS with more accurate future traffic flows so that it can make real-time or long-term traffic management strategies, and it can also be used for navigation apps on mobile phones, in-car navigation, Advanced Driver Assistance Systems (ADAS) and so on.

The rational modelling of complex spatio-temporal relationships in urban traffic data has been a challenging problem in the task of predicting urban traffic flow. As shown in Fig.1, spatio-temporal relationship in traffic network, we can see that the traffic flow is affected not only by its historical traffic flow, but also by the traffic flow of its neighbors from Fig.1(a), traffic network topology diagram. Fig.1(b) is the visualisation of traffic flow at the same moment at 7 different nodes. It can be seen that although the 7 nodes have different traffic flows, the overall trend is similar. At the same time, from the two marker points p and q in Fig.1(b), it can be seen that when there is a sudden change in traffic flow at point p, q will also be affected by the sudden change. And this paper specifically describes these two challenges from spatial and temporal aspects as follows.



(a) (b)

FIGURE 1. Spatio-temporal relationship in traffic network. (a) Traffic network topology diagram. (b) Visualisation of traffic flow at the same moment at 7 different nodes.

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

(1) Spatial aspect. The intricate spatial topological structure of urban roads leads to the traffic flows' spatial interdependence. In the traffic network, the traffic state at an upstream road node affects the traffic flow at a downstream road node which also gives feedback to the traffic flow at the upstream road node. The dynamic performance of traffic flow on a road node is influenced by that node's spatial topological structure, which means the change at the current road node is influenced by the traffic state of its neighboring nodes [1]. To address the spatial features in the traffic flows, traditional methods grid the urban space and extract the spatial dependencies using Convolutional Neural Networks (CNNs) [2]. However, as the number of neighbors of each node in the traffic network is not the same, the traffic network does not belong to the Euclidean structure, there are limitations in the way CNNs perform feature extraction.

(2) Temporal aspect. In urban traffic, traffic flow is stochastic and non-linear since the traffic state can be disturbed by vehicles, pedestrians and other factors [3], which brings challenges to the prediction model. When performing short-term traffic flow prediction tasks, the Recurrent Neural Network (RNN) model with memory capacity and its variants, Long Short-Term Memory (LSTM) model and Gate Recurrent Unit (GRU) model [4] can all play the role of temporal feature learning. And these models are relatively simple, easy to train and respond quickly, but the prediction performance still needs to be improved.

Many researchers have made a great deal of contributions to solving the traffic flow prediction problem. In the early time, most of the traffic flow prediction methods used statistical analysis [5], such as Historical Average (HA), Exponential Smoothing (ES), Auto Regressive Integrated Moving Average (ARIMA), and Kalman Filtering (KF). However, traffic flow is non-linear, stochastic, and chaotic, while statistical analysis methods are suitable for solving linear problems and insensitive to traffic flow data trends. With the development of machine learning and neural networks, many machine learning methods suitable for solving non-linear data prediction have been proposed, such as Support Vector Regression (SVR); and neural network methods [6], such as Back Propagation Neural Network (BPNN), RNN, LSTM, CNN, and Graph Convolutional Network (GCN). For the improvement of the prediction accuracy of individual neural networks, plenty of combinatorial models of a swarm intelligence optimization algorithm to optimize the hyperparameters of a neural network have been created, including an improved particle swarm optimization algorithm to optimize BP [7], an improved whale optimization algorithm to optimize wavelet neural network [8]. In addition, there are some combinatorial neural network models such as CNN-BiLSTM [9], and T-GCN [10] with spatio-temporal feature extraction.

In this paper, a short-term traffic flow prediction model, Spatio-Temporal Residual Graph Convolution Network (STRGCN) is proposed. The STRGCN first uses a Deep Full Residual Graph Convolutional Network (DFRGCN) module to learn the spatial feature, and then uses a Bidirectional Gated Recurrent Unit based on the Attention mechanism (ABi-GRU) to accurately obtain the temporal dependence of traffic flow data. The main contributions of this paper are as follows.

(1) Data preprocessing. This paper processes traffic features which include traffic flow and the topological structure of the traffic network to obtain the normalized traffic flow and the traffic feature matrix.

(2) Traffic flow spatial feature learning. The processed feature matrix is input into the STRGCN model and we use the Deep Full Residual Graph Convolutional Network (DFRGCN) module to capture the spatial correlation between the nodes in the traffic network.

(3) Traffic flow temporal dependence capturing. We use the Bidirectional Gated Recurrent Unit based on the Attention mechanism (ABi-GRU) to learn the temporal dependency of the traffic flow, thus completing the task of accurately predicting the traffic flow at the next time step.

(4) Model evaluation. The STRGCN model is applied to real traffic flow data and compared with other baseline models. At the same time, the experiments on the effect of hyperparameters and ablation study are designed to verify the superiority of the model.

II. RELATED WORK

In the field of traffic flow prediction, many researchers have contributed different research methods, which can be generally divided into methods based on time series, and methods based on machine learning and neural networks in order of development.

With the help of the CiteSpace visualisation tool, we visualised the current state of research on the keyword "traffic flow prediction", using the Web of Science as the data source, and obtained the keyword burst map shown in Fig.2.

It can be seen from the keyword burst map in Fig.2 that from 2008 to 2014, the "time series" method was mainly used, among which "extended Kalman filter" and "cell transmission model" were the dominant methods. In the time series method, Jeffery et al. [11], [12] first applied the HA method to traffic flow prediction. Okutani and Stephanedes [13] proposed the application of KF to traffic flow prediction. Wu et al. [14] combined MA and KF to traffic flow prediction. Lin [15] adopted K-nearest neighbor nonparametric regression prediction to the taxi passenger flow prediction of the capital airport. Yuan and Zhou [16] introduced new parameters to make up for the loss of information caused by difference in traditional models, and constructed a short-term traffic flow prediction model based on ARIMAX. However, with the development of big data and neural networks, prediction methods based on time series often perform poorly when dealing with large amounts of data. The development of machine learning and neural networks has led traffic flow prediction to a new level [6], which is favoured by many researchers.

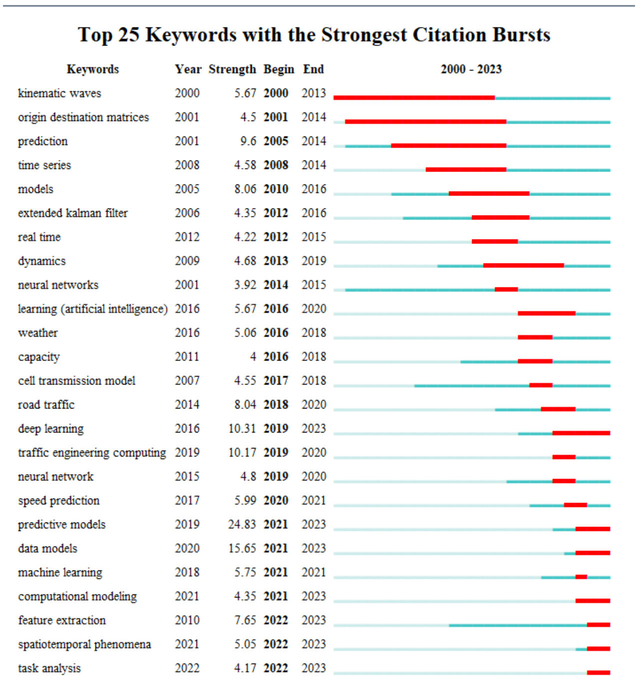


FIGURE 2. Keyword burst map of traffic flow prediction state.

From Fig.2, We can also see that "deep learning", "neural networks", and "machine learning" have gradually become the dominant methods in recent years. For temporal correlation extraction, Chao et al. [17] used BP network to predict vehicle passage times, which was the first application of BP network in traffic flow prediction, and since then, BP network has been widely used in the prediction of various types of traffic flows [18], [19], [20], [21]. Han and Huang [22] used a deep belief network and a kernel extreme learning machine classifier as a prediction model. Ma et al. [23] first applied a LSTM neural network to traffic speed prediction; Chen et al. [24] used a data denoising scheme to suppress potential data outliers and then introduced a LSTM neural network to meet the traffic flow prediction task.

For spatial correlation extraction, Pu et al. [25] designed a model based on a new attentional convolutional neural network with an encoder-decoder framework. Li et al. [26] proposed a diffusion convolutional recurrent neural network applied to traffic flow prediction using a bidirectional random flow on the graph to capture spatial dependencies. Since the traffic network can be seen as a graph structure, traffic flow prediction methods based on GCN were explored, and Yu et al. [27] used a spatio-temporal graph convolutional network to solve the prediction problem. Subsequently, many researchers proposed traffic flow prediction methods such as T-GCN [10] and ASTGCN [1], which are derived methods for GCN. Besides, methods such as attention mechanism [28], [29] and transformer [30] are also commonly used in the field of traffic flow prediction. And taking day of week, weather, and holiday as an entry point for traffic flow prediction is also a way worth exploring [31], [32].

