Short-term Traffic Flow Prediction Based on Spatiotemporal and Periodic Feature Fusion

Qingrong Wang, Xiaohong Chen, Changfeng Zhu, Kai Zhang, Runtian He, Jinhao Fang

***Abstract*—Short-term traffic volume prediction is crux for alleviating traffic gridlock. Considering the insufficiently extracted spatiotemporal and periodic characteristics of traffic stream in existing traffic volume forecast studies, this research presents a short-term traffic volume forecast model (WRNCL-TCL) that considers spatiotemporal and periodic characteristics. Firstly, the wavelet threshold is used to denoise the initial traffic stream data. Secondly, the CNN-LSTM model is employed to capture the spatiotemporal features of the traffic stream. Considering the degradation problem that may be caused by the model with the increase of network depth, add residual neural units based on CNN. We employ the TCN-LSTM model to acquire the periodic characteristics of the traffic stream. Finally, we combine the extracted spatiotemporal and periodic features, and utilize the fully connected layer to obtain the ultimate forecasted results. WRNCL-TCL is applied to real data sets in two different scenarios to validate the forecasting capability of the suggested model. Compared to the benchmark model and the ablation experiment, the consequences suggest that the proposed model exhibits favorable predictive capabilities and can serve as a theoretical foundation for traffic control.**

***Index Terms*—Spatiotemporal characteristics, Periodic characteristics, Short-term traffic volume, Wavelet thresholds**

I. Introduction

With the acceleration of urbanization and process, the number of cars in China continues to increase, and traffic congestion has become one of the focus problems faced by modern society. The advent of the Intelligent Transportation System (ITS) has efficaciously mitigated the issue of traffic congestion, and the short-term traffic volume forecast plays a vital charactar in ITS. Hence, the investigation of theoretical aspects in short-term traffic volume prediction has very significant theoretical value for further promoting the implementation of ITS and alleviating traffic congestion [1].

Earlier, [2] used ACF and PACF to analyze traffic time series and introduced a traffic congestion prediction model based on ARIMA. [3] introduced a recurrent neural network architecture that enables simultaneous prediction of road conditions across multiple segments. By perceiving the interconnection among multiple segments, they can grasp their reciprocal impact and consequently enhance the precision of the prediction model. [4] introduced the denoising scheme into traffic volume forecast and proposed a forecasting model combining the scheme and support vector machine (SVM).

While the preceding techniques have conducted some investigations into short-term traffic volume forecast, they cannot effectively solve the sudden traffic conditions. It is also challenging to acquire the features of long-term memory of traffic stream sequences, which affects the model's prediction accuracy. As deep learning becomes increasingly mature in the field of transportation, it can address the limitations of the aforementioned methods. A Convolutional Neural Networks (CNN) model built upon deep learning was suggested in reference [5]. In reference [6], a method for predicting short-term urban traffic volume was presented, which utilized a data clustering technique with a convolutional neural network (DGCNN). [7] proposed an attentive differential convolutional neural network (ADCNN) model for capturing the higher-order spatiotemporal correlation in traffic stream data. In reference [8], a F-CNN technique was introduced, which incorporated uncertain traffic accident data into a CNN framework. The approach utilized fuzzy methods to represent the attributes of traffic accidents. [9] presented a new variant of CNN training based on balanced optimization of the metaheuristic algorithm, the EO-CNN model.

The aforementioned studies effectively addressed the unexpected traffic conditions, yet did not comprehensively account for the impact of the temporal attributes of traffic stream on the model's prediction. Therefore, [10] introduced a traffic volume prediction model built upon deep learning, employing LSTM. Due to its challenge in handling ultra-long dependencies, LSTM sometimes struggles. To overcome this limitation, in [11], the influential values of extended sequential time steps were connected to the current step, and an attention mechanism was integrated to capture these significant traffic stream values while reducing the impact of anomalous data, thus enhancing the prediction accuracy. [12] proposed a temporal information-enhanced LSTM model for predicting traffic stream on a single road section in response to existing studies that did not fully consider the effect of temporal characteristics on traffic volume forecast. [13] recommended a method for predicting

Manuscript received July 17, 2023; revised October 26, 2023.This work was supported in part by the National Natural Science Foundation of China (No. 71961016, 72161024), "Double-First Class" Major Research Programs, Educational Department of Gansu Province (No. GSSYLXM-04).

Qingrong Wang is a professor at School of Electronic and Information Engineering, Lanzhou Jiaotong University, Lanzhou 730070, China. (e-mail: 329046272@qq.com).

Xiaohong Chen is a postgraduate student at School of Electronic and Information Engineering, Lanzhou Jiaotong University, Lanzhou 730070, China. (Corresponding author to provide e-mail: 2937325132@qq.com).

Changfeng Zhu is a professor at School of Traffic and Transportation, Lanzhou Jiaotong University, Lanzhou 730070, China. (e-mail: cfzhu003@163.com).

Kai Zhang is a postgraduate student at School of Electronic and Information Engineering, Lanzhou Jiaotong University, Lanzhou 730070, China. (e-mail: 1274079870@qq.com).

Runtian He is a doctoral student at School of Traffic and Transportation, Lanzhou Jiaotong University, Lanzhou 730070, China. (e-mail: hrt25678@qq.com).

Jinhao Fang is a doctoral student at School of Traffic and Transportation, Lanzhou Jiaotong University, Lanzhou 730070, China. (e-mail: fangjin_hao@163.com).

short-term traffic stream employing LSTM and hierarchical clustering. Reference [14] dealt with diverse traffic situations and conducted empirical optimization of the construction and arguments of the blended LSTM neural network.

