WILEY | Hindawi

Research Article

A Multifeature Fusion Short-Term Traffic Flow Prediction Model Based on Deep Learnings

Chunxu Chai [iD],[1] **Chuanxiang Ren** [iD],[1] **Changchang Yin** [iD],[2] **Hui Xu**,[1] **Qiu Meng**,[3] **Juan Teng**,[1] **and Ge Gao** [iD][1]

[1]*College of Transportation, Shandong University of Science and Technology, Qingdao 266590, China*
[2]*College of Electrical Engineering and Automation, Shandong University of Science and Technology, Qingdao 266590, China*
[3]*College of Electronic and Information Engineering, Shandong University of Science and Technology, Qingdao 266590, China*

Correspondence should be addressed to Chuanxiang Ren; renchx@sdust.edu.cn

Short-term traffic flow prediction is an important component of intelligent transportation systems, which can support traffic trip planning and traffic management. Although existing predicting methods have been applied in the field of traffic flow prediction, they cannot capture the complex multifeatures of traffic flows resulting in unsatisfactory short-term traffic flow prediction results. In this paper, a multifeature fusion model based on deep learning methods is proposed, which consists of three modules, namely, a CNN-Bidirectional GRU module with an attention mechanism (CNN-BiGRU-attention) and two Bidirectional GRU modules with an attention mechanism (BiGRU-attention). The CNN-BiGRU-attention module is used to extract local trend features and long-term dependent features of the traffic flow, and the two BiGRU-attention modules are used to extract daily and weekly periodic features of the traffic flow. Moreover, a feature fusion layer in the model is used to fuse the features extracted by each module. And then, the number of neurons in the model, the loss function, and other parameters such as the optimization algorithm are discussed and set up through simulation experiments. Finally, the multifeature fusion model is trained and tested based on the training and test sets from the data collected from the field. And the results indicate that the proposed model can better achieve traffic flow prediction and has good robustness. Furthermore, the multifeature fusion model is compared and analyzed against the baseline models with the same dataset, and the experimental results show that the multifeature fusion model has superior predictive performance compared to the baseline models.

1. Introduction

With the development of urbanization, the number of population and motor vehicles in cities is increasing. While the demand for travel, especially in the morning and evening rush hours, often makes the road utilization rate saturated, resulting in urban "traffic diseases." In this case, in order to solve the urban "traffic disease," intelligent transportation system (ITS) was developing [1–4]. And with the development of big data technology, ITS has started to change into data-driven ITS [5]. Among them, short-term traffic flow prediction is one of the core components of ITS, which provides the basis for traffic management, traffic control, and traffic guidance, as well as support for travel decision of travelers. However, short-term traffic flow has complex stochastic and nonlinear characteristics, which brings great challenges to traffic flow prediction. And how to accurately predict short-term traffic flow has been a hot topic of concern for scholars in the field of traffic engineering.

The methods proposed in the early studies on short-term traffic flow forecasting mainly consist of three main methods, parametric methods, nonparametric methods, and combined methods, which include both parametric and nonparametric methods. Parametric methods include the autoregressive integrated moving average model (ARIMA) and its variants [6, 7]. Nonparametric methods include K-nearest neighbor nonparametric regression methods (KNN) [8], Kalman filters (KF) [9], support vector

machines (SVR) [10], and artificial neural networks (ANN) [11]. Combined methods are a combination of two or more methods [12–14].

However, due to the development of data-driven ITS, especially the development and widespread use of traffic information collection technologies, such as induction detectors, geomagnetic detectors, radio frequency identification technology, radar detection, video detection, and floating vehicle detection [15–19], provide a large amount of data for traffic flow prediction. In this case, there are difficulties for parametric and nonparametric methods to deal with big traffic data. Therefore, deep learning methods [20–22], which have powerful data feature mining and nonlinear data fitting capabilities, have been applied to traffic flow prediction and achieved some results [23–26]. However, existing deep learning-based methods for traffic flow prediction mainly consider the spatial and temporal correlation of traffic flow, without fully considering the complex characteristics of traffic flow such as daily and weekly periodicity. In addition, although some combined deep learning methods use several different single models to extract multiple features of traffic flow, such as spatiotemporal correlation and periodicity, in fact, the spatiotemporal correlation and periodicity of traffic flow are a whole and should be considered comprehensively in prediction model. Based on this, this paper designs a multifeature fusion model based on deep learning methods that considers the periodic features of traffic flow for traffic flow prediction, and the main contributions are summarized as follows:

(1) A fusion feature model considering the periodic features of traffic flow is proposed, namely, multifeature fusion model. In the model, the CNN-BiGRU module is designed, which treats the spatiotemporal features of traffic flow as a whole, where 1DCNN and BiGRU are used to extract the local trend features and long temporal dependencies trend features of traffic flow, respectively.

(2) In multifeature fusion model, two two-layer BiGRU modules are designed to extracting the daily and weekly periodicity features of traffic flow, respectively.

(3) In order to improve the prediction performance of the multifeature fusion model, an attention mechanism is designed for the CNN-BiGRU and the two-layer BiGRU modules to adaptively make each module pay attention to the importance of the temporal and periodic features at different times.

(4) The multifeature fusion model is validated by simulating the traffic flow collected in the field, and the experiments' results show that the prediction performance of the multifeature fusion model is better than that of the baseline model.

2. Literature Review

In general, existing traffic flow prediction methods can be classified into parametric methods, nonparametric methods, deep learning methods, and combined methods.

2.1. Parametric Methods. The parametric method is a modelling approach where the structure of the model is predetermined based on theory, and the parameters of the model can be calibrated by realistic traffic flow data. Levin and Tsao [27] applied a time series analysis method to predict the morning peak period traffic on a motorway and found that the ARIMA (0,1,1) model was statistically significant. Zhang et al. [28] developed a hybrid model, where spectral analysis techniques are invoked to extract the daily and weekly periodicity of traffic flows, and the ARIMA model is used to extract the general time trend characteristics of traffic flows. Subsequently, a number of ARIMA variants were applied in traffic flow prediction. For instance, Kohonen self-organizing ARIMA, an autoregressive sliding average model with seasonality, and spatiotemporal autoregressive sliding average model were also used for traffic flow forecasting and achieved good results [29–31].

2.2. Nonparametric Methods. Due to the strong randomness and nonlinearity of the state changes in traffic flow, the traffic flow prediction results using parametric methods have a certain degree of deviation from the actual traffic flow. Therefore, some nonparametric methods gradually replace parametric methods in traffic flow prediction. Specifically, Ryu et al. [32] proposed a traffic flow prediction model that considering the spatiotemporal information associated with the predicted road section. The spatiotemporal information with the highest correlation to the predicted road section is first selected using a greedy algorithm, and then the traffic flow is predicted using KNN. Yan and Lv [33] proposed a hybrid classification and regression tree k-nearest neighbor model to predict short-term taxi demand. Okutani and Stephanedes [34] proposed two prediction models based on Kalman filter theory to predict traffic flow on streets within Nagoya. Guo et al. [35] proposed a hierarchical Kalman filter-based autoregressive moving average and generalized autoregressive conditional heteroskedasticity model for traffic flow velocity prediction. Hu et al. [36] proposed a hybrid model to forecast the short-term traffic flow based on particle swarm optimization (PSO) and support vector regression (SVR), in which PSO is used to find the optimal parameters of the SVR model. Lu and Zhou [37] proposed a Kalman filter traffic flow prediction model that takes into account structural deviations, where a polynomial is used to describe the evolutionary trend of structural deviations in traffic flow, and a Kalman filter model is used to describe the historical trend of traffic flow. Jiang et al. [38] proposed a support vector machine model with radial basis functions as kernel functions to predict traffic flow speed, and the experiment results showed that the prediction accuracy of the model was better than that of the traditional model. Wang and Shi [39] proposed a chaotic wavelet analysis-support vector machine model (C-WSVM), and the results showed that the C-WSVM model has better prediction performance and practicality. Feng et al. [40] proposed a new short-term traffic flow prediction model based on adaptive multicore support vector machine with spatiotemporal correlation. Wang et al. [41] proposed a combined support vector machine model to forecast short-term metro ridership, which includes a vector machine overall online model (SVMOOL)

and a vector machine partial online model (SVMPOL). The SVMOOL model obtains the periodic characteristics of passenger flow, and SVMPOL obtains the nonlinear characteristics of traffic flow.