Meanwhile, we can also see from Fig.2 that in recent years, the current methods used for traffic flow prediction are still dominated by "deep learning" and "neural networks". And "feature extraction", "spatiotemporal phenomena", and "task analysis" are still the key objects of current traffic flow prediction. In this paper, we propose a model based on a spatio-temporal residual graph convolutional network for short-term traffic flow prediction.

III. PROBLEM DESCRIPTION

A. PROBLEM STATEMENT

In traffic flow theory, the three main parameters that reflect traffic flow are traffic flow, speed, and road occupancy [33], where traffic flow usually indicates the flow of vehicles running on the road [34].

B. PROBLEM DEFINITION

Traffic network: The traffic network can be defined as an undirected graph $G = \{V, E, \mathbf{A}\}$ [35], where $V(G)$ means the set of vertices of G, which is the node set in the traffic network; $E(G)$ means the set of edges of G, which is the road set in the traffic network; and $\mathbf{A} \in \mathbb{R}^{N\times N}$ represents the set of arcs of G, which is the adjacency matrix in the traffic network, indicating the connection between any two nodes, and N represents the number of nodes.

Mapping function: The task of spatio-temporal short-term traffic flow prediction is to use the historical traffic flow data of length T_R to predict the traffic flow data of the next period. It means that the traffic flow $Y_{t-T_R+1}, Y_{t-T_R+2}, \ldots, Y_t$ is used to predict the traffic flow Y_{t+1} at the next period, and the mapping function can be expressed as shown in

$$Y_{t+1} = f\left(\left\{Y_{t-T_R+1}, Y_{t-T_R+2}, \ldots, Y_t\right\}; G\right). \tag{1}$$

IV. METHODOLOGY

In the short-term traffic flow prediction tasks, temporal prediction models such as RNN, LSTM, and GRU are usually conducive to temporal dependence learning, however, the global temporal features are weakly captured. The spatial feature in the traffic network, as a Granger cause of spatio-temporal traffic flow prediction [36], [37], is equally important for prediction accuracy. For the problem of capturing the spatial feature of traffic flow, GCN has received more and more attention from researchers since the traffic flow data does not belong to Euclidean data. Therefore, we propose a short-term traffic flow prediction model based on a Spatio-Temporal Residual Graph Convolution Network (STRGCN) to accurately predict the traffic flow at the next moment in the future. the overall model framework of the STRGCN model is described in Fig.3.

As we can see from Fig.3, the traffic data in the traffic network is generally divided into traffic flow, speed, and road occupancy, the object of this paper is the traffic flow among them, which is collected by sensors distributed on the road. Firstly, we need to preprocess the traffic flow data, i.e.,

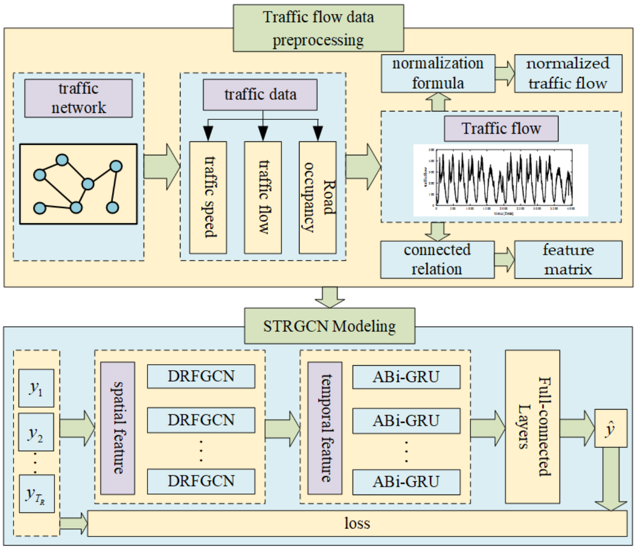


FIGURE 3. STRGCN model framework.

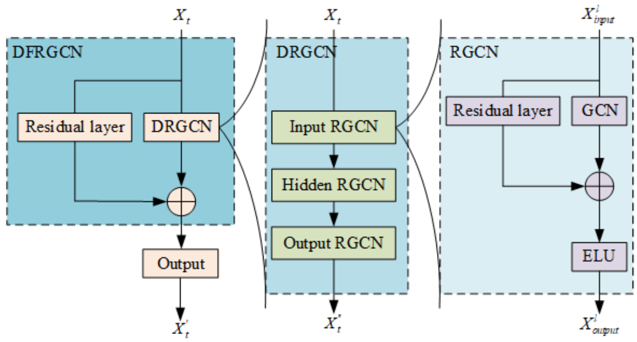


FIGURE 4. DFRGCN model framework.

we normalize the traffic flow data to generalize the statistical distributivity of the sample and solved to obtain the feature matrix between the nodes. After that, the STRGCN traffic flow prediction model is constructed, in which the Deep Full Residual Graph Convolutional Network (DRFGCN) module is used to obtain the spatial dependence, and the Bidirectional Gated Recurrent Unit based on the Attention mechanism (ABi-GRU) is used to learn the historical information between traffic flow, and the traffic flow prediction value is obtained through the fully connected layer to calculate the loss function value.

A. SPATIAL CORRELATION MODELLING

It is a core problem that effective extraction of the spatial feature of traffic network in the spatio-temporal traffic flow prediction. Generally, as the layers of the network (GCN) deepen, the noise information may become more obvious. In order to avoid this situation, the DFRGCN introduces a full residual structure that optimizes the serial GCN into a DFRGCN capable of parallel computation to improve the prediction accuracy of the model while enhancing the model training efficiency. The DFRGCN's framework module is shown in Fig.4.

The DFRGCN module is composed of a Deep Residual Graph Convolution Network (DRGCN) module and a residual layer. And the DRGCN module captures the spatial correlation through three Residual Graph Convolution Network (RGCN) modules. The RGCN module is the core submodule of the DFRGCN, which reduces the error spread and accumulation problems by adding a residual network. After the residual layer, the traffic flow features are activated by the ELU activation function to increase the nonlinear fitting capability of the model, and the final output of the RGCN module is obtained.

1) RESIDUAL GRAPH CONVOLUTION NETWORK MODULE

Graph Convolutional Network (GCN) is a neural network based on graph structure that can effectively handle data in non-Euclidean spaces like traffic flow [38]. The core idea of GCN is to use the Laplacian matrix of a graph to define convolutional operations on the graph, which enables aggregation and propagation of node features [39].

Each graph structure can be represented by an adjacency matrix $\mathbf{A}$, and each hidden layer in the network can be represented as a nonlinear function as shown in

$$H^{(l+1)} = f\left(H^{(l)}, \mathbf{A}\right) \tag{2}$$

where $\mathbf{A}$ is the adjacency matrix in the traffic network; when $l = 0$, $H^{(l)} = H^{(0)}$ denotes the input layer; when $l = L-1$, $H^{(l+1)} = H^{(L)}$ denotes the output layer, and L is the number of layers. $f\left(H^{(l)}, \mathbf{A}\right)$ is shown in

$$f\left(H^{(l)}, \mathbf{A}\right) = \sigma\left(\hat{\mathbf{D}}^{-\frac{1}{2}}\hat{\mathbf{A}}\hat{\mathbf{D}}^{-\frac{1}{2}}H^{(l)}\mathbf{W}^{(l)}\right) \tag{3}$$

where $\sigma(\bullet)$ is the nonlinear activation function, $\mathbf{W}^{(l)}$ is the weight matrix, $\hat{\mathbf{D}}^{-\frac{1}{2}}\hat{\mathbf{A}}\hat{\mathbf{D}}^{-\frac{1}{2}}$ is a symmetric normalization of $\mathbf{A}$, and $\hat{\mathbf{A}} = \mathbf{A} + \mathbf{I}$, $\hat{\mathbf{D}}$ is the degree matrix of $\hat{\mathbf{A}}$.

The computational process of GCN can be understood as follows. Firstly, the adjacency matrix is normalized so that each node's neighbor information has the same weight. Then the normalized adjacency matrix is multiplied by the node feature matrix of the current layer to obtain the weighted sum of each node and its neighbor features. Finally, the weighted sum is multiplied by the weight matrix and the node feature matrix of the next layer is obtained by the activation function. In this way, each node can fuse the information of itself and its neighbors to learn the feature representation at a higher level.