Short-term traffic volume prediction is highly susceptible to the influence of spatiotemporal attributes, but the above simplex models do not fully consider their spatial and temporal characteristics. Therefore, related scholars later proposed combined forecasting models. [15] proposed a short-term traffic volume forecast method combining an ARIMA and LSTM model. [16] recommended a CLM model founded on CNN and LSTM models, using CNN to obtain the daily and weekly cyclical characteristics of traffic speed in the target area. The LSTM layer extracts the spatiotemporal characteristics of the CNN output. [17] proposed an improved CNN and LSTM method for short-term traffic volume forecast to address the matter that existing models do not fully acquire the spatiotemporal attributes of the traffic stream. In literature [18], a traffic volume forecast technique that merges KNN and LSTM models was suggested. Reference [19] introduced a predictive model for short-term traffic stream, which integrated an attention mechanism and a 1DCNN-LSTM network.

The above built-up prediction model efficaciously acquire the spatiotemporal features of the traffic stream. However, the models may bring degradation problems with the increase of network depth. As a result, reference [20] suggested an algorithm for predicting urban short-term traffic stream built upon a CNN-ResNet-LSTM model. This approach utilizes CNN to acquire the brush-fire spatial attributes of the traffic stream. It incorporates multiple residual neural units to increase the depth of the network and enhance the prediction precision of the model. The temporal attributes of traffic stream data are captured using LSTM. In reference [21], a novel multi-scale ConvLSTM-Resnet network was suggested to acquire the local and global spatial pertinences of traffic stream data slices at every instant, which are treated as "traffic frames." The experimental findings illustrate the model's superior predictive performance.

While the preceding methods consider the spatiotemporal attributes of traffic stream data, they do not address the impact of noise in the raw data on traffic volume forecast. Therefore, In [22], a wavelet reconstruction-based convolutional neural network (WT-2DCNN) model was introduced to address the problem of yawp in the initial traffic stream data, which can result in substantial prediction errors. Reference [23] utilized wavelet transformation (WT) to break down the time series of traffic stream into various frequency ranges, aiming to alleviate the effect of noise on the traffic volume prediction. Reference [24] recommended a WBLA model based on wavelet transform and compared it with other models that do not consider noise. The results demonstrated that the WBLA model exhibits superior predictive proficiency. [25] recommended a combination of WT and kernel limit learning machine (KLM) for the instability and complexity of urban rail traffic changes.) and kernel extreme learning machine (KELM) hybrid prediction model W-KELM.

When accounting for the periodic characteristics of traffic stream data, if a single LSTM is applied to obtain its periodic characteristics, the prediction results of the model may have time lags due to the shortcomings of the LSTM itself. Therefore, some researchers have combined LSTM models with other neural networks and used them together to capture the spatiotemporal features or periodic properties of traffic stream. [26] recommended a prediction method combining temporal convolutional networks (TCN) and LSTM for the difficulty of capturing the data sequences' nonlinear properties, correlation, and cyclicality simultaneously by a single model. [27] used Savitzky-Golay to eliminate yawp from traffic stream data, TCN to adjust the temporal peculiarities of the data, and LSTM to acquire long-term dependencies in the time series. In reference [28], a proposed model for short-term load forecast incorporated an attention mechanism within a TCN-LSTM architecture.

In view of this, this paper addresses the strengths and weaknesses of the models proposed in the above study. It offers a combined forecasting model WRNCL-TCL that takes into account temporal and spatial characteristics and periodic properties, whose primary contributions are:

(1) To address the problem that the original traffic stream data can be contaminated by noise, wavelet threshold theory is implemented to decompose and reconstruct the original traffic stream data.

(2) In light of the spatio-temporal attributes of traffic stream, CNN and LSTM theories are introduced. The CNN model is utilized for obtain the spacial characteristics of the traffic flow and LSTM to capture the temporal dependency of the traffic flow.

(3) The model's performance may deteriorate as network layers increase. In this paper, multiple residual neural units are added to the CNN to deepen the network depth and enhance the precision of model prediction.

(4) To acquire the cyclical properties of traffic stream data, a TCN-LSTM model is developed by integrating the strengths of the TCN and LSTM models. This approach enables the full extraction of the daily and weekly cyclic patterns in traffic stream data.

II. Short-term Traffic Volume Forecast Model

Short-term traffic flow forecasting effectively uses historical traffic stream data to forecast traffic conditions in a certain period. The results can offer real time and effective traffic information to travelers, helping them to be able to make better route choices, thus reducing travel time and improving traffic congestion.

A. *Characteristics of the traffic stream*

Traffic stream data exhibits spatial and temporal attributes, where the present traffic stream is influenced by not only the preceding traffic volume but also the traffic volume across different road segments. Therefore, as shown in Equation (1), the spatio-temporal matrix of the traffic stream is constructed:

$$X_{t\times n}=\begin{bmatrix} x_1^1 & x_1^2 & \cdots & x_1^n \\ x_2^1 & x_2^2 & \cdots & x_2^n \\ \vdots & \vdots & \vdots & \vdots \\ x_t^1 & x_t^2 & \cdots & x_t^n \end{bmatrix} \tag{1}$$

Where: X_t^n is the traffic stream at the observation point n at the time t. Furthermore, the traffic stream data displays daily and weekly cyclic patterns. During weekdays, the traffic flow shows obvious peak commuting hours, and during non-working days, the traffic flow is relatively flat. As shown in Equation (2), the traffic flow matrix with daily periodicity is constructed as follows:

$$D=\begin{bmatrix} x_1^1 & x_1^2 & \cdots & x_1^n \\ x_2^1 & x_2^2 & \cdots & x_2^n \\ \vdots & \vdots & \vdots & \vdots \\ x_d^1 & x_d^2 & \cdots & x_d^n \end{bmatrix} \tag{2}$$

where: X_d^n denotes the traffic volume at the identical time t on the prior day at the observation point n. The traffic stream data from the preceding week and the current week demonstrate comparable patterns, reflecting the weekly cyclical nature of traffic volume. Therefore, the historical traffic volume matrix with weekly periodicity is constructed as shown in Equation (3):

$$W=\begin{bmatrix} x_1^1 & x_1^2 & \cdots & x_1^n \\ x_2^1 & x_2^2 & \cdots & x_2^n \\ \vdots & \vdots & \vdots & \vdots \\ x_w^1 & x_w^2 & \cdots & x_w^n \end{bmatrix} \tag{3}$$

where: X_w^n represents the traffic volume at the observation point n at the identical moment of the previous week.