ANN [42] was regarded as another popular method for traffic flow prediction due to its ability to handle large amounts of multidimensional data, flexibility of model structure, and learning and generalization capabilities. And ANN combined with error backpropagation algorithm, i.e., Backpropagation Neural Network (BPNN) [43], was gradually applied to traffic flow prediction, and subsequently, a short-time traffic flow prediction model incorporating wavelet analysis and BP neural network approach [44] was applied to short-time traffic flow prediction. Then, an adaptive differential evolution algorithm optimized BPNN [45] was applied to short-time traffic flow prediction models. All these methods have achieved good results.

2.3. Deep Learning Methods.
With the development of data collection and processing technology, traffic big data has emerged. However, the traditional nonparametric methods have difficulties in processing multisource data [46], and the short-term traffic flow prediction methods have started to shift from nonparametric methods to deep learning methods [24, 26, 47, 48]. For instance, Huang et al. [49] designed a combined prediction model including a deep belief network with unsupervised learning at the bottom and a multitask learning (MTL) layer for supervised prediction, in which the top multitask learning layer can leverage the weight sharing in the DBN to provide better results in support of prediction. Lv et al. [50] proposed a stacked autoencoder model that is trained in a greedy hierarchical approach for training to learn traffic flow features.

One of the difficulties in short-term traffic flow prediction is to obtain spatiotemporal correlation between traffic flow data. In terms of temporal characteristics, recurrent neural networks (RNNs) are a deep learning structure mainly applied to process time series data. RNNs have the function of temporal memory and can be applied to the field of correlation prediction of time series data [51]. However, traditional RNNs cannot tap the long-term dependence properties among traffic flow data due to the gradient disappearance and gradient explosion problems, so Ma et al. [52] applied long- and short-term memory (LSTM) to the traffic flow prediction. Subsequently, Zhao et al. [53] proposed a two-dimensional LSTM network consisting of many memory units with considered spatiotemporal correlations, and the experimental results showed that the proposed network had better prediction performance compared with traditional prediction methods. Wang et al. [54] proposed a deep learning framework based on paths. In the framework, the road network is divided into critical paths, and then the bidirectional long and short-term memory network is used to model the traffic flow of each critical path. Cui et al. [55] proposed a stacked bidirectional and unidirectional LSTM network structure for predicting road network traffic with missing values. Zheng and Huang [56] proposed a traffic flow prediction model based on LSTM,

and experimental results showed that the prediction performance of the proposed model outperformed the classical model. GRU, which is a well-known variant structure of the LSTM, has also been applied to traffic flow prediction [57].

In terms of spatial properties, CNN is also a typical structure in deep learning. It is a feedforward neural network for solving problems with grid-like structured data, which not only can reduce the complexity of the model while accurately extracting data features, but also can better extract spatial correlations between traffic flow data [58]. Zhang et al. [59] proposed a CNN model for short-term traffic flow prediction, where the optimal input to the model is a spatial-temporal feature selection algorithm, and experimental results showed that the model outperformed the baseline model. An et al. [60] used a fuzzy convolutional neural network based traffic flow prediction method, which for the first time applied CNN to uncertain traffic incident information and used a fuzzy approach to generalize traffic incident characteristics. Tian et al. [61] proposed a hybrid lane occupancy prediction model called 2LayersCapsNet, which combines an improved capsule network and CNN.

2.4. Combined Methods.
Combined models should be useful when a single specified model fails to exhibit good predicting performance, which is a common situation in complex data forecasting [46]. It is difficult for a single forecasting model to capture both the strong complexity and the strong variability of traffic flow, so the proposal of a combined predicting model is necessary. Specifically, to exploit the good linear fitting capability of ARIMA models and the powerful nonlinear relational mapping capability of artificial neural network models, Li et al. [62] proposed a combined ARIMA and radial basis function artificial neural network model to predict short-term traffic flows. Yao et al. [63] proposed a linear hybrid method and a nonlinear hybrid method to predict short-term traffic flows and classified the traffic flow data into similar, unstable, and irregular components. Among them, autoregressive integrated moving average and generalized autoregressive conditional heteroskedasticity models were used to predict the similar and fluctuating components, and Markov models with state membership and wavelet neural networks were used to predict the irregular component. Li et al. [64] analyzed the correlation between the predicted and historical time windows based on the grey correlation coefficient method and used the rank index method to establish a combined prediction model based on ARIMA, BPNN, and SVR developed. A neural network training algorithm combining exponential smoothing and the Levenberg-Marquardt algorithm was proposed to improve the neural networks generalization previously used for short-term traffic predicting [65]. Liu et al. [13] proposed a hybrid forecasting model based on a combination of neural network and KNN methods for short-term traffic predicting. Gu et al. [66] proposed a model incorporating deep learning to predict lane level speeds. In the model, firstly use entropy-based grey correlation analysis to select the lanes with the highest correlation with the predicted lanes to extract spatial features, and secondly,

combine LSTM and GRU to build a two-layer deep learning framework to extract temporal features of traffic flow. The experiments results showed that the model outperformed the baseline model in prediction. Ma et al. [67] proposed a novel deep learning-based approach to daily traffic flow prediction incorporating contextual factors. Firstly, a specific CNN is used to extract daytime and intraday traffic flow features, secondly, the extracted features are used as input to an LSTM to learn the temporal features of the traffic flow, and finally, the traffic flow is predicted by combining the contextual information of historical days. Experiments results showed that the robustness and prediction performance of the model outperformed the benchmark model.

With the development of deep learning, especially the proposed and successful application of attention mechanism [68], it has received attention from scholars in the field of traffic, and some results of applying it in combination with CNN or variant RNN (LSTM and GRU) for short-term traffic flow prediction have emerged. For example, Liu et al. [69] proposed a CNN model based on an attention mechanism to predict traffic flow speed, where the input to the model is a three-dimensional data matrix consisting of traffic flow speed, flow rate, and time occupation, and the extraction of spatiotemporal features is done by convolutional units, and the proposed model has better prediction performance when compared with existing models for simulation experiments. Wu et al. [70] proposed a traffic flow prediction model including a data preprocessing module and a traffic flow prediction module, where the data preprocessing module is to repair missing values in the dataset, and the traffic flow prediction module is a model of a combined LSTM deep learning method based on an attention mechanism, and experimental results show that the prediction performance of the model outperforms other deep learning methods (RNN and CNN). Ma et al. [71] proposed a fuzzy logic-based hybrid model based on the complementary advantages of nonparametric and deep learning methods. Firstly, the model uses two submodels, KNN and LSTM, to extract features on the spatiotemporal correlation of traffic flow and the influence of specific contextual factors on traffic flow, and secondly, dynamic weights based on the fusion mechanism are used to optimize the hybrid model, and simulation experiments show that the model has better prediction and robustness than other state-of-the-art models. Ren et al. [72] proposed a combined deep learning prediction (CDLP) model, which consists of two parallel single deep learning models, that is, a CNN-LSTM-attention model and a CNN-GRU-attention model. In addition, a dynamic optimal weighting combination algorithm was proposed to combine the outputs of the two single models, and experimental results showed that this model has better prediction performance and robustness than the state-of-the-art prediction models.

In summary, as the research on short-term traffic flow prediction continues to grow, combined prediction models have received more and more attention, and in particular, the application of combined deep learning models has achieved greater success. However, most of the researches are based on the fusion of multiple single combination methods or just obtaining a fusion model of simple spatiotemporal characteristics of traffic flow, which cannot reflect the unified whole of spatiotemporal correlation and periodicity of traffic flow. In this paper, we analyze the complex characteristics of traffic flow, including the relationship between spatiotemporal and periodic features, and apply CNN, Bidirectional GRU, and Attention mechanism to build a multifeature fusion model for short-time traffic flow prediction.

3. Method

3.1. CNN. CNN is a deep feed-forward neural network, which mainly consists of a convolutional layer, a pooling layer, and a fully connected layer [73]. The convolutional layer is the most important part of the CNN, where the local features of the input data are obtained in the form of sliding filters, and the number of convolutional kernels in the convolutional layer corresponds to the number of output features in the convolutional layer. Typically, CNN models contain multiple convolutional layers, and the network can generate an excessive number of parameters. To reduce the number of parameters, the pooling layer usually performs a downsampling operation with the output features of the convolutional layer while keeping the overall features unchanged, in order to extract important features and prevent overfitting of the model. The fully connected layer is usually at the end of the CNN, and its main role is to spread the features obtained by convolution and pooling into a feature vector for classification and regression.