As the number of GCN layers increases, the noise information will also accumulate. Therefore, this paper introduces a residual structure to solve this problem, thus forming a Residual Graph Convolution Network (RGCN) module, whose output features are shown in

$$\vec{h}_{i,\mathrm{res}} = W_{\mathrm{res}}\vec{h}_i \tag{4}$$

$$\vec{h}_i' = \sigma\left(\vec{h}_{i,\mathrm{res}} + \hat{\mathbf{D}}^{-\frac{1}{2}}\hat{\mathbf{A}}\hat{\mathbf{D}}^{-\frac{1}{2}}H^{(l)}\mathbf{W}^{(l)}\right) \tag{5}$$

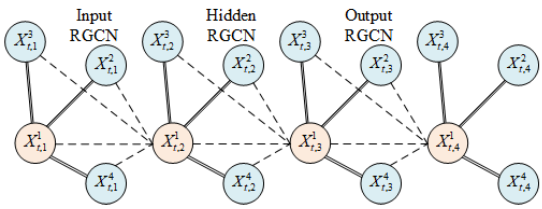


FIGURE 5. The diffusion model of DRGCN.

where $\vec{h}_i \in \mathbb{R}^F$ is the current node input features, F is the number of features in each node; $W_{\text{res}} \in \mathbb{R}^{F \times F}$ is the linear transformation matrix, meaning that the input data undergoes a linear transformation; $\vec{h}_{i,res} \in \mathbb{R}^F$ is the feature matrix combined with the residual layer, and $\vec{h}'_i$ is the output of the RGCN network after activation function.

2) DEEP RESIDUAL GRAPH CONVOLUTION NETWORK MODULE

The Deep Full Residual Graph Convolution Network (DFRGCN) module is composed of the Deep Residual Graph Convolution Network (DRGCN) module and a residual layer, which can avoid the loss of local features and the diffusion of errors through the parallel computation of the residual layer. As Fig.5 shows the diffusion model of DRGCN, the module contains three layers of input, hidden, and output RGCN modules.

From Fig.5, we can see that the central node of the DRGCN module is capable of learning third-order neighborhood nodes' information. The central node of the "Input RGCN" in Fig.5 already contains information about first-order neighborhood nodes. And similarly, we can see that the central node of the "Hidden RGCN" contains information about second-order neighborhood nodes, and the central node of the "Output RGCN" contains information about third-order neighborhood nodes. It helps to better learn spatial correlation by learning information about third-order neighborhood nodes. The calculation formula of the multilayer RGCN module in the DRGCN module is shown in

$$\vec{h}_i^{l+1} = \sigma\left(W_{\text{res}}^l \vec{h}_i^l + \hat{\mathbf{D}}^{-\frac{1}{2}} \hat{\mathbf{A}} \hat{\mathbf{D}}^{-\frac{1}{2}} H^{(l)} \mathbf{W}^{(l)}\right) \tag{6}$$

where $\vec{h}_i^l$ is the input feature of the layer l of the DRGCN module.

In this paper, we use $DRGCN(\mathbf{A}, X_t)$ to represent the DRGCN module, where $\mathbf{A}$ is the topological structure of the traffic network; $X_t \in \mathbb{R}^{N \times F}$ is the input features of the traffic flow. Therefore, the formula of the DFRGCN module is shown in

$$DFRGCN(\mathbf{A}, X_t) = \sigma(DRGCN(\mathbf{A}, X_t) + W_{\text{res}} X_t) \tag{7}$$

B. TEMPORAL CORRELATION MODELLING

Although the DFRGCN module has the spatial feature extraction capability, it is slightly less capable of learning temporal correlation. Therefore, this section proposes a Spatio-Temporal Residual Graph Convolutional Network (STRGCN) traffic flow prediction model based on the DFRGCN module to improve the temporal correlation learning capability. The STRGCN short-term traffic flow prediction framework is shown in Fig.6. In temporal correlation learning, it includes the following two core aspects: 1) extracting and capturing temporal correlations in traffic flow data by the bidirectional gated recurrent unit, and 2) dynamically aggregating the hidden layer state of the bidirectional gated recurrent unit by attention mechanism. In this way, we obtain the traffic flow prediction value at the next time step to achieve accurate short-term prediction.

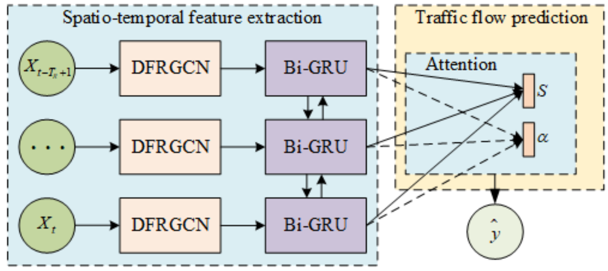


FIGURE 6. STRGCN short-term traffic flow prediction framework.

1) GATED RECURRENT UNIT

The Gated Recurrent Unit (GRU) is a variant of the recurrent neural network proposed in 2014 and it is easy to compute. The mathematical model of GRU is shown in

$$r_t = \sigma\left(W_r\left(x||h_{t-1}\right)\right) \tag{8}$$

$$u_t = \sigma\left(W_u\left(x||h_{t-1}\right)\right) \tag{9}$$

$$h'_t = tanh(W_{h'}(x||\left(r_t{*}h_{t-1}\right))) \tag{10}$$

$$h_t = (1 - u_t){*}h_{t-1} + u_t{*}h'_t \tag{11}$$

where x is the input vector of the GRU, W_r, W_u, and $W_{h'}$ are the weight matrices, r_t and u_t represent the reset gate and the update gate respectively. h_{t-1} is the hidden layer state at time $t-1$, h'_t is the candidate hidden layer state at time t, h_t is the hidden layer state at current time t, and tanh(•) is an activation function [40].

2) BIDIRECTIONAL GATED RECURRENT UNIT

In the GRU network, feature extraction is always performed in a unidirectional temporal order from front to back. It means that the state of the hidden layer at the current t moment contains only the past and present traffic information. In order to enable the model to focus on future time information, we use Bidirectional GRU (Bi-GRU) to model the temporal correlation.

The structure of Bi-GRU is composed of two unidirectional GRU models with opposite directions, as shown in Fig.7. The forward unidirectional GRU model performs feature extraction in temporal order, specifically from time step $t - T_R + 1$ to time step t. In contrast, the reverse unidirectional GRU model performs feature extraction in the opposite order of

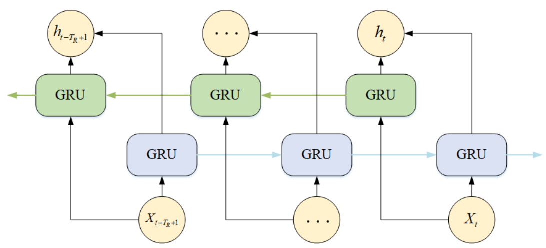


FIGURE 7. The model structure of Bi-GRU.

the temporal order, specifically from time step t to time step $t - T_R + 1$ [41].

The traffic flow of each historical time step will be input to the forward and reverse unidirectional GRU models, respectively. And the hidden layer state obtained from the traffic flow of each time step through the two directions of the unidirectional GRU model will be integrated in a fixed way, and finally obtain the hidden layer state incorporating the past, present, and future traffic flow information. The mathematical model of the Bi-GRU is shown in

$$h_t = W_{BiGRU}\left(\overrightarrow{h}_t \parallel \overleftarrow{h}_t\right) \tag{12}$$

$$\overrightarrow{h}_t = GRU\left(x_t, \overrightarrow{h}_{t-1}\right) \tag{13}$$

$$\overleftarrow{h}_t = GRU\left(x_t, \overleftarrow{h}_{t+1}\right) \tag{14}$$

where W_{BiGRU} is the linear transformation of the state of the hidden layer of Bi-GRU; || denotes the connection operation of the matrices. And we use $\overrightarrow{h}_t$ and $\overleftarrow{h}_t$ to denote the hidden layer state of the unidirectional GRU model in the forward and reverse direction at time step t respectively. Thus, the hidden layer state value of Bi-GRU at the current time step t can be calculated from the three components, the input matrix x_t of the traffic flow at the time step t, the hidden layer state value $\overrightarrow{h}_{t-1}$ of the forward unidirectional GRU model at the time step $t-1$, and the hidden layer state value $\overleftarrow{h}_{t+1}$ of the reverse unidirectional GRU model at the time step $t+1$.