B. Wavelet Threshold Denoising

The primordial traffic stream data is vulnerable to noise pollution, which can impact the model's forecasting performance. Therefore, this study addresses this issue by introducing the wavelet thresholding denoising theory and applying it to the traffic flow data through wavelet transform to generate information of lower and higher frequencies. Then, the low-frequency information is further decomposed based on the maximum decomposition scale of the traffic flow data until the maximum scale is reached. Since the wavelet coefficients obtained after decomposition contain essential information and noise, a threshold is set to retain the coefficients above the threshold as important information and set the coefficients below the threshold to zero as noise, thus achieving the denoising objective. The wavelet thresholding denoising and decomposition reconstruction process are illustrated in Fig. 1.

Firstly, the traffic stream data containing noise is represented by Equation (4):

$$N(k)=O(k)+\varepsilon e(k) \tag{4}$$

Where: $O(k)$ is the traffic flow data without noise contamination, ε represents the standard deviation of the noise coefficient, $e(k)$ represents the noise, and $N(k)$ is the signal contaminated by noise. From Fig. 1, the basic steps of wavelet threshold denoising are:

Step 1 Wavelet decomposition: determine the largest scale at which the traffic stream data needs to be decomposed and subsequently perform multi-level decomposition on the original data.

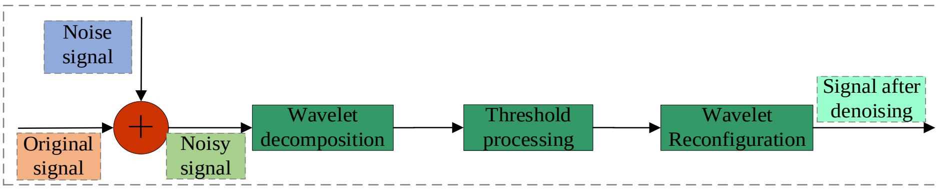
Step 2 Thresholding: establish an appropriate threshold and apply it to quantify the coefficients of each layer using a thresholding function.

Step 3 Wavelet reconstruction: The wavelet coefficients are reconstructed after thresholding to realize the denoising of the initial data.

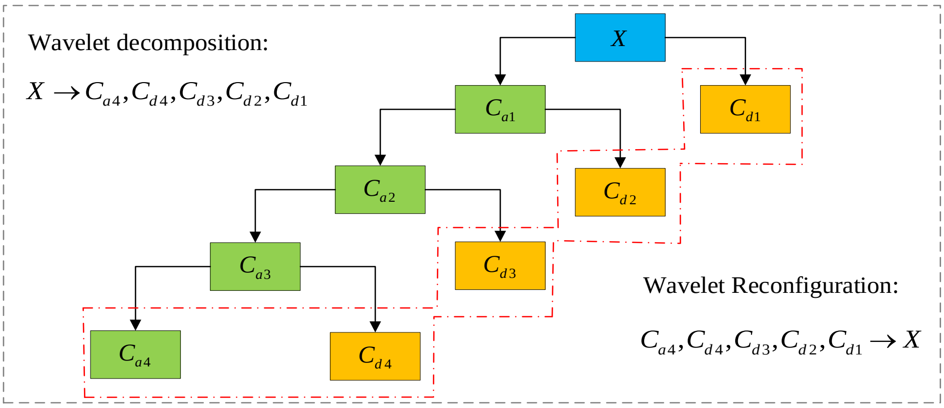
As shown in (b) in Fig. 1, this paper divides the raw traffic stream data into four layers and performs wavelet decomposition for each layer, the expressions of which are shown below:

$$WT(l,p)=\frac{1}{\sqrt{p}}\int_{-\infty}^{+\infty} O(t)*\psi\left(\frac{t-p}{l}\right)dt \tag{5}$$

Where: l is the decomposition scale, p is the translation, t represents the point-in-time, and $O(t)$ is the raw data.



(a) Wavelet threshold denoising process



(b) Wavelet Decomposition and Reconstruction

Fig. 1 Wavelet Threshold Denoising

Wavelet analysis produces information of lower and higher frequencies. The low-frequency information is commonly referred to as the "approximation component," while the high-frequency information is known as the "detail component." In Fig. 1(b), C_a represents the approximation component and C_d represents the detail component. This study employs a fixed threshold to process the decomposed signals. Finally, the approximate components of scale four and the detailed components of each scale are wavelet reconstructed.

C. Spatio-temporal feature extraction

1) Convolutional Neural Networks

Convolutional Neural Networks (CNN) is a more classical deep learning network framework with powerful feature extraction capability [29]. Given that the traffic conditions at nearby observation sites affects the traffic stream, the research utilizes CNN to obtain the spatial characteristics of traffic stream data. The CNN's network design is depicted in Fig. 2.