3.2. Bidirectional GRU. In order to address the shortcomings of traditional RNNs, which ignore the long-term dependence of time series, LSTM and GRU have been proposed one after another. GRU and LSTM networks have not only the function of short-term memory, but also the function of long-term memory. In particular, the GRU is a further simplification of the LSTM [74], from the three gating units of the LSTM to two gate structures (update gate and reset gate), which further improves the operational efficiency of the network due to the simplified number of gates. The structure of the GRU unit is shown in Figure 1, where the purple line indicates the update gate, and the red line indicates the reset gate, defined as z_t and r_t respectively.

The role of the update gate in the GRU is to determine whether the hidden layer state h_{t-1} is updated to a new hidden layer state h_t, and the role of the reset gate is to control the extent to which the hidden layer state h_{t-1} is discarded at moment $t-1$. Equation (1) represent the computation process for each state within each time step in the GRU.

$$
\begin{aligned}
z_t &= \sigma\left(W_z X_t + U_z h_{t-1} + b_z\right), \\
r_t &= \sigma\left(W_r X_t + U_r h_{t-1} + b_r\right), \\
g_t &= \tanh\left(W_g X_t + U_g\left(r_t \circ h_{t-1}\right) + b_g\right), \\
h_t &= \left(1 - z_t\right) \circ h_{t-1} + z_t \circ g_t,
\end{aligned}
\tag{1}
$$

Where $\circ$ represents the Hadamard product, X_t represents the input at moment t, W_z, W_h and W_g represent the weight
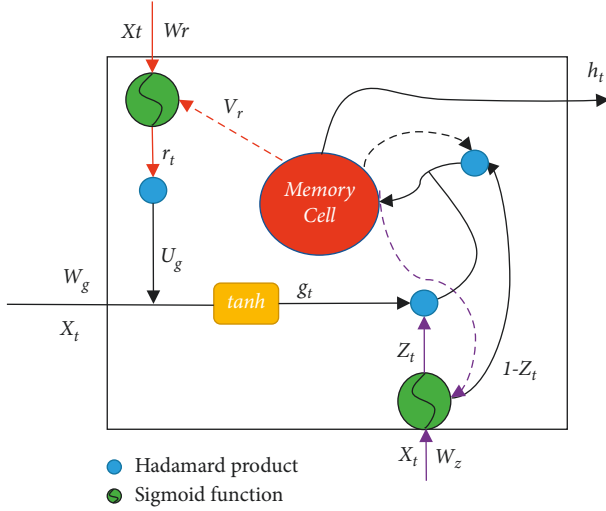
FIGURE 1: The structure of GRU.

matrix associated with the input, U_z, U_h and U_g represent the weight matrix associated with h_{t-1}, and b_z, b_r, b_h and b_g represent the bias.

Based on GRU network, bidirectional GRU network has been further developed [75]. The structure of a bidirectional GRU network is made up of two GRU layers stacked in different directions, which is shown in Figure 2. In the figure, x_t is the input to the GRU, h_f is the output of the forward GRU layer, and h_b is the output of the reverse GRU layer. The input to the BiGRU network contains two time series from the past and the future, and in each moment, the input time series is fed into the two opposite GRU layers, and the outputs $[h_1, h_2, h_3, h_4]$ are obtained by the joint determination of these two reverse GRU layers.

At each time node x_t, this network has two hidden layers containing opposite order. The neurons in one hidden layer are ordered from left to right, and the other hidden layer is ordered from right to left. To ensure that there are two hidden layers at any moment t, the network consumes twice the amount of storage to store parameters such as weights and offsets. The final output of the network is the fusion of the outputs of the two hidden layers to produce the final output. In addition, there is no information interaction between the two opposite hidden layers, and they are computed independently, but the state output vectors of both are combined at the final output to ensure that the unfolding graph is acyclic.

3.3. Attention Mechanism. The attention mechanism uses a method of assigning different weights to the input features of a model in order to highlight the important factors that influence the model. The function of the attention mechanism can be understood as the process of filtering important information from multiple pieces of information, focusing on the important information and ignoring the unimportant information. The process of focusing on the important information is also the process of calculating the weight coefficients, and the more important the information, the larger the weight coefficient assigned. The process of calculating the context vectors and weights for the application of the attention mechanism to a deep learning model is as follows:

Assuming that the output state of the hidden layer of the deep learning model is $h_1, h_2, \ldots, h_i, \ldots, h_t$, the context vector can be calculated as C_t:

$$C_t = \sum_{i=1}^{t} \alpha_{t,i} h_i, \tag{2}$$

where $\alpha_{t,i}$ denotes the attention parameter, the corresponding weight of h_i, and the sum of the weights is 1. The attention parameter can be calculated as

$$\alpha_{t,i} = \frac{\exp\left(e_{t,i}\right)}{\sum_{i=1}^{T} \exp\left(e_{t,i}\right)}, \tag{3}$$

where $e_{t,i}$ is the alignment model, which scores the input at moment i and the output at moment t. It is calculated as follows:

$$e_{t,i} = \tanh\left(W_a s_{t-1} + U_a h_i + b_a\right), \tag{4}$$

where W_a, U_a, and b_a are the parameters of the feedforward neural network, and s_{t-1} can be calculated as follows:

$$s_{t-1} = g\left(s_{t-2}, y_{t-2}, c_{t-1}\right), \tag{5}$$

where $g(\cdot)$ denotes the deep learning network.

Based on (5), the output of the attention mechanism can be calculated as

$$y_t = \mathrm{softmax}\left(s_t\right), \tag{6}$$

where softmax is the activation function.

4. Model

Realistic short-term traffic flow often exhibits complexity and randomness, which requires traffic flow prediction models that can tap into multiple features of traffic flow. CNNs can extract local trend features of traffic flows, while bidirectional GRU networks can obtain long-term dependent features of traffic flows not only in the past, but also in the future and can achieve temporal feature extraction by fusing past and future features. By fusing past and future features, temporal feature extraction can be achieved. At the same time, the attention mechanism enables the model to focus on important features. Based on this, this paper proposes a short-term traffic flow model based on a deep learning method of multifeature fusion, which consists of a CNN-BiGRU-attention module and two BiGRU-attention modules, and the model structure is shown in Figure 3. The CNN-BiGRU-attention module is composed of CNN, BiGRU network, and attention sequentially connected, where the CNN is composed of one convolutional layer. The CNN-BiGRU-attention module extracts traffic flow features
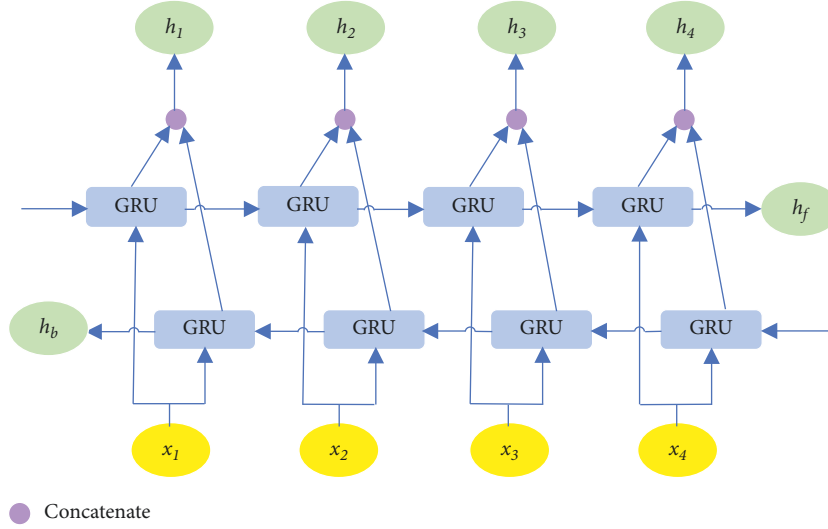
Figure 2: The structure of bidirectional GRU network.

by considering the local trend features extracted by CNN and the time-dependent features extracted by BiGRU as a whole. The two BiGRU-attention modules are used to obtain the weekly and daily features of the traffic flow data, respectively.

In addition, from a layer perspective, the model consists of an input layer, a hidden layer, a feature fusion layer, and an output layer. The input layer contains a parallel composition of historical time series, daily and weekly series, where the historical time series X_T is a sequence of traffic flows from time $t-n$ to t and can be represented as

$$X_T = \left(x_{t-n}, x_{t-n+1}, \ldots, x_t\right)^T, \tag{7}$$

where x_t is the traffic flow at time t.

The daily periodic traffic sequence X_T^d can be expressed as

$$X_T^d = \left(x_{t-n}^d, x_{t-n+1}^d, \ldots, x_t^d\right)^T, \tag{8}$$

where x_t^d indicates the traffic flow x_t corresponding to the previous day.