3) BIDIRECTIONAL GATED RECURRENT UNIT BASED ON THE ATTENTION MECHANISM

The bidirectional gated recurrent unit enables each hidden layer state to contain past, present, and future traffic information. In order to accurately predict the traffic flow at time step $t+1$ using the historical traffic flow information from time steps $t - T_R + 1$ to t, all the hidden layer states need to be efficiently integrated to obtain the predicted traffic flow at time step $t+1$.

In the T_R time steps of the hidden layer states, it is not all state values are of the same importance. At a certain moment, only some of the data are important, and the degree of importance varies. In this case, the attention mechanism can play its advantage to help the model focus on some important state values, and thus the model can achieve better training results. Therefore, we use the attention mechanism to integrate the hidden layer states over T_R time steps.

The ABi-GRU is divided into the following three parts: 1) the bidirectional gated recurrent unit, through which the hidden layer states are obtained incorporating past, present, and future traffic flow information; 2) the attention mechanism, through which the hidden layer states are integrated over T_R time steps; 3) the fully connected layer, through which the integrated information is converted to the final traffic flow prediction value. In summary, the computational idea of the ABi-GRU is shown as follows.

Step 1: The hidden layer state h is calculated for T_R time steps using a bidirectional gated recurrent unit

$$h_t = BiGRU\left(x_t, \overrightarrow{h}_{t-1}, \overleftarrow{h}_{t+1}\right) \tag{15}$$

Step 2: The attention mechanism is applied to integrate the hidden layer state of T_R time steps to obtain the integrated hidden layer state s

$$e_i = \mathtt{LeakyReLU}\left(w_i h_i + b_i\right) \tag{16}$$

$$\alpha_i = \text{softmax}_i(e_i) = \frac{\exp(e_i)}{\sum_{i=1}^{T_R} \exp(e_i)} \tag{17}$$

$$s = \sum_{i=1}^{T_R} \alpha_i h_i \tag{18}$$

where h_i is the hidden layer state in T_R time steps, w_i is the weight coefficient, b_i denotes the bias, e_i denotes the attention coefficient of the weight of the hidden layer state at the time step i among all hidden layer states, and α_i denotes the normalized e_i by the softmax function.

Step 3: The fully connected layer is used to obtain the integrated hidden layer state s, and the final accurate prediction value is shown in

$$\hat{y} = w_o s + b_o \tag{19}$$

where w_o denotes the weight coefficient of the fully connected layer, b_o denotes the bias of the fully connected layer, and $\hat{y}$ denotes the predicted value of the model.

V. EXPERIMENTAL RESULTS

A. DATASETS

To evaluate the predictive effectiveness of the STRGCN model proposed in this paper, the performance of the STRGCN is evaluated using real freeway datasets: PeMS04, PeMS07, and PeMS08 [42]. These datasets are freeway datasets collected by the California Department of Transportation Performance Measurement System, and the time interval is 5 minutes. In this paper, traffic flow is used as the object of study. All datasets are divided into training, validation, and test sets by the ratio of 6:2:2. The detailed division of these datasets is shown in Table 1.

In order to generalize the statistical distributivity of the sample, we normalize the traffic flow by the "max-min"

TABLE 1. Dataset description.

Dataset	Time range	Loop detector	Training set	Validation set	Test set
PeMS04	2018.1.1-2018.2.28	307	10188	3396	3396
PeMS07	2017.5.1-2017.8.6	883	16927	5643	5643
PeMS08	2016.7.1-2016.8.31	170	10706	3569	3569

normalization. And the calculation formula is shown in

$$y_i' = \frac{y_i - \min(y_i)}{\max(y_i) - \min(y_i)} \tag{20}$$

where y_i denotes the ith traffic flow data, y_i' denotes the ith traffic flow data after normalization; max(•) and min(•) denote the maximum and minimum traffic flow data values respectively.

B. EVALUATION METRICS

The prediction effectiveness of a model can be accurately measured by calculating the error between the true and predicted values of traffic flow. In the field of traffic flow prediction, most papers prefer to use Mean Absolute Error (MAE), Root Mean Square Error (RMSE) and Mean Absolute Percentage Error (MAPE) as the evaluation metrics. Because they can evaluate the model by scaling up the error, and better reflect the accuracy and stability of the model. In this paper, we use MAE, RMSE, and MAPE to assess the performance of prediction models [43]. The formulas of these three evaluation metrics are shown in

$$\mathrm{MAE} = (\sum_{i=1}^{N} |y_i - \hat{y}_i|)/N \tag{21}$$

$$\mathrm{RMSE} = \sqrt{\sum_{i=1}^{N} (y_i - \hat{y}_i)^2/N} \tag{22}$$

$$\mathrm{MAPE} = \frac{100\%}{N} \sum_{i=1}^{N} \left| \frac{y_i - \hat{y}_i}{y_i} \right|. \tag{23}$$

where y_i is the true traffic flow value, $\hat{y}_i$ is the predicted value, and N is the sample size. A lower value of the evaluation metrics means that the model has higher accuracy and better prediction capability.

C. PARAMETER DISCUSSION

We use the Huber loss function in this paper [44]. And the parameters for the experiment are set as follows, 1) optimizer is Adam, 2) batch size is 24, 3) learning rate is 0.002, 4) maximum iterations is 200, 5) the number of layers of the DFRGCN module is 3, 6) the number of Bi-GRU hidden neuron is 64, and 7) the historical time length is 12.

D. COMPARED METHODS

To verify the prediction performance of the STRGCN short-term traffic flow prediction model, we compare the STRGCN model with the following baseline prediction models, which include 1) Historical Averaging (HA) model, 2) Long Short-Term Memory (LSTM) network, 3) Gated Recurrent Unit (GRU) network, 4) Graph Convolutional Network (GCN), 5) Graph WaveNet for deep spatial-temporal graph modeling (Graph WaveNet) [45], 6) Spatio-Temporal Graph Convolutional Networks: a deep learning framework for traffic forecasting (STGCN) [27], and 7) Attention based Spatial-Temporal Graph Convolutional Networks for traffic flow forecasting (ASTGCN) [1].

E. EXPERIMENTAL RESULTS

1) EXPERIMENTAL RESULTS AND ANALYSIS

The prediction results of the STRGCN model at node 150 of PeMS04, node 100 of PeMS07, and node 150 of PeMS08 are visualized in Fig.8 and Fig.9. Fig.8 shows the traffic flow visualization results for a week, and Fig.9 shows the traffic flow visualization results for a day. From Fig.8 and Fig.9, it can be seen that the traffic flow prediction values of the STRGCN model can follow the true values well, and it can effectively capture the spatio-temporal features and dynamic changes of traffic flow. From the local amplification of Fig.8 and Fig.9, it can also be seen that the STRGCN model is effective in predicting traffic flow during sudden changes, peaks, and troughs periods in traffic flow, and the predicted values can follow these situations to a certain extent.

Table 2 is the average performance of different methods on the PeMS04, PeMS07, and PeMS08 datasets. As we can see that the STRGCN model has the lowest prediction error among the compared methods for all three datasets, indicating that the STRGCN model has better prediction performance.

Specifically from Table 2, the first three prediction methods (HA, LSTM, and GRU) are prediction methods for temporal correlation only, and GCN is a prediction method for spatial correlation only.

In terms of temporal correlation, HA is an early traffic flow prediction method based on statistical analysis, and LSTM and GRU are recent traffic flow prediction methods based on neural networks. It can be seen that the prediction performance of the neural networks prediction methods is better than that of the statistical analysis prediction method on all three datasets. On the PeMS04 dataset, the prediction error of GRU is slightly smaller than that of LSTM, and compared with HA, the prediction error of GRU is reduced by 30.66%, 24.85%, and 32.24% at MAE, RMSE, and MAPE, respectively. On the PeMS07 dataset, the prediction error of GRU is slightly smaller than that of LSTM, and compared with HA, the prediction error of GRU is reduced by 39.03%, 34.31%, and 43.05% at MAE, RMSE, and MAPE, respectively. And on the PeMS08 dataset, the prediction error of LSTM is