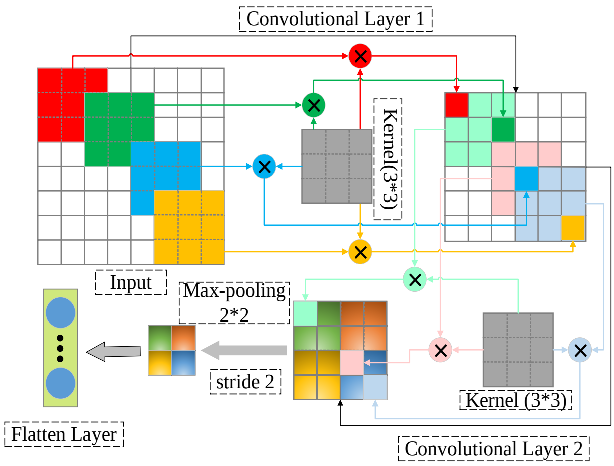


Fig. 2 Network structure of CNN

To adequately acquire the spatial attributes of the traffic stream data, this research transforms the traffic stream data into a traffic stream spatio-temporal matrix, which serves as the input for the CNN model and generates the feature matrix by the convolution operation. The convolution operation is shown in Equation (6):

$$c_i = f(conv(w_c x_i + b_c)) \tag{6}$$

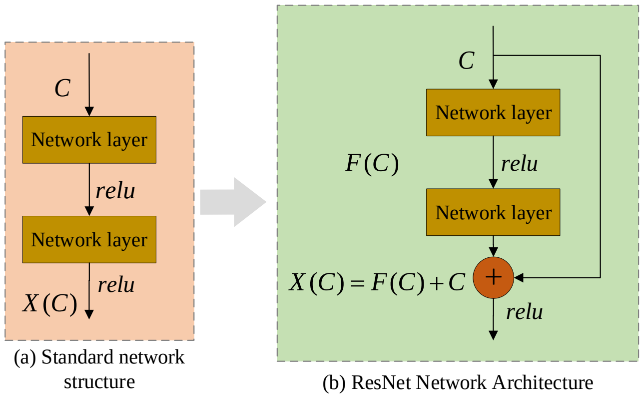
Where: f denotes the activation function, $conv$ represents the convolution operation, x_i and c_i denotes the input and output of the convolution layer, w_c and b_c is the weights and biases of the convolution layer. The research employs two convolutional layers to retain the spatial attributes of the traffic stream data, and only after the second convolutional operation is a pooling layer utilized.

$$C_i = pooling(c_i) \tag{7}$$

where $pooling$ denotes the pooling operation, c_i and C_i represent the input and output of this layer.

2) Residual neural unit

As the count of layers in the deep learning network grows, the model's accuracy will eventually reach a plateau or even decline, leading to increased training difficulty and model degradation. And the proposed residual neural network (ResNet) effectively solves the problem. The ResNet's architecture is depicted in Fig. 3(b).



(a) Standard network structure

(b) ResNet Network Architecture

Fig. 3 ResNet network structure diagram

To tackle the issue of model degradation in this study, a residual unit is incorporated into the convolutional layer, as illustrated in Equation (8).

$$X_i^{(l+1)} = F(C_i^{(l)}; \theta^{(l)}) + C_i^{(l)} \quad l = 1, 2, \cdots, M \tag{8}$$

where: $C_i^{(l)}$ and X_i^{l+1} are the inputs and outputs of the residual neural network, F are the residual functions, and $\theta^{(l)}$ are all learnable parameters of the residual units at the layer.

3) Long short-term memory

Sequential data is effectively processed through the utilization of Recurrent Neural Networks (RNNs). Nevertheless, RNNs are susceptible to gradient vanishing and exploding during the model training process, thus making them less effective in handling long sequence data. LSTM, a distinctive RNN structure, is an enhancement of RNN that primarily resolves the matters of vanishing and exploding gradients during training with long sequences. Hence, the study utilizes LSTM to acquire the temporal features of short-term traffic stream data that CNN disregards. The LSTM's design is displayed in Fig. 4.

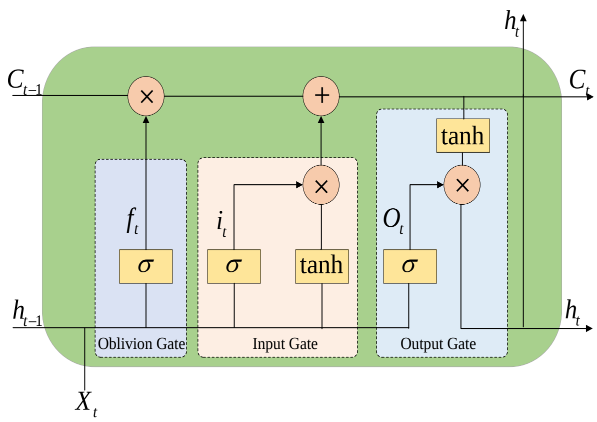


Fig. 4 LSTM structure diagram

Firstly, the information from the current moment is fed to the forgetting gate. It regulates the degree of information forgotten regarding the cellula state from the preceding moment. As the value of f_t approaches 0, more information is discarded, while as it approaches 1, more information is retained. The precise equation is as follows:

$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f) \tag{9}$$

Where: f_t denotes the forgetting gate, σ denotes the activation function, W_f denotes the weight matrix between the input vector and the forgetting gate, h_{t-1} represents the latent state at the preceding time step, x_t denotes the input tensor at the current moment, and b_f is the bias vector.

Second, the useful information retained by the forgetting gate is fed to the input gate. It mainly controls how many messages of the candidate state of the present time need to be saved. As seen in Fig. 4, the input gate needs to input both h_{t-1} from the preceding moment and X_t from the present moment into the $Sigmoid$ function and adjust the value within the range of 0 to 1 to determine the information that requires updating. On the other hand, it is necessary to input h_{t-1} and X_t into tanh to form the cell's state at the present time. Next, the multiplication of i_t and c_t is calculated, and the resulting value of $Sigmoid$ decides the relevant information to retain from the output value of tanh .