The weekly periodic traffic flow sequence can be expressed as

$$X_T^w = \left(x_{t-n}^w, x_{t-n+1}^w, \ldots, x_t^w\right)^T, \tag{9}$$

where x_t^g indicates the traffic flow x_t corresponding to the previous week.

The hidden layer contains three parallel CNN-BiGRU-attention layers with two BiGRU-attention layers. The 1DCNN is chosen as the convolution layer of the model due to the one-dimensional and periodic nature of the traffic flow sequence. The dropout layer is followed by the feature fusion layer, where the features of the traffic flow are fused and output to the output layer for prediction.

5. Experiment

5.1. Data Processing and Dataset. The collected cross-sectional traffic flow at the intersection of Shandong Road and

Minjiang Road in Qingdao, China, is used as the data set, containing 101 consecutive days of traffic flow data from February 1 to May 12, 2019, and a total of 29,088 raw pieces of data, and the interval for these data is 5 minutes. Then, the Lagrangian interpolation method is used to process the missing data and abnormal data. The data are then normalized using the maximum-minimum normalization method to obtain the dataset for the model. A total of 87 days of data from February 1 to April 28 in the dataset are used as the training set, and a total of 14 days of data from April 29 to May 12 are used as the test set.

5.2. Experimental Environment and Model Evaluation Index Selection. The software and hardware conditions of the experimental environment in this paper are shown in Table 1.

In order to evaluate the performance of the fused feature model, three evaluation metrics were chosen, namely, MAPE, MAE, and RMSE, which are calculated as follows:

$$MAPE = \frac{1}{n} \sum_{i=1}^{n} \left| \frac{y_i - \widehat{y}_i}{y_i} \right| \times 100,$$

$$MAE = \frac{1}{n} \sum_{i=1}^{n} |y_i - \widehat{y}_i|, \tag{10}$$

$$RMSE = \left(\sum_{i=1}^{n} \frac{(y_i - \widehat{y}_i)^2}{n} \right)^{1/2},$$

where n is the total number of samples in the test set, y_i is the actual value of the ith sample, and $\widehat{y}_i$ is the predicted value of the ith sample.

5.3. Model Parameter Settings

5.3.1. Loss Function Setting. The loss function quantifies how close a given neural network is to the ideal situation it is trained for. The mean absolute error function and the mean
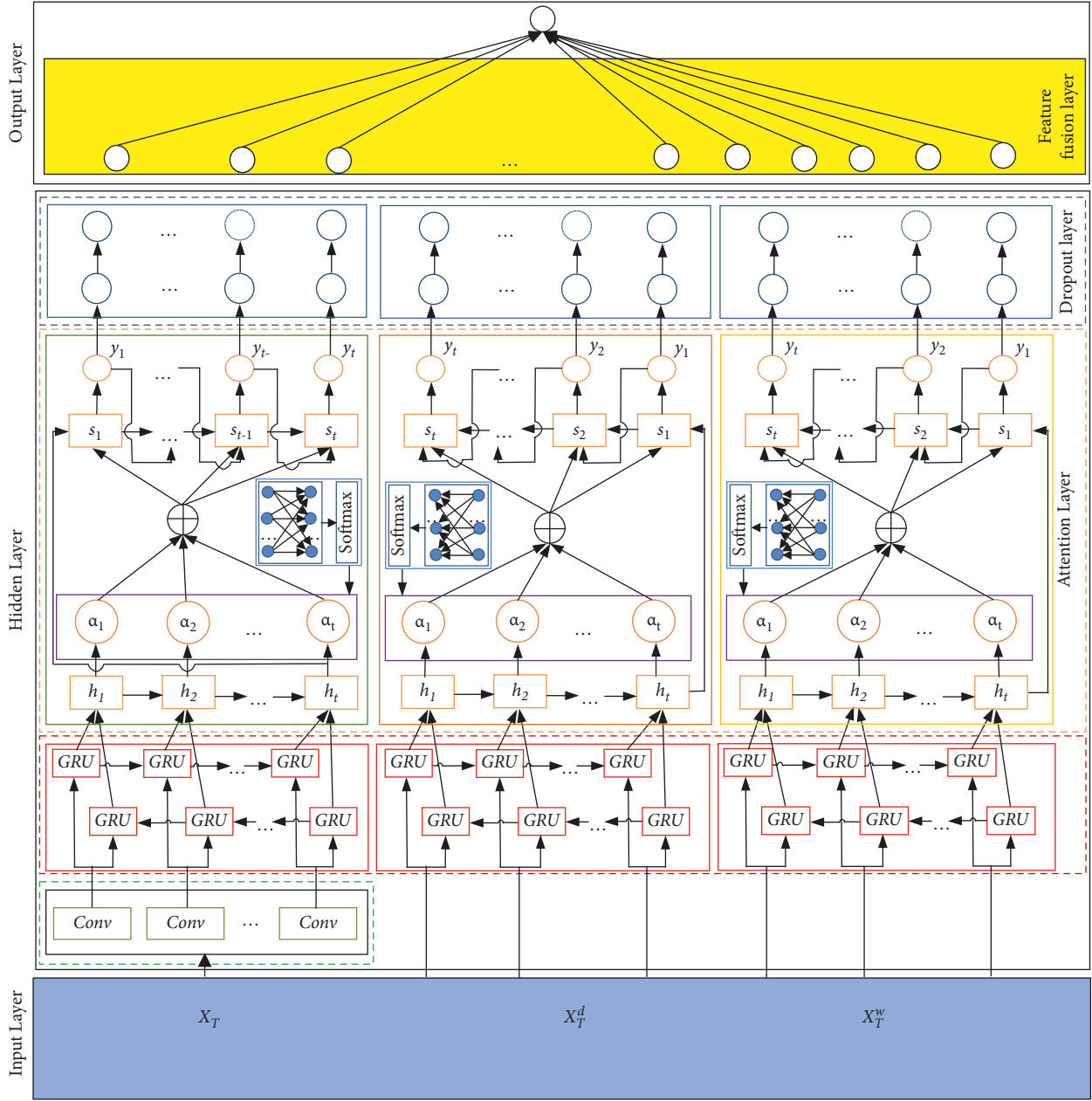
FIGURE 3: Structure of the multi-feature fusion model.

TABLE 1: Experimental environment.

Software and hardware configuration	Configuration parameter
CPU	Intel i5-8250U @1.60 GHz
RAM	8G
Programming language	Python 3.7.0
Deep learning framework	TensorFlow 1.14
Deep learning library	Keras 2.3.1

square error function are generally used. Due to the convenience of calculating, the mean square error function is chosen as the loss function in the fusion feature model, and the calculation formula is as follows:

$$L(y, \hat{y}) = \frac{1}{n} \sum_{i=1}^{n} (y_i - \hat{y}_i)^2, \quad (11)$$

where y_i is the actual value of the ith sample, $\hat{y}_i$ is the predicted value of the ith sample, and n is the number of samples.

5.3.2. Setting the Number of Neurons in the Model. Before the model is trained, the number of neurons in the input and hidden layers of the model should be set (the model in the paper is based on a sequence of historical traffic flows to predict the traffic flow value at the next moment, so the number of neurons in the output layer of the model is set to 1; refer to Section 4 for details). The following is the process of setting the number of neurons in the input and hidden layers.

To obtain the appropriate number of neurons for the input layer, we select 6, 12, 18, and 24 as the number of neurons for the input layer to train the model and obtain the optimal number of neurons for the input layer by error analysis of the test set. Similarly, for the setting of the number of neurons in the BiGRU layer, four neuron numbers of 16, 32, 64, and 128 are chosen to train the model. The optimal number of neurons in each input and hidden layer of the neural network is determined by error analysis of the test set. Meanwhile, in the 1DCNN layer, the convolutional operation to extract features is implemented through convolutional kernels, and the size of kernel is set 2 ∗ 1, i.e. filters = 64 and kernel_size = 2. The ReLu function was chosen as the activation function for the convolutional layer. It is calculated as follows:

$$\text{Re}lu(x) = \begin{cases} x, & x > 0, \\ 0, & x \leq 0, \end{cases} \tag{12}$$

where x is the input to the activation function.

In the Dropout layer, the neuron loss rate is set to 20%. In addition, epoch is set to 300 rounds, and the batch size is set to 256.

For the error analysis of the test set, MAPE is selected as the main evaluation metric, and MAE and RMSE are selected as auxiliary evaluation metrics. The results of the evaluation metrics for the test set with different numbers of neurons in the model input layer and the bidirectional GRU network, including MAPE, MAE, and RMSE, were obtained, as shown in Table 2.