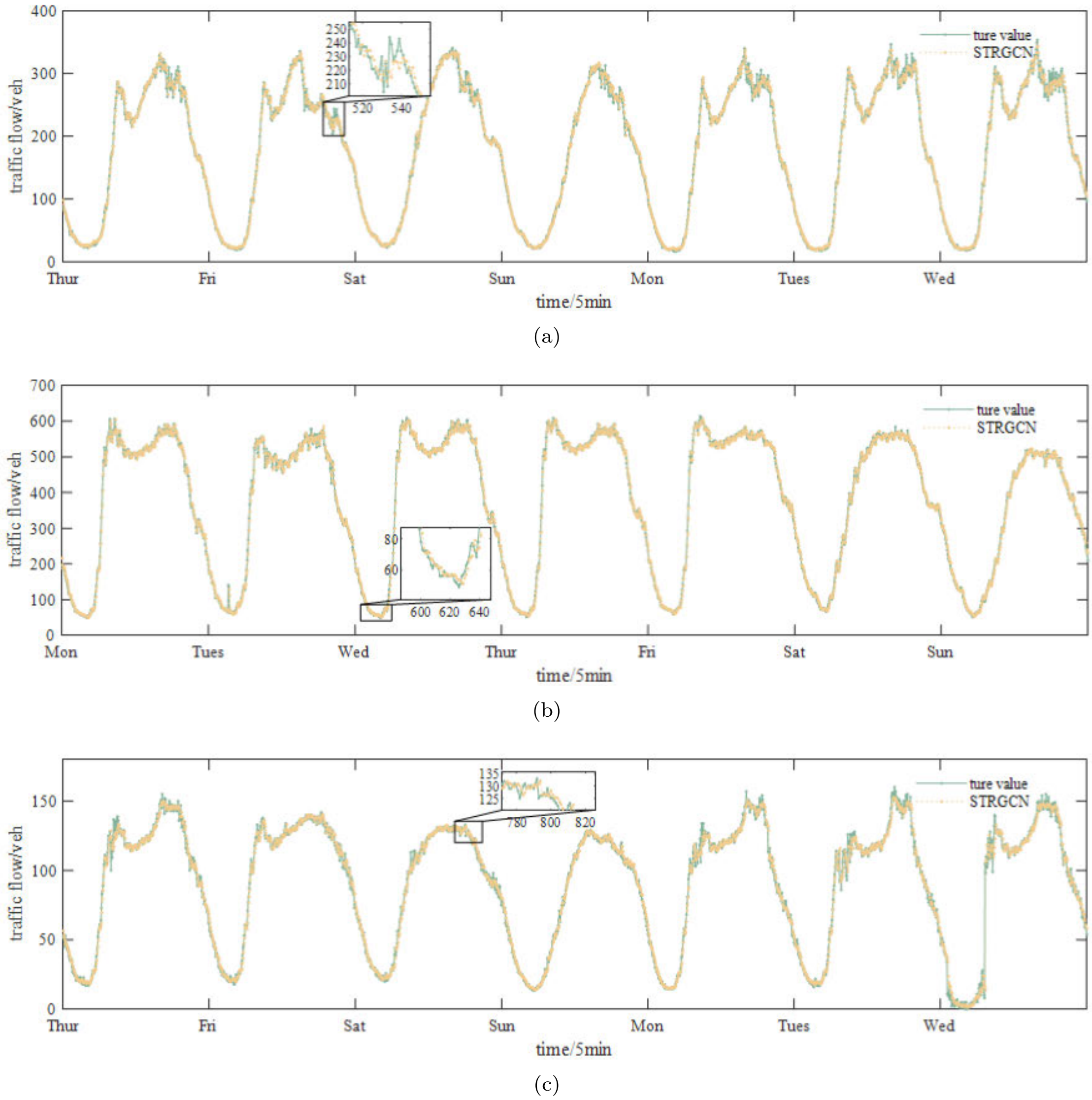


FIGURE 8. Visualisation of a week's traffic flow prediction. (a) Node 150 of PeMS04. (b) Node 100 of PeMS07. (c) Node 150 of PeMS08.

TABLE 2. The average performance of different methods on the PeMS04, PeMS07, and PeMS08 datasets.

Dataset	Evaluation metrics	HA	LSTM	GRU	GCN	Graph WaveNet	STGCN	ASTGCN	STRGCN
PeMS04	MAE	25.89	17.98	17.95	31.80	17.62	19.21	18.32	**17.27**
	RMSE	38.27	28.77	28.76	47.28	28.28	30.81	30.65	**27.92**
	MAPE(%)	18.02	12.23	12.21	24.17	11.82	12.25	12.09	**11.54**
PeMS07	MAE	29.80	18.20	18.17	34.92	18.12	20.81	18.58	**17.67**
	RMSE	43.78	28.77	28.76	52.81	28.51	33.96	32.67	**28.29**
	MAPE(%)	13.59	7.76	7.74	16.09	7.86	8.56	8.12	**7.48**
PeMS08	MAE	21.34	13.93	13.89	34.61	14.37	16.58	14.23	**13.21**
	RMSE	31.96	21.43	21.45	50.41	21.85	25.64	21.43	**20.64**
	MAPE(%)	13.29	8.89	8.98	27.89	8.56	9.83	9.42	**8.37**

slightly smaller than that of GRU, and compared with HA, the prediction error of LSTM is reduced by 34.72%, 32.95%, and 33.11% at MAE, RMSE, and MAPE, respectively. From the above comparison, it can be seen that the neural networks based on the time series prediction method are able to learn the temporal relationship between traffic flow data better than the statistical analysis methods.

Since the prediction errors of LSTM and GRU in Table 2 are only slightly different on the three datasets, this paper uses these two models on the PeMS08 for prediction comparison

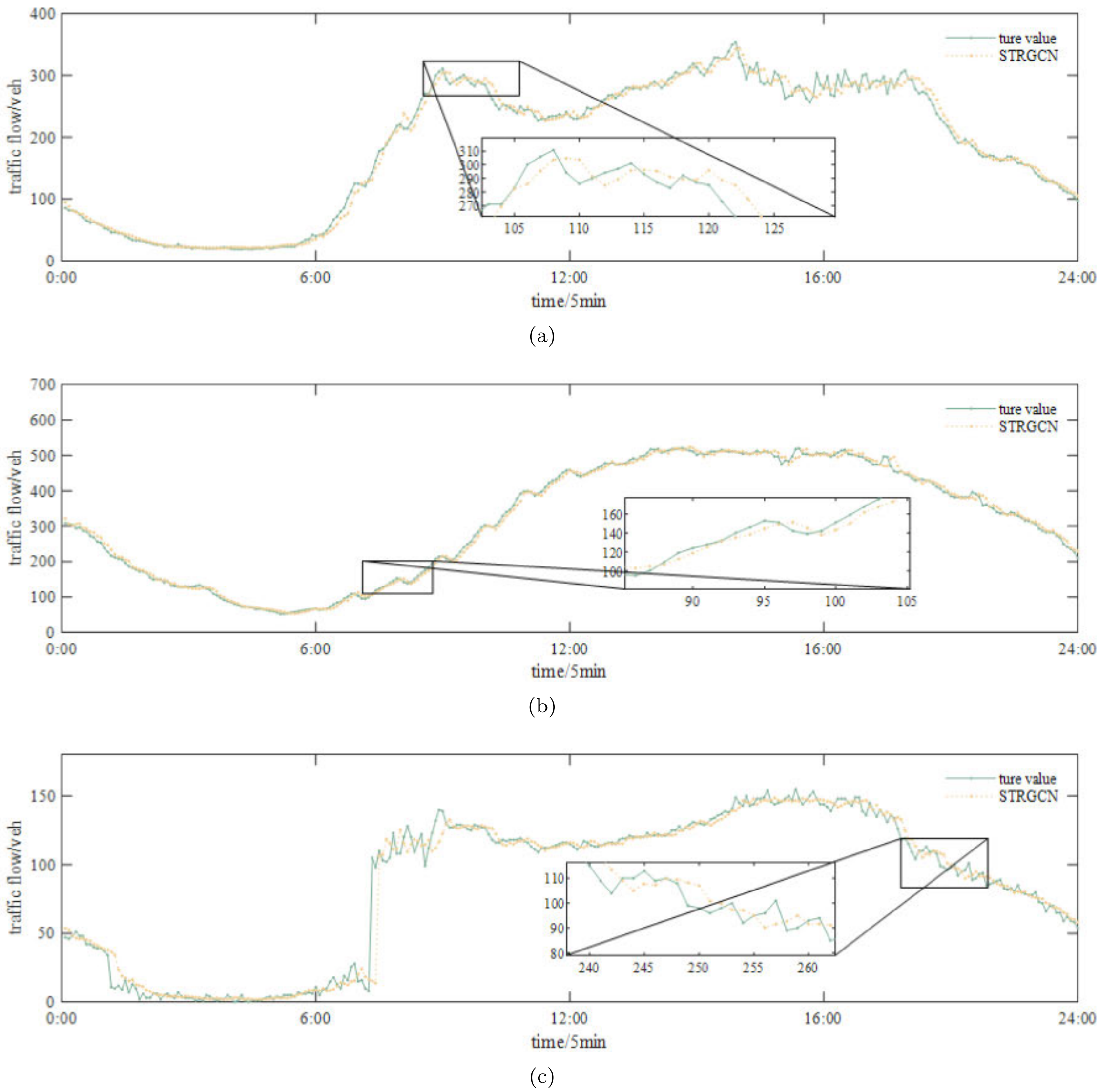


FIGURE 9. Visualisation of a day's traffic flow prediction. (a) Node 150 of PeMS04. (b) Node 100 of PeMS07. (c) Node 150 of PeMS08.

and obtains the duration and the number of parameters when training one generation, which is shown in Table 3.