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i) \tag{10}$$

$$\tilde{C}_t = \tanh(W_c \cdot [h_{t-1}, x_t] + b_c) \tag{11}$$

where i_t denotes the input gate, W_t denotes the weight matrix between the input vector and the input gate, b_i denotes the bias vector, $\tilde{C}_t$ denotes the temporary cell state, tanh denotes the activation function, W_c denotes the weight matrix between the input vector and the temporary cell state, and b_c is the bias vector. Finally, the output gate is employed to regulate the degree of information from the current cell state that should be conveyed to the current hidden layer state, namely:

$$o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o) \tag{12}$$

$$C_t = f_t * C_{t-1} + i_t * \tilde{C}_t \tag{13}$$

$$h_t = o_t * \tanh(C_t) \tag{14}$$

Where: $o(t)$ is the output gate, W_o deotes the weight matrix that connects the input vector to the output gate, b_o is the bias vector, C_t denotes the cell state at the present step, h_t is the current hidden state, and $*$ denotes the Hadamard product of the matrix. LSTM models are usually employed to acquire the temporal features of the traffic stream. However, a single LSTM model with a time lag phenomenon leads to low performance of traffic flow prediction.

D. Periodic feature extraction

Temporal Convolutional Networks (TCN) can output a series of arbitrary lengths to the same length and be used in time series prediction. It differs from CNN in that TCNs use causal convolution, null convolution, and residual linking, making it possible to extract temporal features from time series and achieve prediction. Besides, TCN can efficiently address the performance decline of deep networks while training.

1) Causal convolution

Causal convolution is a model with rigorous time constraints that possesses a unidirectional structure and does not take into account future information. Suppose given the input series $\{x'_{1^p}, x'_{2^p}, \cdots, x'_{t^p}\}$, in predicting $\{\hat{X}_{1^p}, \hat{X}_{2^p}, \cdots, \hat{X}_{t^p}\}$, only the already observed series $\{x'_{1^p}, x'_{2^p}, \cdots, x'_{t^p}\}$ can be used, but not $\{x'_{t+1}, x'_{t+2}, \cdots\}$. The expression for the causal convolution is shown in Equation(15):

$$(\hat{X}_{1^p}, \hat{X}_{2^p}, \cdots, \hat{X}_{t^p}) = f(x'_{1^p}, x'_{2^p}, \cdots, x'_{t^p}) \tag{15}$$

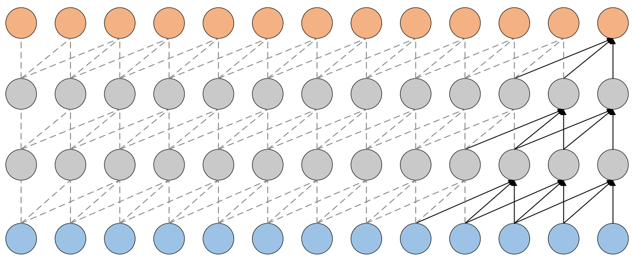


Fig. 5 Causal convolution structure diagram

Where $\hat{X}_{t^p}$ is only related to the moment t and the past input sequence, not future information. The design of the causal convolution is illustrated in Fig. 5.

2) Extended causal convolution

For causal convolution, if large samples of time series are to be considered, the number of convolution layers is bound

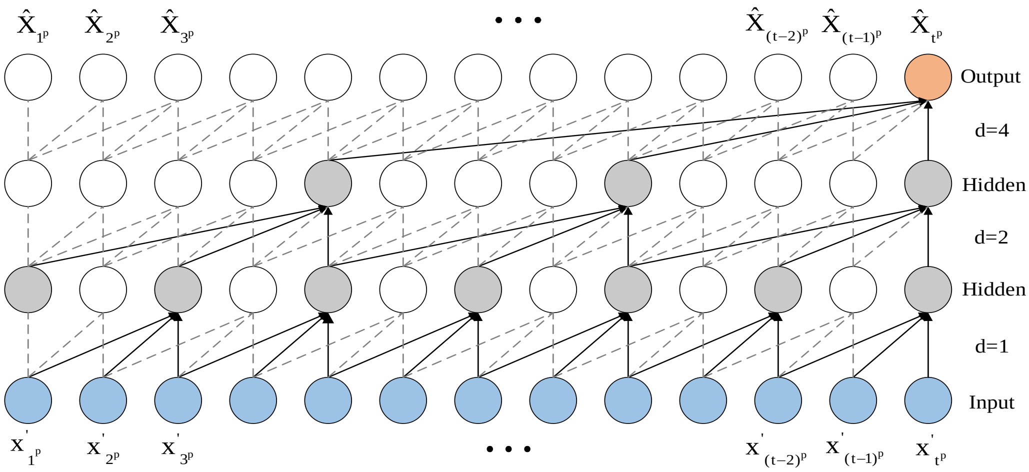


Fig. 6 Extended causal convolution structure diagram

to increase, or a larger convolution kernel is used to increase the convolution field of perception. This makes the training gradient of the network disappear, the training becomes more complicated, and the fitting effect is not good. To address this problem, the TCN model uses Dilated Causal Convolution (DCC) to augment the convolutional field of perception without significantly increasing the computational cost by skipping some of the inputs so that the convolutional kernel can be utilized for regions larger than the length of the convolutional kernel itself. The expression is shown in (16):

$$F(s)=\sum_{i=1}^{k} f(i)\cdot X'_{t^{p}_{s-d\cdot i}} \tag{16}$$

Where: $F(s)$ is the convolution result of the s element in the series $\{x'_{1^p}, x'_{2^p}, \cdots, x'_{t^p}\}$, $f(i)$ is the convolution filter, and d denotes the unfolding factor. The inputs to the extended causal convolution are sampled at intervals, and their sampling rate is determined by the expansion factor d. When d =1, it indicates that each point in the input sequence is sampled. When d =2, every second point in the input sequence is sampled. When d =4, every fourth point in the input sequence is sampled. As the network depth increases, the value of d becomes larger. Thus, extended causal convolution allows the ssuitable window size to expand exponentially as the number of layers increases, thereby achieving a larger sensory field. The construction of the extended causal convolution is shown in Fig. 6.