From Table 2, it can be found that the model has the strongest generalization ability when the number of neurons in the input layer is 12, and the number of neurons in the BiGRU network is 128, so we choose 12 and 128 as the numbers of neurons in the input layer of the model and the BiGRU network.

5.3.3. Optimization Algorithm Setup.

In the training process of deep learning models, optimization algorithms are used to iteratively optimize the parameters generated in the training model in order to reduce the value of the loss function, so that the training process of the model becomes stable as the number of iterations increases. The mainstream optimization algorithms include RMSProp and Adam, both of which are applied to train the fused feature model, and the optimization algorithm is selected based on the generalization capability of the model as an indicator. The RMSProp algorithm and Adam algorithm are used to train the fusion feature model, respectively, and the results of the three evaluation metrics are obtained, as shown in Table 3.

As can be seen in Table 3, MAPE, MAE, and RMSE are all smaller than the RMSProp algorithm when the CDLP model is trained using the Adam algorithm. The results indicate that the Adam algorithm is more efficient than the RMSProp algorithm and is selected as the optimization algorithm for multifeature fusion model.

5.4. Results and Analysis.

After determining the parameters of the model, the designed training and test sets are used to

TABLE 2: Evaluation indicator results under different neuron numbers of BiGRU network.

Neuron number in BiGRU network	MAPE (%)	MAE	RMSE
(6,16)	7.69	12.80	17.80
(6,32)	6.27	11.30	15.86
(6,64)	5.98	11.01	15.59
(6,128)	5.89	11.03	15.59
(12,16)	7.61	12.85	17.73
(12,32)	6.19	11.08	15.64
(12,64)	5.53	10.79	15.36
(12,128)	5.52	10.46	14.82
(18,16)	6.43	11.40	15.83
(18,32)	6.34	11.33	17.99
(18,64)	6.04	10.81	15.26
(18,128)	5.87	10.59	15.08
(24,16)	8.01	12.39	16.82
(24,32)	6.22	11.33	15.89
(24,64)	5.52	10.41	14.96
(24,128)	5.85	10.56	14.96

TABLE 3: Comparison of three evaluation indicators between RMSProp and Adam.

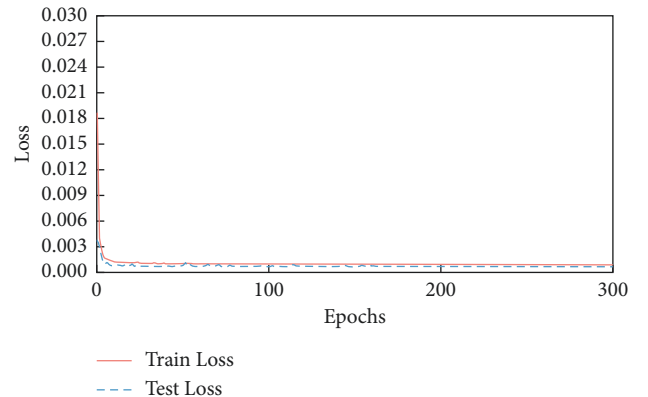| Optimization algorithm	MAPE (%)	MAE	RMSE
Adam	5.52	10.46	14.82
RMSProp	6.37	11.63	16.22

FIGURE 4: Loss function curves for the training and test sets of multifeature fusion model.

validate the predictive performance of the multifeature fusion model. The loss function curves generated by the model during the training process are shown in Figure 4. From Figure 4, it can be found that as the epoch increases, the loss function curves of the training and test sets decrease rapidly and steadily and finally converge to a constant 0, indicating that the design of the multifeature fusion model is reasonable.

Figure 5 shows the prediction results of the multifeature fusion model in the test set. It can be found that the multifeature fusion model can fit the actual traffic flow in the test set very well; specifically, the absolute error of the model at each moment is found to be between [−60,60] from the error curve graph.

In addition, to further verify the robustness of the multifeature fusion model, Figure 6 shows the MAPE plot of
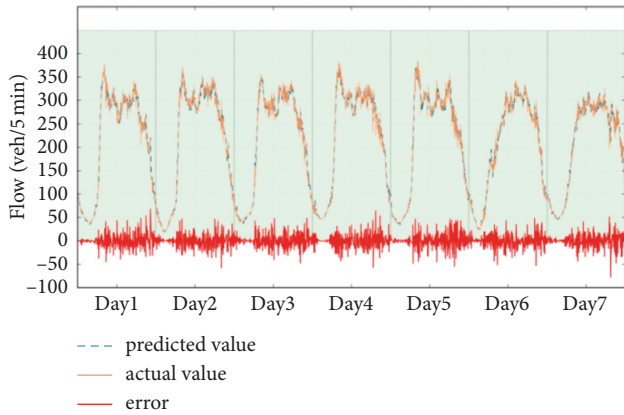
FIGURE 5: Prediction results from multi-feature fusion model under test set.
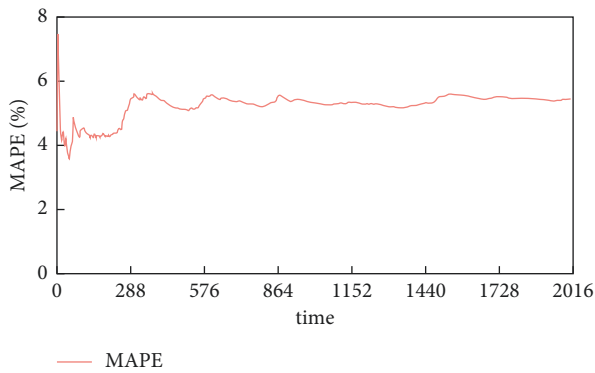

FIGURE 6: MAPE in the test set.

the model in the test set. As can be seen from the graph, the trend of the MAPE curve gradually decreases from the maximum value to the inflection point and then slowly increases and gradually converges to 5.52%, which indicates that the fused feature model has good robustness and low error, further indicating that the multifeature fusion model can better achieve traffic flow prediction.

To further validate the feasibility of the multifeature fusion model, the ability of the multifeature fusion model in extracting long-term dependent features and local features of the traffic is first observed. The Conv-BiGRU module (includes other modules) is selected as the comparison model. The structure of the module consists of a parallel layer of a convolutional layer and a BiGRU network, and the function of the module is to extract local trend features and long-term dependent features of the traffic flow individually. The model finally fuses the long-term dependent features, local trend features, and periodicity (including daily and weekly periodicity) of the traffic flow through the feature fusion layer and then predicts them. Second, the impact of periodic features on the multifeature fusion model is verified. Short-term traffic flows usually exhibit strong periodicity, and the advantage of the model is that it takes into account the periodicity of traffic flows by using two BiGRU-attention modules to extract the daily and weekly periodicity of traffic flows, respectively. The model containing only one

module of CNN-BiGRU-attention is used as a comparison model for validation. Third, the periodicity usually includes daily and weekly periodicity, and the models considering only daily and weekly periodicity, respectively, are used as comparison models for validation. Fourthly, a model that does not contain attention mechanisms in each module is considered as a comparison model for validation. Based on these comparison models and the multifeature fusion models mentioned above, the corresponding MAPE results were obtained by training and testing, as shown in Figure 7.

From Figure 7, it can be found that the maximum, minimum, and median values of the multifeature fusion model containing the CNN-BiGRU module are smaller than those containing the Conv-BiGRU module, indicating that the feature extraction capability of the CNN-BiGRU module is better than that of the Conv-BiGRU module. This is because the local trend features and long-term dependent features of the traffic flow are intertwined and interact with each other. Furthermore, the maximum, minimum, and median values of the multifeature fusion model are smaller than those of the CNN-BiGRU-attention model with only one module, because the periodic features play an important role in the prediction of traffic flow in the short-term traffic flow. In addition, from Figure 7, it also can be found that the MAPE of the multifeature fusion model is smaller than that of the feature fusion model without the attention mechanism. This indicates that the attention mechanism in multifeature fusion model improves the prediction accuracy by focusing on the important features extracted from each module.

Finally, the proposed multifeature fusion model is compared with existing baseline models. The baseline models include the LSTM model, GRU model, CNN-LSTM-attention model, CNN-GRU-attention model, and CDLP model [72]. The LSTM model and GRU model are composed of one input layer, two hidden layers (LSTM layer and GRU layer), and one output layer. The CNN-LSTM-attention model is composed of an input layer, a hidden layer, and an output layer, where the hidden layer is composed of a convolutional layer, two LSTM layers, and an attention mechanism layer connected sequentially, and the structure of the CNN-GRU-attention model is the same as that of the CNN-LSTM-attention. The parameters of the five benchmark models are set as in the multifeature fusion model.