It can be seen from Table 3 that the number of parameters of the LSTM model is larger than that of the GRU model, and the duration of training one generation is also larger than that of GRU. Therefore, although both LSTM and GRU are good at the task of extracting temporal correlations, LSTM has more training parameters. With no significant difference in the overall training results, more training parameters would impose a greater training burden on the overall model. It further proves that it is more practical and effective to use GRU model for temporal correlation extraction in this paper.

Combined with Table 1, it can be analyzed that the STRGCN model is 3.79%, 2.92%, and 5.49% lower than those of the GRU model on the PeMS04 dataset at MAE, RMSE, and MAPE, respectively. On the PeMS07 dataset, the STRGCN model is 2.75%, 1.63%, and 3.36% lower than those of the GRU model at MAE, RMSE, and MAPE, respectively. And on the PeMS08 dataset, the STRGCN model is 4.90%, 3.78%, and 6.79% lower than those of the GRU model at MAE, RMSE, and MAPE, respectively.

In terms of spatial correlation, the STRGCN model considers temporal correlation compared to the GCN model, which significantly improves the prediction results. On the PeMS04 dataset, the STRGCN model is reduced by 45.69%, 40.95%, and 52.25% at MAE, RMSE, and MAPE respectively, compared with the GCN model that only considers spatial correlation. On the PeMS07 dataset, the STRGCN model is reduced by 49.40%, 46.40%, and 53.51% at MAE, RMSE, and MAPE respectively. And on the PeMS08 dataset, the STRGCN model is reduced by 61.83%, 59.06%, and 69.99% at MAE, RMSE, and MAPE respectively.

TABLE 3. The prediction results' comparison of LSTM and GRU models on the PeMS08 dataset.

Model	MAE	RMSE	MAPE	Duration/epoch(min)	Number of parameters
LSTM	13.93	21.43	8.89	0.11	16961
GRU	13.89	21.45	8.98	0.10	12737

And we can see that, compared with the three classical models of Graph WaveNet, STGCN, and ASTGCN, the short-term traffic flow prediction effect of STRGCN is also better than these three models, which proves the superiority of the STRGCN model. In short, combining the results of Fig.8, Fig.9, and Table 2, it can be concluded that the STRGCN model with the integration of temporal correlation and spatial correlation can better perform the task of short-term spatio-temporal traffic prediction.

2) EFFECT OF HYPERPARAMETERS

In order to measure the effect of different parameters on the prediction effect of the model, this paper designs a comparison experiment for two hyperparameters, the number of DFRGCN layers (layer) and the number of Bi-GRU hidden neurons (hid_size). And the layer in the STRGCN model is set to 1, 2, 3 and 4, and the hid_size is set to 8, 16, 32, 64, and 128, respectively.

Fig.10 shows the effect of the number of DFRGCN layers on the STRGCN model, where (a), (b), and (c) are the experiments conducted on the PeMS04, PeMS07, and PeMS08 datasets, respectively. From Fig.10 (a), we can see that on the PeMS04 dataset, the minimum values of MAE and RMSE are obtained at layer = 3, and the minimum value of MAPE is obtained at layer = 4. From Fig.10 (b), we can see that on the PeMS07 dataset, the minimum values of MAE, RMSE, and MAPE are all obtained at layer = 3; at the same time, from Fig.10 (b), we can see that the three evaluation metrics have the greatest rate of change from layer = 2 to layer = 3. As can be seen in Fig.10 (c), on the PeMS08 dataset, the minimum values of MAE, RMSE, and MAPE are obtained at layer = 4, followed by the corresponding MAE, RMSE, and MAPE at layer = 3. In summary, the prediction errors of the model at layer = 3 and layer = 4 are smaller in all three datasets, while the training time of the model at layer = 3 is 0.28 min, 1.31 min, and 0.26 min for one generation on the three datasets, respectively, and the training time of the model at layer = 4 is 0.3 min, 1.44 min, and 0.26 min for one generation on the three datasets, respectively. As the number of the layers increases, the training parameters of the model also increase, which also puts a greater burden on the training of the model. Therefore, compared with layer = 4, the model with layer = 3 can obtain better results with less cost, which has a better application value.

Fig.11 shows the effect of the number of Bi-GRU hidden layer neurons (hid_size) on the STRGCN model. From Fig.11, we can see that the model only obtains the minimum MAPE on the PeMS08 dataset when hid_size = 128, and the rest of the error values are all obtained when hid_size = 64. Meanwhile, in terms of model time complexity analysis, the rate of change of the duration when training one generation at hid_size = 64 compared to hid_size = 32 is smaller than the rate of change of the duration when training one generation at hid_size = 128 compared to hid_size = 64 on all three datasets. For example, on the PeMS04 dataset, the training one generation time for hid_size = 64 is 0.28 min compared to 0.27 min for hid_size = 32, which is an increase of 3.7%, while the training one generation time for hid_size = 128 is 0.48 min compared to 0.28 min for hid_size = 64, which is an increase of 71.43%.

3) ABLATION STUDY

The STRGCN model contains two main modules for temporal and spatial correlation extraction, and to verify the role of each module, we design ablation study on the PeMS04, PeMS07, and PeMS08 datasets. The details of the ablation models are described as follows.

1) STRGCN-T model: Based on the STRGCN model, the ABi-GRU for learning temporal correlation is removed, and only the DFRGCN module incorporating the full residual structure is retained.

2) STRGCN-S model: Based on the STRGCN model, the DFRGCN module capturing spatial correlation is removed, and only the ABi-GRU is retained.

Table 4 shows the experimental results of the STRGCN models as well as the ablation models on the three datasets. From Table 4, it can be seen that the STRGCN model is reduced by 33.27%, 28.21%, and 34.95% at MAE, RMSE, and MAPE, respectively, on the PeMS04 dataset compared to the STRGCN-T model. On the PeMS07 dataset, the STRGCN model is reduced by 40.53%, 35.41%, and 43.97% at MAE, RMSE, and MAPE, respectively. And on the PeMS08 dataset, the STRGCN model is reduced by 37.45%, 34.81%, and 35.52% at MAE, RMSE, and MAPE, respectively. It shows that the ABi-GRU in the STRGCN model plays an important role in temporal correlation extraction.

The STRGCN-S model removes the DFRGCN module for extracting spatial correlations and uses ABi-GRU to capture temporal correlations. Compared with the GRU model, the STRGCN-S model has an improvement in short-term prediction capability to a certain extent. Compared with the STRGCN-S model, the STRGCN model is reduced by 3.47%, 2.89%, and 4.63% at MAE, RMSE, and MAPE, respectively, on the PeMS04 dataset. On the PeMS07 dataset, the STRGCN model is reduced by 5.61%, 2.98%, and 5.67% at MAE, RMSE, and MAPE respectively. And on the PeMS08 dataset, the STRGCN model is reduced by 2.72%, 3.73%, and 4.01% at MAE, RMSE, and MAPE, respectively.

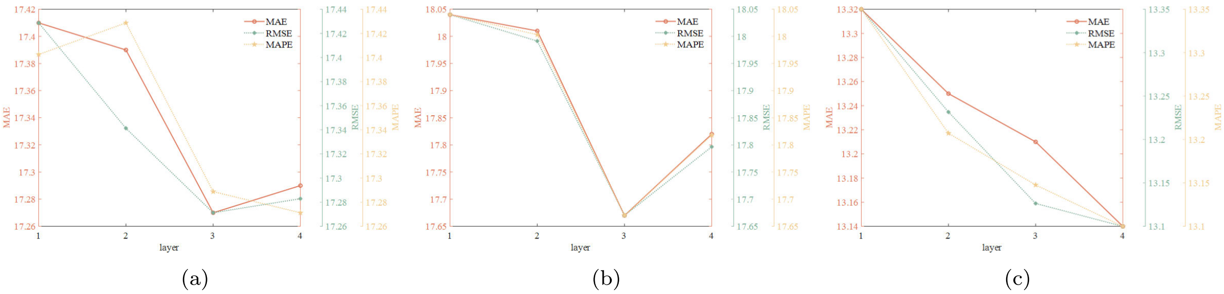


FIGURE 10. Model prediction error with different DFRGCN layers. (a) PeMS04. (b) PeMS07. (c) PeMS08.