3) Residual links

Residual connections prove to be an valid technique for training deep neural networks, allowing information to be propagated across layers. A residual block is comprised of two convolutional layers and a non-linear mapping, incorporating WeightNorm and Dropout layers in each layer to enforce network regularization. The combined use of residual links and extended causal convolution in TCN modeling effectively improves the features learning capability and robustness of the TCN model. The residual block's network architecture is displayed in Fig. 7.

Influenced by people's travel patterns, the traffic stream data is cyclical. Due to the limitations of LSTM, a standalone LSTM model for traffic volume forecast may exhibit temporal delays. Hence, this study merges two models with distinct architectures, TCN and LSTM, to create a TCN-LSTM model, as depicted in Fig. 8. Leveraging the strengths of TCN and LSTM's different structures, the TCN-LSTM model is utilized to acquire the daily and weekly periodicities of traffic stream data to enhance the accuracy of time series forecasting.

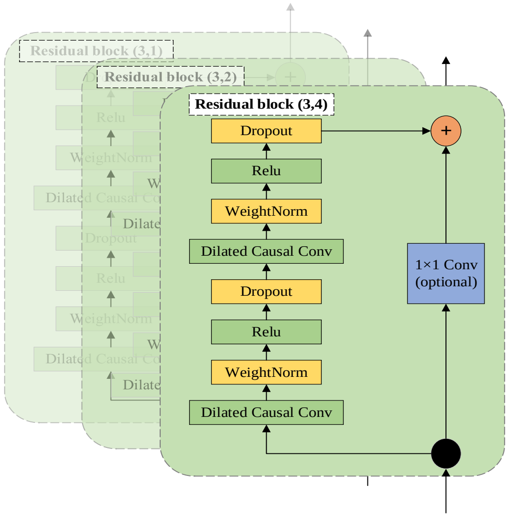


Fig. 7 Diagram of the network architecture for residual block

E. Construction of WRNCL-TCL model

In response to the analysis of the above models, this paper combines their respective advantages. It proposes a combined forecasting model (WRNCL-TCL) that considers spatio-temporal characteristics and periodic properties.

Firstly, for the problem that the collected raw traffic stream data is polluted by noise, wavelet threshold theory is introduced to realize the denoising of traffic stream data by decomposing and reconstructing the original traffic stream data. Secondly, this research considers the spatiotemporal characteristics of traffic stream and introduces CNN and LSTM models. The CNN model acquires the spatial attributes of the traffic volume data, whereas the LSTM model extract the temporal attributes of the traffic volume