The prediction errors in terms of prediction performance metrics for the different models are shown in Table 4, from which it can be found that the multifeature fusion model has the lowest prediction error. This is because the LSTM and GRU models mainly consider the temporal characteristics of traffic flow, that is, the long-short time dependence, while the CNN-GRU-attention model and the CNN-LSTM-attention model mainly consider the spatial and temporal characteristics of traffic flow, which is better than the LSTM and GRU models in terms of prediction error. The prediction performance of the CNN-GRU-attention model and the CNN-LSTM-attention model is better than that of the LSTM and GRU models, because the CNN-GRU-attention model and the CNN-LSTM-attention model mainly consider the spatial and temporal characteristics of traffic flow and consider the
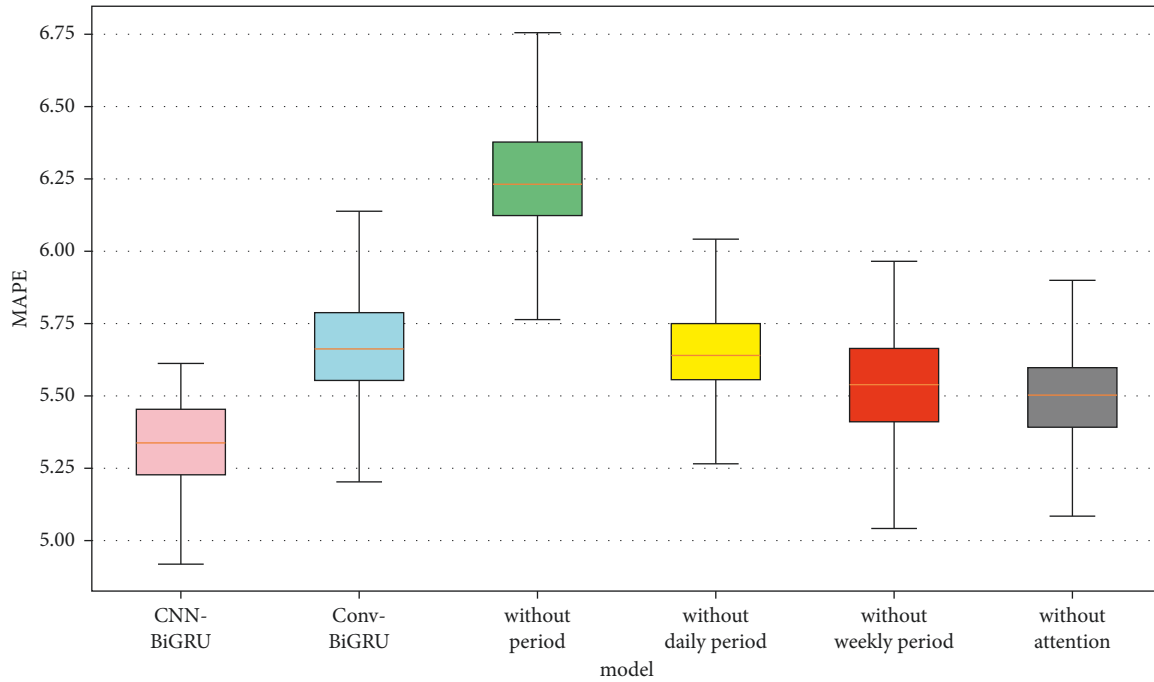
Figure 7: Comparison of MAPE results for different models.

Table 4: Comparison of three evaluation indexes of the multifeature fusion model and baseline models.

Model	MAPE (%)	RMSE	MAE
LSTM	13.32	20.35	15.26
GRU	12.27	19.70	14.73
CNN-LSTM-attention	6.34	15.32	10.93
CNN-GRU-attention	5.71	15.53	10.81
CDLP	5.62	15.17	10.63
Multi-feature fusion model	5.52	14.82	10.46

Table 5: Comparison of model training time of the multifeature fusion model and baseline models.

Model	Training time (min)
LSTM	32
GRU	38
CNN-LSTM-attention	47
CNN-GRU-attention	45
CDLP	47
Multi-feature fusion model	45

spatial characteristics of the model more than that of the LSTM and GRU models. The CDLP model is a combined prediction model based on the CNN-LSTM-attention model and the CNN-GRU-attention model, which also considers only the spatiotemporal characteristics of the traffic flow. The multifeature fusion model extracts the spatiotemporal, weekly, and daily characteristics of the traffic flow by using three different modules of the combined deep learning method, so the prediction performance of the multifeature fusion model is better than that of the baseline model.

In addition, the training time of the different models are shown in Table 5. It can be found that the training time of the multifeature fusion model is the same as that of the CNN-GRU-attention model in the combined model with higher prediction accuracy, but the MAPE, RMSE, and MAE of the model are reduced by 0.19%, 0.71, and 0.35, respectively, which are better than those of the CNN-GRU-attention model. Furthermore, the training time of the multifeature fusion model is smaller than that of the CNN-LSTM-attention model and the CDLP model, while the prediction accuracy is improved in both cases, which can be reflected in Table 4. This is because the model uses the CNN-BiGRU-attention module, in which GRU is a simplification of the LSTM, so the training time for the multifeature fusion model is less than that of the CNN-LSTM-attention model and the CDLP model (which uses the CNN-LSTM-attention module). Therefore, the multifeature fusion model has superior prediction performance.

6. Conclusion and Future Work

Short-term traffic flow prediction is one of the core components in intelligent transportation systems. In order to solve the problem of not extracting multiple features of traffic flow in traffic flow prediction, in this paper, a multifeature fusion model consisting of a CNN-BiGRU module with an attention mechanism and two BiGRU modules with an attention mechanism is proposed. Moreover, the parameters in the multifeature fusion model including the number of neurons, the optimization algorithm, and other parameters are obtained by experimental calibration.

Through experiments, it is found that the CNN-BiGRU-attention module can effectively capture the local trend features and long-term dependent features of the traffic flow, and the two BiGRU-attention modules can effectively capture the daily and weekly cycle features of the traffic flow.

At the same time, the attention mechanism improves the prediction accuracy of the model by focusing on the importance of the features acquired in each module, and the feature fusion layer of the model allows the features extracted from each module to be fused to predict future traffic flow trends.

Finally, extensive experimental results have shown that the predictive performance of the multifeature fusion model is superior to that of the baseline models for the same dataset.

In this work, we investigate traffic flow prediction using only cross-sectional traffic flows as the object of study. However, in real life, road network traffic flows usually exhibit extremely complex characteristics, and it is difficult for traditional CNN and BiGRU networks to fetch short-time traffic flow features under complex road networks. Therefore, similar graph neural network examples, such as spatiotemporal synchronous graph convolutional neural networks [76], provide a solution to the problem of short-term traffic flow prediction in complex and large road networks, which is difficult to be solved by traditional combined CNN-GRU models; therefore, it will be reserved for our future work and offers a new alternative approach for traffic prediction. In addition, the prediction of short-term traffic flows is often influenced by weather, traffic accidents, and major events, so the study of short-term traffic flow prediction considering special events will be left as another study for our future research.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

References

[1] P. Ersoy and G. Börühan, "Intelligent transportation systems and their applications in road transportation industry in Turkey," in *Proceedings of the 12th International Conference on Logistics & Sustainable Transport*, pp. 11–13, Celje, Slovenia, June 2015.

[2] N. Cui, B. Chen, K. Zhang, Y. Zhang, X. Liu, and J. Zhou, "Effects of route guidance strategies on traffic emissions in intelligent transportation systems," *Physica A: Statistical Mechanics and Its Applications*, vol. 513, pp. 32–44, 2019.

[3] I. O. Olayode, L. K. Tartibu, M. O. Okwu, and U. F. Uchechi, "Intelligent transportation systems, un-signalized road intersections and traffic congestion in Johannesburg: a systematic review," *Procedia CIRP*, vol. 91, pp. 844–850, 2020.

[4] A. Richter, M.-O. Löwner, R. Ebendt, and M. Scholz, "Towards an integrated urban development considering novel intelligent transportation systems," *Technological Forecasting and Social Change*, vol. 155, Article ID 119970, 2020.

[5] J. Zhang, F.-Y. Wang, K. Wang, W.-H. Lin, X. Xu, and C. Chen, "Data-driven intelligent transportation systems: a survey," *IEEE Transactions on Intelligent Transportation Systems*, vol. 12, no. 4, pp. 1624–1639, Dec. 2011.