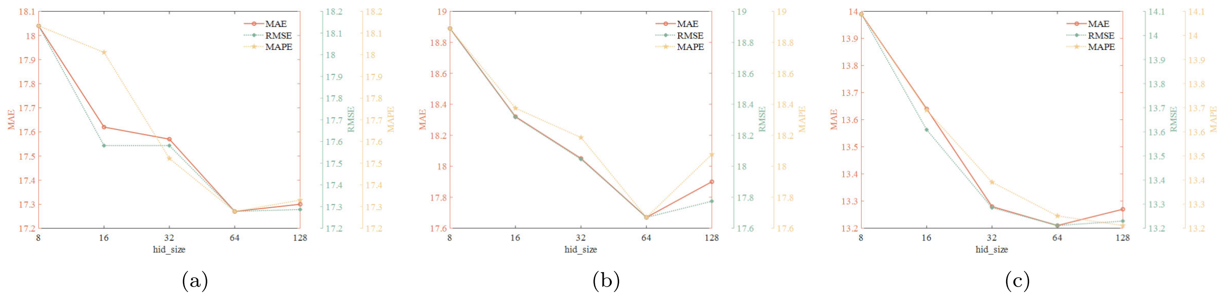


FIGURE 11. Model prediction error with different numbers of Bi-GRU hidden layer neurons. (a) PeMS04. (b) PeMS07. (c) PeMS08.

TABLE 4. Results of the ablation study on the PeMS04, PeMS07, and PeMS08 datasets.

Dataset	Evaluation metrics	GRU	GCN	STRGCN-T	STRGCN-S	STRGCN
PeMS04	MAE	17.95	31.80	25.88	17.89	**17.27**
	RMSE	28.76	47.28	38.89	28.75	**27.92**
	MAPE(%)	12.21	24.17	17.74	12.1	**11.54**
PeMS07	MAE	18.17	34.92	29.71	18.72	**17.67**
	RMSE	28.76	52.81	43.8	29.16	**28.29**
	MAPE(%)	7.74	16.09	13.35	7.93	**7.48**
PeMS08	MAE	13.89	34.61	21.12	13.58	**13.21**
	RMSE	21.45	50.41	31.66	21.44	**20.64**
	MAPE(%)	8.98	27.89	12.98	8.72	**8.37**

This indicates that the DFRGCN module plays a role in extracting spatial correlation, which helps to improve the prediction accuracy of the model.

In summary, we can infer that the STRGCN model learns the spatial structure of the traffic network through the DFRGCN model, and the temporal correlation of traffic flow is obtained through the ABi-GRU, which is combined with the spatio-temporal correlation extraction to improve the prediction performance of the model.

VI. CONCLUSION AND PROSPECT

In this paper, a Spatio-Temporal Residual Graph Convolution Network (STRGCN) is proposed for short-term traffic flow prediction. To address the challenges of the spatio-temporal short-term traffic flow prediction task and the problems of existing networks. The model first uses a deep full residual graph convolution network to capture the spatial structure of the traffic network, and then uses a bidirectional gated recurrent unit based on the attention mechanism to learn the temporal correlation of traffic flow, and finally conducts experimental analysis on the PeMS04, PeMS07, and PeMS08 datasets. Meanwhile, the experiments on the effect of hyperparameters and ablation study are conducted on the STRGCN model on these three datasets to demonstrate the role of hyperparameters and each module in the model. The results show that the STRGCN model has better prediction performance on these three datasets compared with the baseline models and can efficiently capture the spatio-temporal features of the traffic network and accomplish more accurate short-term traffic flow prediction. The model provides a new idea and technique in the field of traffic flow prediction, with strong application potential and value.

In practice, traffic flow is usually affected by weather, special holidays, etc. and thus shows different features. Therefore, in future traffic flow prediction work, we can take these factors into account to further validate our proposed STRGCN model.

REFERENCES

[1] S. Guo, Y. Lin, N. Feng, C. Song, and H. Wan, "Attention based spatial-temporal graph convolutional networks for traffic flow forecasting," in *Proc. AAAI*, vol. 33, 2019, pp. 922–929, doi: 10.1609/aaai.v33i01.3301922.

[2] X. Shi, H. Qi, Y. Shen, G. Wu, and B. Yin, "A spatial–temporal attention approach for traffic prediction," *IEEE Trans. Intell. Transp. Syst.*, vol. 22, no. 8, pp. 4909–4918, Aug. 2021, doi: 10.1109/TITS.2020.2983651.

[3] L. Zhu and N. Laptev, "Deep and confident prediction for time series at Uber," in *Proc. IEEE Int. Conf. Data Mining Workshops (ICDMW)*, Nov. 2017, pp. 103–110, doi: 10.1109/ICDMW.2017.19.

[4] A. Yu, H. Yang, T. Xu, B. Yu, Q. Yao, Y. Li, T. Peng, H. Guo, J. Li, and J. Zhang, "Long-term traffic scheduling based on stacked bidirectional recurrent neural networks in inter-datacenter optical networks," *IEEE Access*, vol. 7, pp. 182296–182308, 2019, doi: 10.1109/ACCESS.2019.2959303.

[5] J. Liu and W. Guan, "A summary of traffic flow forecasting methods," *J. Highway Transp. Res. Develop.*, vol. 21, no. 3, pp. 82–85, 2004.

[6] F. M. Yang, "A review of traffic flow prediction methods based on artificial neural networks," *J. Highway Transp. Res. Develop.*, vol. 37, no. 1, pp. 130–135, 2020.

[7] Q. F. Ma, "Improved PSO optimized BP neural network short-term traffic flow prediction," *Comput. Simul.*, vol. 36, no. 4, pp. 94–98, 2019.

[8] Q. Yu, Y. Chen, Q. Zhang, L. Li, and W. Ma, "Short-term traffic flow prediction based on IWOA-WNN," in *Proc. 33rd Chin. Control Decis. Conf. (CCDC)*, May 2021, pp. 899–904.

[9] M. Méndez, M. G. Merayo, and M. Nuñez, "Long-term traffic flow forecasting using a hybrid CNN-BiLSTM model," *Eng. Appl. Artif. Intell.*, vol. 121, May 2023, Art. no. 106041, doi: 10.1016/j.engappai.2023.106041.

[10] 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, doi: 10.1109/TITS.2019.2935152.

[11] D. J. Jeffery, K. Russam, and D. I. Robertson, "Electronic route guidance by autoguide: The research background," *Traffic Eng. Control.*, vol. 28, pp. 525–529, Oct. 1987.

[12] X. J. Zhang, Y. Tao, and G. N. Zhang, "ACapsGRU-based short-term traffic flow forecasting research," *J. Huazhong Univ. Sci. Technol.*, vol. 50, no. 4, pp. 51–56, 2022.

[13] I. Okutani and Y. J. Stephanedes, "Dynamic prediction of traffic volume through Kalman filtering theory," *Transp. Res. B, Methodol.*, vol. 18, no. 1, pp. 1–11, 1984, doi: 10.1016/0191-2615(84)90002-X.

[14] W. Wu, H. D. Liu, and W. G. Zhou, "KF-MA model for short-term traffic flow prediction based on SCATS data," *Appl. Mech. Mater.*, vols. 44–47, pp. 3418–3422, Dec. 2010, doi: 10.4028/WWW.SCIENTIFIC.NET/AMM.44-47.3418.

[15] C. Lin, "Research on short-term traffic flow prediction algorithm based on K-nearest neighbour non-parametric regression," M.S. thesis, Dept. Commun. Inf. Eng., Electron. Sci. Technol. Univ., Chengdu, China, 2015.

[16] P. C. Yuan and T. L. Zhou, "ARIMAX-based short-term forecasting model for urban road traffic flows," *Intell. Comput. Appl.*, vol. 11, no. 10, pp. 12–19, 2021.

[17] Y. Chao, C. Wang, and X. Wang, "Short-term traffic flow prediction on urban roads based on spatio-temporal node selection and deep learning," *J. Comput. Appl.*, vol. 40, no. 5, pp. 1488–1493, 2020.

[18] J. Zhang, S. Zhao, Y. Wang, and X. Zhu, "Improved social emotion optimization algorithm for short-term traffic flow forecasting based on back-propagation neural network," *J. Shanghai Jiaotong Univ.*, vol. 24, no. 2, pp. 209–219, Apr. 2019.