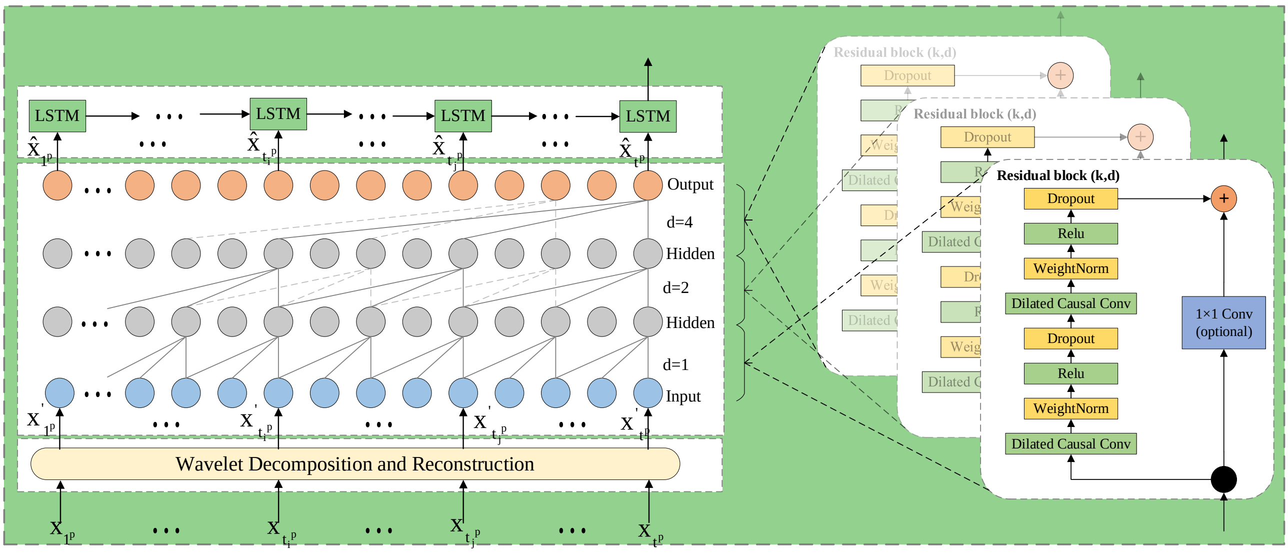


Fig. 8 Structure of TCN-LSTM model

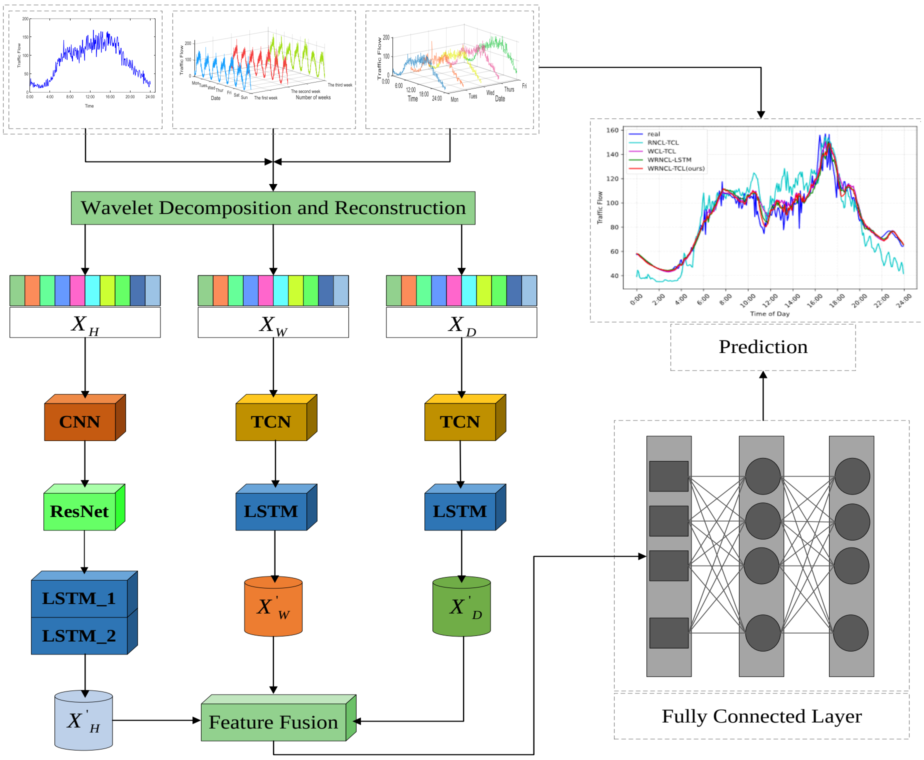


Fig. 9 Structural diagram of WRNCL-TCL model

data. Simultaneously, for the network degradation problem, the residual neural unit is added built upon the CNN model. Finally, since the single LSTM model has a time lag, the TCN model is introduced to combine these two models with different structures to build a TCN-LSTM model to obtain the periodic peculiarity of traffic stream data. The extracted spatio-temporal features and cyclic patterns are then fused through feature fusion, and the ultimate prediction outcome is achieved using a fully connected layer. Fig. 9 illustrates the model architecture of WRNCL-TCL.

The primary objective of model WRNCL-TCL is to enhance the accuracy of short-term traffic volume prediction, thereby minimizing the disparity between the predicted and actual values. The mean square error (MSE) is employed as the loss function for the WRNCL-TCL model, measuring the difference between the forecasted and real values:

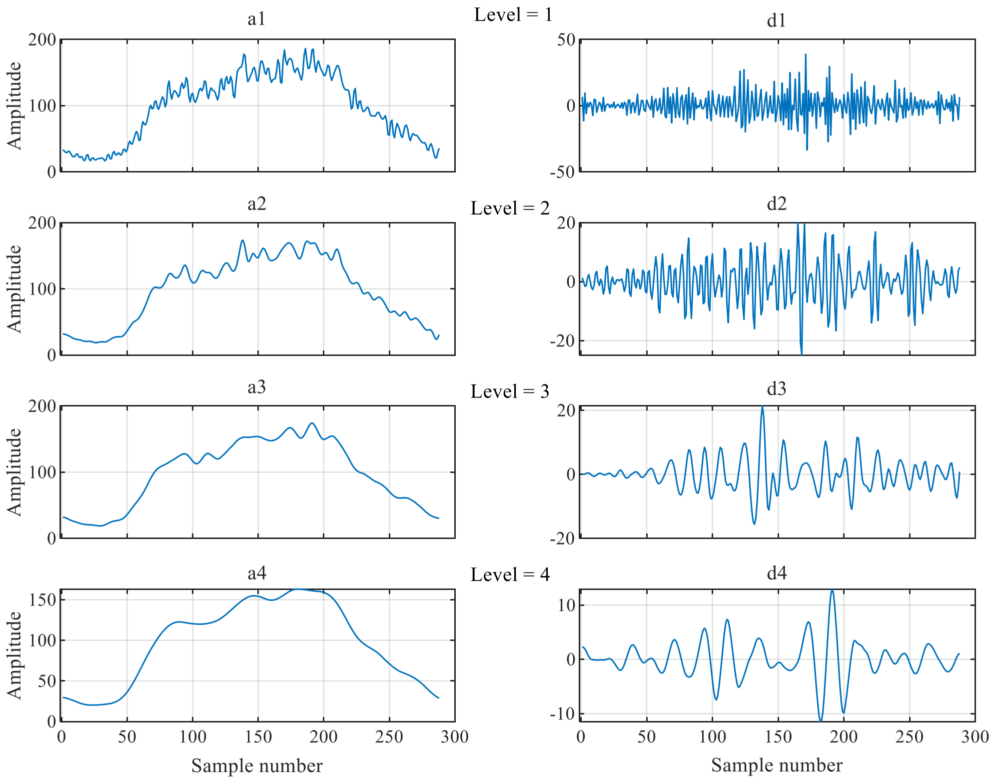


Fig. 10 Traffic flow data decomposition visualization

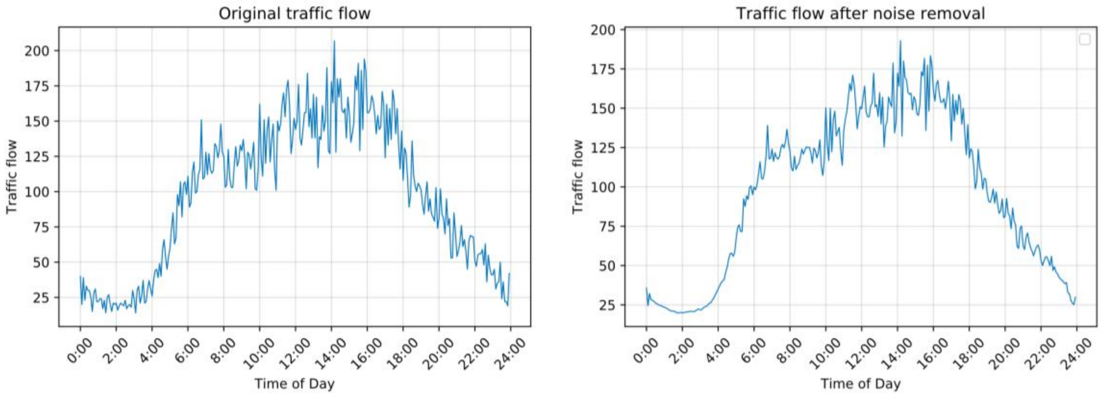


Fig. 11 Comparison of traffic flow data before decomposition (left) and after reconstruction (right)