[6] M. M. Hamed, H. R. Al-Masaeid, and Z. M. B. Said, "Short-term prediction of traffic volume in urban arterials," *Journal of Transportation Engineering*, vol. 121, no. 3, pp. 249–254, May 1995, 10.1061/(ASCE)0733-947X.

[7] S. V. Kumar and L. Vanajakshi, "Short-term traffic flow prediction using seasonal ARIMA model with limited input data," *Eur. Transp. Res. Rev.,* vol. 7, no. 3, p. 21, 2015.

[8] L. Zhang, Q. Liu, W. Yang, N. Wei, and D. Dong, "An improved k-nearest neighbor model for short-term traffic flow prediction," *Procedia - Social and Behavioral Sciences*, vol. 96, pp. 653–662, 2013.

[9] R. E. Kalman, "A new approach to linear filtering and prediction problems," *Journal of Basic Engineering*, vol. 82, no. 1, pp. 35–45, 1960.

[10] T. D. Toan and V.-H. Truong, "Support vector machine for short-term traffic flow prediction and improvement of its model training using nearest neighbor approach," *Transportation Research Record: Journal of the Transportation Research Board*, vol. 2675, no. 4, pp. 362–373, 2021.

[11] M. G. Karlaftis and E. I. Vlahogianni, "Statistical methods versus neural networks in transportation research: differences, similarities and some insights," *Transportation Research Part C: Emerging Technologies*, vol. 19, no. 3, pp. 387–399, 2011.

[12] D. Dehuai Zeng, J. Jianmin Xu, J. Jianwei Gu, L. Liyan Liu, and G. Gang Xu, "Short term traffic flow prediction using hybrid ARIMA and ANN models," in *Proceedings of the 2008 Workshop on Power Electronics and Intelligent Transportation System*, pp. 621–625, Guangzhou, China, August 2008.

[13] Z. Liu, J. Guo, J. Cao, Y. Wei, and W. Huang, "A hybrid short-term traffic flow forecasting method based on neural networks combined with k-nearest neighbor," *Promet - Traffic & Transportation*, vol. 30, no. 4, pp. 445–456, 2018.

[14] J. Wang, W. Deng, and Y. Guo, "New bayesian combination method for short-term traffic flow forecasting," *Transportation Research Part C: Emerging Technologies*, vol. 43, pp. 79–94, Jun. 2014.

[15] H. Wang, W. Quan, and W. Y. Ochieng, "Smart road stud based two-lane traffic surveillance," *Journal of Intelligent Transportation Systems*, vol. 24, no. 5, pp. 480–493, 2020.

[16] S. F. Wong, H. C. Mak, C. H. Ku, and W. I. Ho, "Developing advanced traffic violation detection system with RFID technology for smart city," in *Proceedings of the 2017 Ieee International Conference on Industrial Engineering and Engineering Management (ieem)*, pp. 334–338, Singapore, December 2017.

[17] L. A. Klein, M. R. Kelley, and M. K. Mills, "Evaluation of traffic detection technologies for IVHS," *SPIE Proceedings*, vol. 2344, 1995.

[18] J. Versavel and B. Boucke, "Video for traffic data and incident detection by traficon," 1998, http://trid.trb.org/view.aspx?id=507034.

[19] D. F. Llorca, M. A. Sotelo, S. Sánchez, M. Ocaña, J. M. Rodríguez-Ascariz, and M. A. García-Garrido, "Traffic data collection for floating car data enhancement in V2I networks," *EURASIP Journal on Applied Signal Processing*, vol. 2010, no. 1, Article ID 719294, pp. 1-13, Dec. 2010.

[20] L. Theis, W. Shi, A. Cunningham, and F. Huszár, "Lossy Image Compression with Compressive Autoencoders," Mar. 2017, http://arxiv.org/abs/1703.00395.

[21] L. Gao, Z. Guo, H. Zhang, X. Xu, and H. T. Shen, "Video captioning with attention-based LSTM and semantic consistency," *IEEE Transactions on Multimedia*, vol. 19, no. 9, pp. 2045–2055, 2017.

[22] A. Graves and N. Jaitly, "Towards end-to-end speech recognition with recurrent neural networks," in *Proceedings of the International conference on machine learning*, pp. 1764–1772, Beijing, China, 2014.

[23] Y. Hou, Z. Deng, and H. Cui, "Short-term traffic flow prediction with weather conditions: based on deep learning algorithms and data fusion," *Complexity*, vol. 2021, pp. 1–14, Article ID 6662959, 2021.

[24] Y. Wu, H. Tan, L. Qin, B. Ran, and Z. Jiang, "A hybrid deep learning based traffic flow prediction method and its understanding," *Transportation Research Part C: Emerging Technologies*, vol. 90, pp. 166–180, May 2018.

[25] C. Chen, Z. Liu, S. Wan, J. Luan, and Q. Pei, "Traffic flow prediction based on deep learning in internet of vehicles," *IEEE Transactions on Intelligent Transportation Systems*, vol. 22, no. 6, pp. 3776–3789, 2021.

[26] L. Han and Y. S. Huang, "Short-term traffic flow prediction of road network based on deep learning," *IET Intelligent Transport Systems*, vol. 14, no. 6, pp. 495–503, 2020.

[27] M. Levin and Y.-D. Tsao, "On forecasting freeway occupancies and volumes," *Transportation Research Record*, vol. 773, pp. 47–49, Jan. 1980.

[28] Y. Zhang, Y. Zhang, and A. Haghani, "A hybrid short-term traffic flow forecasting method based on spectral analysis and statistical volatility model," *Transportation Research Part C: Emerging Technologies*, vol. 43, pp. 65–78, Jun. 2014.

[29] M. Van Der Voort, M. Dougherty, and S. Watson, "Combining kohonen maps with arima time series models to forecast traffic flow," *Transportation Research Part C: Emerging Technologies*, vol. 4, no. 5, pp. 307–318, Oct. 1996.

[30] B. M. Williams and L. A. Hoel, "Modeling and forecasting vehicular traffic flow as a seasonal ARIMA process: theoretical basis and empirical results," *Journal of Transportation Engineering*, vol. 129, no. 6, pp. 664–672, 2003.

[31] P. Duan, G. Mao, W. Liang, and D. Zhang, "A unified spatio-temporal model for short-term traffic flow prediction," *IEEE Transactions on Intelligent Transportation Systems*, vol. 20, no. 9, pp. 3212–3223, 2019.

[32] U. Ryu, J. Wang, T. Kim, S. Kwak, and J. U, "Construction of traffic state vector using mutual information for short-term traffic flow prediction," *Transportation Research Part C: Emerging Technologies*, vol. 96, pp. 55–71, 2018.

[33] Z. Yan and S. Lv, "Short-term forecast model of taxi demand based on time and space heterogeneity," *Journal of Intelligent and Fuzzy Systems*, vol. 41, no. 2, pp. 4175–4186, 2021.

[34] I. Okutani and Y. J. Stephanedes, "Dynamic prediction of traffic volume through Kalman filtering theory," *Transportation Research Part B: Methodological*, vol. 18, no. 1, pp. 1–11, 1984.

[35] J. Guo and B. M. Williams, "Real-time short-term traffic speed level forecasting and uncertainty quantification using layered kalman filters," *Transportation Research Record: Journal of the Transportation Research Board*, vol. 2175, no. 1, pp. 28–37, Jan. 2010.

[36] W. Hu, L. Yan, K. Liu, and H. Wang, "A Short-term traffic flow forecasting method based on the hybrid PSO-SVR," *Neural Processing Letters*, vol. 43, no. 1, pp. 155–172, 2016.

[37] C.-C. Lu and X. Zhou, "Short-term highway traffic state prediction using structural state space models," *Journal of Intelligent Transportation Systems*, vol. 18, no. 3, pp. 309–322, 2014.

[38] H. Jiang, Y. Zou, S. Zhang, J. Tang, and Y. Wang, "Short-term speed prediction using remote microwave sensor data: machine learning versus statistical model," *Mathematical Problems in Engineering*, vol. 2016, pp. 1–13, Article ID 9236156, 2016.

[39] J. Wang and Q. Shi, "Short-term traffic speed forecasting hybrid model based on chaos-wavelet analysis-support vector machine theory," *Transportation Research Part C: Emerging Technologies*, vol. 27, pp. 219–232, Feb. 2013.

[40] X. Feng, X. Ling, H. Zheng, Z. Chen, and Y. Xu, "Adaptive multi-kernel SVM with spatial-temporal correlation for short-term traffic flow prediction," *IEEE Transactions on Intelligent Transportation Systems*, vol. 20, no. 6, pp. 2001–2013, 2019.