[19] H. K. Xu, W. Zhao, and M. Yang, "Improved BPNN-based traffic accident duration prediction for highways," *J. East China Jiaotong Univ.*, vol. 37, no. 5, pp. 60–65, 2020.

[20] X. Miao, Z. Y. Wang, and B. Wu, "Two-layer BPNN prediction model for bus arrival times considering the state of the preceding section," *J. Transp. Syst. Eng. Inf. Technol.*, vol. 20, no. 2, pp. 127–133, 2020.

[21] C. Adam, A. Aliotti, F. D. Malliaros, and P.-H. Cournède, "Dynamic monitoring of software use with recurrent neural networks," *Data Knowl. Eng.*, vol. 125, Jan. 2020, Art. no. 101781, doi: 10.1016/j.datak.2019.101781.

[22] L. Han and Y. Huang, "Short-term traffic flow prediction of road network based on deep learning," *IET Intell. Transp. Syst.*, vol. 14, no. 6, pp. 495–503, Jun. 2020, doi: 10.1049/iet-its.2019.0133.

[23] X. Ma, Z. Tao, Y. Wang, H. Yu, and Y. Wang, "Long short-term memory neural network for traffic speed prediction using remote microwave sensor data," *Transp. Res. C, Emerg. Technol.*, vol. 54, pp. 187–197, May 2015, doi: 10.1016/j.trc.2015.03.014.

[24] X. Chen, H. Chen, Y. Yang, H. Wu, W. Zhang, J. Zhao, and Y. Xiong, "Traffic flow prediction by an ensemble framework with data denoising and deep learning model," *Phys. A, Stat. Mech. Appl.*, vol. 565, Mar. 2021, Art. no. 125574, doi: 10.1016/j.physa.2020.125574.

[25] B. Pu, Y. Liu, N. Zhu, K. Li, and K. Li, "ED-ACNN: Novel attention convolutional neural network based on encoder–decoder framework for human traffic prediction," *Appl. Soft Comput.*, vol. 97, Dec. 2020, Art. no. 106688, doi: 10.1016/j.asoc.2020.106688.

[26] Y. G. Li, R. Yu, and C. Shahabi, "Diffusion convolutional recurrent neural networks: Data-driven traffic forecasting," in *Proc. Int. Conf. Learn. Represent.*, 2017, pp. 1–16.

[27] B. Yu, H. Yin, and Z. Zhu, "Spatio-temporal graph convolutional networks: A deep learning framework for traffic forecasting," in *Proc. 27th Int. Joint Conf. Artif. Intell.*, Jul. 2018, pp. 3634–3640.

[28] G. Shen, W. Zhou, W. Zhang, N. Liu, Z. Liu, and X. Kong, "Bidirectional spatial–temporal traffic data imputation via graph attention recurrent neural network," *Neurocomputing*, vol. 531, pp. 151–162, Apr. 2023, doi: 10.1016/j.neucom.2023.02.017.

[29] Y. Gu and L. Deng, "STAGCN: Spatial–temporal attention graph convolution network for traffic forecasting," *Mathematics*, vol. 10, no. 9, p. 1599, May 2022, doi: 10.3390/math10091599.

[30] X. Wang, R. Zeng, F. Zou, L. Liao, and F. Huang, "STTF: An efficient transformer model for traffic congestion prediction," *Int. J. Comput. Intell. Syst.*, vol. 16, p. 2, Jan. 2023, doi: 10.1007/s44196-022-00177-3.

[31] D. Ma, X. B. Song, J. Zhu, and W. Ma, "Input data selection for daily traffic flow forecasting through contextual mining and intra-day pattern recognition," *Exp. Syst. Appl.*, vol. 176, Aug. 2021, Art. no. 114902, doi: 10.1016/j.eswa.2021.114902.

[32] D. Ma, X. Song, and P. Li, "Daily traffic flow forecasting through a contextual convolutional recurrent neural network modeling inter- and intra-day traffic patterns," *IEEE Trans. Intell. Transp. Syst.*, vol. 22, no. 5, pp. 2627–2636, May 2021, doi: 10.1109/TITS.2020.2973279.

[33] Y. M. Zhang, *Analysis and Application of Traffic Flow Time Series*. Beijing, China: Science Press, 2019.

[34] D. H. Sun, C. Tian, and W. N. Liu, *Theory and Methods of Coordinated Road Traffic Flow*. Beijing, China: Science Press, 2019.

[35] J. C. Wang, Y. Zhang, and Y. L. Hu, "A review of traffic prediction based on graph convolutional networks," *J. Beijing Univ. Technol.*, vol. 47, no. 8, pp. 954–970, 2021.

[36] C. W. J. Granger, "Investigating causal relations by econometric models and cross-spectral methods," *Econometrica*, vol. 37, no. 3, pp. 424–438, Aug. 1969.

[37] Z. R. Wei, "Deep learning-based traffic forecasting," M.S. thesis, Dept. Electron. Inf. Eng., Beijing Jiaotong Univ., Beijing, China, 2019.

[38] N. Li, S. Jia, and Q. Li, "Traffic message channel prediction based on graph convolutional network," *IEEE Access*, vol. 9, pp. 135423–135431, 2021, doi: 10.1109/ACCESS.2021.3114691.

[39] B. B. Xu, K. T. Cen, and J. J. Huang, "A review of graph convolutional networks," *Chin. J. Comput.*, vol. 43, no. 5, pp. 755–780, 2020.

[40] Y. Y. Wu, L. N. Zhao, and Z. X. Yuan, "CNN-GRU ship traffic flow prediction model based on attention mechanism," *J. Dalian Maritime Univ.*, vol. 49, no. 1, pp. 75–84, 2023.

[41] S. Wang, C. Shao, J. Zhang, Y. Zheng, and M. Meng, "Traffic flow prediction using bi-directional gated recurrent unit method," *Urban Informat.*, vol. 1, no. 1, p. 16, Dec. 2022, doi: 10.1007/s44212-022-00015-z.

[42] C. Wang, R. Tian, and J. Hu, "A trend graph attention network for traffic prediction," *Inf. Sci.*, vol. 623, pp. 275–292, Apr. 2023, doi: 10.1016/j.ins.2022.12.048.

[43] Z. Zheng, Y. Yang, J. Liu, H.-N. Dai, and Y. Zhang, "Deep and embedded learning approach for traffic flow prediction in urban informatics," *IEEE Trans. Intell. Transp. Syst.*, vol. 20, no. 10, pp. 3927–3939, Oct. 2019, doi: 10.1109/TITS.2019.2909904.

[44] F. Rempe, G. Huber, and K. Bogenberger, "Spatio-temporal congestion patterns in urban traffic networks," *Transp. Res. Proc.*, vol. 15, pp. 513–524, Jan. 2016, doi: 10.1016/j.trpro.2016.06.043.

[45] Z. Wu, S. Pan, G. Long, J. Jiang, and C. Zhang, "Graph WaveNet for deep spatial–temporal graph modeling," in *Proc. 28th Int. Joint Conf. Artif. Intell.*, Aug. 2019, pp. 1907–1913.

QINGYONG ZHANG received the B.E. and Ph.D. degrees from the School of Automation, Wuhan University of Technology, Wuhan, China, in 2009 and 2021, respectively.

She is currently a Professorial Senior Laboratory Master with the School of Automation, Wuhan University of Technology. Her current research interests include machine learning, data mining, and intelligent transport.

HUIWEN XIA received the B.E. and M.E. degrees from the School of Automation, Wuhan University of Technology, Wuhan, China, in 2014 and 2017, respectively.

She is currently a Lecturer with the School of Automation, Wuhan University of Technology. Her current research interests include machine learning and neural networks.

MEIFANG TAN received the B.E. degree from the School of Automation, Wuhan University of Technology, Wuhan, China in 2021, where she is currently pursuing the master's degree.

Her current research interests include machine learning and neural networks.

WANFENG CHANG received the B.E. degree from the School of Automation, Wuhan University of Technology, Wuhan, China, in 2021, where he is currently pursuing the master's degree.

His current research interests include machine learning and neural networks.

CHANGWU LI received the B.E. degree from the School of Automation, Wuhan University of Technology, Wuhan, China, in 2020, where he is currently pursuing the master's degree.

His current research interests include machine learning and neural networks.

MINGLONG LI received the B.E. degree from the School of Electrical and New Energy, China Three Gorges University, Yichang, China, in 2022. He is currently pursuing the master's degree with the School of Automation, Wuhan University of Technology, Wuhan, China.

His current research interests include intelligent control and intelligent systems.