[41] X. Wang, N. Zhang, Y. Zhang, and Z. Shi, "Forecasting of short-term metro ridership with support vector machine online model," *Journal of Advanced Transportation*, vol. 2018, pp. 1–13, Article ID 3189238, 2018.

[42] D. Park and L. R. Rilett, "Forecasting multiple-period freeway link travel times using modular neural networks," *Transportation Research Record: Journal of the Transportation Research Board*, vol. 1617, no. 1, pp. 163–170, Jan. 1998.

[43] G. Xiaojian and Z. Quan, "A traffic flow forecasting model based on BP neural network," in *Proceedings of the 2009 2nd International Conference on Power Electronics and Intelligent Transportation System (PEITS)*, vol. 3, pp. 311–314, Shenzhen, China, December 2009.

[44] C. Yanchong, H. Darong, and Z. Ling, "A short-term traffic flow prediction method based on wavelet analysis and neural network," in *Proceedings of the 28th Chinese Control and Decision Conference (2016 Ccdc)*, Yinchuan, China, May 2016, https://www.webofscience.com/wos/alldb/full-record/WOS:000383222307055.

[45] S. Li, Z. Zhang, Y. Liu, and Z. Qin, "A short-term traffic flow reliability prediction method considering traffic safety," *Mathematical Problems in Engineering*, vol. 2020, pp. 1–9, Article ID 6682216, 2020.

[46] E. I. Vlahogianni, M. G. Karlaftis, and J. C. Golias, "Short-term traffic forecasting: where we are and where we're going," *Transportation Research Part C: Emerging Technologies*, vol. 43, pp. 3–19, Jun. 2014.

[47] L. Qu, W. Li, W. Li, D. Ma, and Y. Wang, "Daily long-term traffic flow forecasting based on a deep neural network," *Expert Systems with Applications*, vol. 121, pp. 304–312, May 2019.

[48] J. Yao and Y. Ye, "The effect of image recognition traffic prediction method under deep learning and naive Bayes algorithm on freeway traffic safety," *Image and Vision Computing*, vol. 103, Article ID 103971, 2020.

[49] W. Huang, G. Song, H. Hong, and K. Xie, "Deep architecture for traffic flow prediction: deep belief networks with multitask learning," *IEEE Transactions on Intelligent Transportation Systems*, vol. 15, no. 5, pp. 2191–2201, Oct. 2014.

[50] Y. Lv, Y. Duan, W. Kang, Z. Li, and F.-Y. Wang, "Traffic flow prediction with big data: a deep learning approach," *IEEE Transactions on Intelligent Transportation Systems*, vol. 16, no. 2, pp. 1–9, 2014.

[51] M. Sameen and B. Pradhan, "Severity prediction of traffic accidents with recurrent neural networks," *Applied Sciences*, vol. 7, no. 6, p. 476, 2017.

[52] 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," *Transportation Research Part C: Emerging Technologies*, vol. 54, pp. 187–197, May 2015.

[53] Z. Zhao, W. Chen, X. Wu, P. C. Y. Chen, and J. Liu, "LSTM network: a deep learning approach for short-term traffic

forecast," *IET Intelligent Transport Systems*, vol. 11, no. 2, pp. 68–75, 2017.

[54] J. Wang, R. Chen, and Z. He, "Traffic speed prediction for urban transportation network: a path based deep learning approach," *Transportation Research Part C: Emerging Technologies*, vol. 100, pp. 372–385, 2019.

[55] Z. Cui, R. Ke, Z. Pu, and Y. Wang, "Stacked Bidirectional and Unidirectional LSTM Recurrent Neural Network for Forecasting Network-wide Traffic State with Missing Values," May 2020, http://arxiv.org/abs/2005.11627.

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

[57] K. Zhang, L. Wu, Z. Zhu, and J. Deng, "A multitask learning model for traffic flow and speed forecasting," *IEEE Access*, vol. 8, Article ID 80707, 2020.

[58] 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, 2017.

[59] W. Zhang, Y. Yu, Y. Qi, F. Shu, and Y. Wang, "Short-term traffic flow prediction based on spatio-temporal analysis and CNN deep learning," *Transportmetrica: Transportation Science*, vol. 15, no. 2, pp. 1688–1711, 2019.

[60] J. An, L. Fu, M. Hu, W. Chen, and J. Zhan, "A novel fuzzy-based convolutional neural network method to traffic flow prediction with uncertain traffic accident information," *IEEE Access*, vol. 7, Article ID 20708, 2019.

[61] R. Tian, J. Bi, Q. Zhang, and Y. Liu, "Research on lane occupancy rate forecasting based on the capsule network," *IEEE Access*, vol. 8, Article ID 38776, 2020.

[62] K. Li, C. Zhai, and J. Xu, "Short-term traffic flow prediction using a methodology based on ARIMA and RBF-ANN," in *Proceedings of the 2017 Chinese Automation Congress (Cac)*, pp. 2804–2807, Jinan, China, October 2017, https://www.webofscience.com/wos/alldb/fullrecord/WOS:000427816102167.

[63] R. Yao, W. Zhang, and L. Zhang, "Hybrid methods for short-term traffic flow prediction based on ARIMA-GARCH model and wavelet neural network," *Journal of Transportation Engineering, Part A: Systems*, vol. 146, no. 8, Article ID 04020086, Aug. 2020.

[64] W. Li, S. Chen, X. Wang, C. Yin, and Z. Huang, "A hybrid approach for short-term traffic flow forecasting based on similarity identification," *Modern Physics Letters B*, vol. 35, no. 13, Article ID 2150212, May 2021.

[65] K. Y. Chan, T. S. Dillon, J. Singh, and E. Chang, "Neural-network-based models for short-term traffic flow forecasting using a hybrid exponential smoothing and levenberg-marquardt algorithm," *IEEE Transactions on Intelligent Transportation Systems*, vol. 13, no. 2, pp. 644–654, Jun. 2012.

[66] Y. Gu, W. Lu, L. Qin, M. Li, and Z. Shao, "Short-term prediction of lane-level traffic speeds: a fusion deep learning model," *Transportation Research Part C: Emerging Technologies*, vol. 106, pp. 1–16, Sep. 2019.

[67] 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 Transactions on Intelligent Transportation Systems*, vol. 22, no. 5, pp. 2627–2636, May 2021.

[68] D. Bahdanau, K. Cho, and Y. Bengio, "Neural Machine Translation by Jointly Learning to Align and Translate," May 2016, http://arxiv.org/abs/1409.0473.

[69] Q. Liu, B. Wang, and Y. Zhu, "Short-term traffic speed forecasting based on attention convolutional neural network for arterials," *Computer-Aided Civil and Infrastructure Engineering*, vol. 33, no. 11, pp. 999–1016, 2018.

[70] P. Wu, Z. Huang, Y. Pian, L. Xu, J. Li, and K. Chen, "A combined deep learning method with attention-based LSTM model for short-term traffic speed forecasting," *Journal of Advanced Transportation*, vol. 2020, pp. 1–15, Article ID 8863724, 2020.

[71] D. Ma, B. Sheng, X. Ma, and S. Jin, "Fuzzy hybrid framework with dynamic weights for short-term traffic flow prediction by mining spatio-temporal correlations," *IET Intelligent Transport Systems*, vol. 14, no. 2, pp. 73–81, 2020.

[72] C. Ren, C. Chai, C. Yin et al., "Short-term traffic flow prediction: a method of combined deep learnings," *Journal of Advanced Transportation*, vol. 2021, pp. 1–15, Article ID 9928073, 2021.

[73] J. Gu, Z. Wang, J. Kuen et al., "Recent advances in convolutional neural networks," *Pattern Recognition*, vol. 77, pp. 354–377, 2018.

[74] K. Cho, B. V Merrienboer, C Gulcehre et al., "Learning Phrase Representations Using RNN Encoder-Decoder for Statistical Machine Translation," Sep. 2014, http://arxiv.org/abs/1406.1078.

[75] D. Chen, X. Yan, X. Liu, S. Li, L. Wang, and X. Tian, "A multiscale-grid-based stacked bidirectional GRU neural network model for predicting traffic speeds of urban expressways," *IEEE Access*, vol. 9, pp. 1321–1337, 2021.

[76] C. Song, Y. Lin, S. Guo, and H. Wan, "Spatial-temporal synchronous graph convolutional networks: a new framework for spatial-temporal network data forecasting," *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 34, no. 01, pp. 914–921, 2020.