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(\hat{y}_i - y_i)^2 \tag{17}$$

Where: y_i is the observed value, $\hat{y}_i$ is the forecasted value, and n denotes the quantity of chosen specimens.

III. Simulation Experiments

A. Data source

The dataset utilized in this study was derived from the Performance Measurement System (PeMS) [30] supported by the California Department of Transportation (Caltrans), and a total of traffic flow data, both weekday and non-weekday, was gathered between September 18, 2017, and March 4, 2018. In this paper, the WRNCL-TCL model will be trained and evaluated from two different scenarios: freeways and urban arterials where the freeways are located in SR99-S Zone 10, the study section contains seven sensors, each of which collects traffic flow data with approximately 24 weeks, and the test set is composed of one week's worth of traffic flow data, whereas the training dataset comprises the remaining data. The main city street is located at 1980 Street in Oakland's Section 4, which contains 7 sensors, and the traffic stream data collected by each sensor and its division is the same as the freeway data set.

B. Implementation details

1) Evaluation Indicators

During the laboratorial phase, to appraise the precision of the WRNCL-TCL model accurately, the assessment criteria employed in this study comprise root mean square error (RMSE), mean absolute error (MAE), mean absolute percentage error (MAPE), and coefficient of determination (R2). Their respective definitions are provided as follows:

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(\hat{y}_i - y_i)^2} \tag{18}$$

$$MAE = \frac{1}{n}\sum_{i=1}^{n}\left|\hat{y}_i - y_i\right| \tag{19}$$

$$MAPE = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{y_i - \hat{y}_i}{y_i}\right| \times 100\% \tag{20}$$

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(\hat{y}_i - y_i)^2}{\sum_{i=1}^{n}(\hat{y}_i - \bar{y}_i)^2} \tag{21}$$

Where: y_i denotes the observed value, $\hat{y}_i$ represents the forecasted value, $\bar{y}_i$ denotes the average value, and n denotes the quantity of chosen specimens.

2) Wavelet Threshold Denoising

To tackle the issue of yawp contamination in the acquired raw traffic stream data, this study introduces wavelet threshold theory to execute data denoising by decomposing and reconstructing the raw traffic stream data. As shown in Fig. 10, the traffic stream data of a certain day is selected for visualization and analysis.

From Fig. 10, the raw traffic stream data is decomposed into 4-scale decomposition into approximate coefficients and detail coefficients. After the decomposition is completed, the low-frequency coefficients of scale four and the preceding high-frequency coefficients are reconstructed for denoising. A comparison of the original traffic stream data before decomposition and after reconstruction is shown in Fig. 11.

C. Parameter setting and analysis

1) Parameter Setting

The CNN convolutional layer number is 2, and the count and size of convolutional kernels in each layer are set to 15 and 3, respectively. The TCN model's residual blocks are constructed with 3 layers, with each residual block containing 15 convolutional kernels of size 3. The dilation factor of the expandable causal convolution in the first residual block is set to 1, while in the second and third residual blocks, it is 2 and 4, severally. Additionally, dropout is 0.5. The learning rate, epoch, and batch size are 0.001, 50, and 128, correspondingly.

2) Parameter Analysis

During the model prediction process, the learning rate, batch size, number of residual neural network layers, number of residual blocks in the TCN model, and training iterations have a relatively momentous influence on the forecasted precision of the WRNCL-TCL model. This research determines their specific values through an experimental search approach.

① The impact of learning rate on model performance

TABLE I
Effect of Learning Rate on WRNCL-TCL Model

Learning rate	RMSE	MAE	MAPE	R^2
0.0001	6.742	4.448	0.047	0.936
0.0005	6.688	4.249	0.045	0.939
0.001	**6.366**	**4.033**	**0.042**	**0.941**
0.005	7.037	4.676	0.049	0.934

TABLE I demonstrates the effect of the learning rate on the predictive preccision of the model. The model's performance is assessed by experimenting with different learning rates from the predefined set {0.0001, 0.0005, 0.001, 0.005}. The outcome suggests that with the learning rate is 0.001, the assessment metrics RMSE, MAE, and MAPE of WRNCL-TCL model have the lowest values, while the R^2 value is the highest. This suggests that the model exhibits the best predictive performance.

② Impact of batch size on model performance

TABLE II
EFFECT OF BATCH SIZE ON WRNCL-TCL MODEL

batch size	RMSE	MAE	MAPE	R^2
32	6.978	4.590	0.049	0.932
64	7.138	4.804	0.051	0.929
128	**6.366**	**4.033**	**0.042**	**0.941**
256	6.394	4.399	0.049	0.941

TABLE II depicts the effect of batch size on the model's forecasting accuracy. The model's efficacy was evaluated by varying the batch size within the set {32, 64, 128, 256}, with the learning rate fixed at 0.001. The consequence indicates that when the batch size 128, the WRNCL-TCL model demonstrates the best predictive performance. At this point, its evaluation metrics, including the least values for RMSE, MAE, and MAPE, and the maximum value for R2.

③ Residual neural network layer count

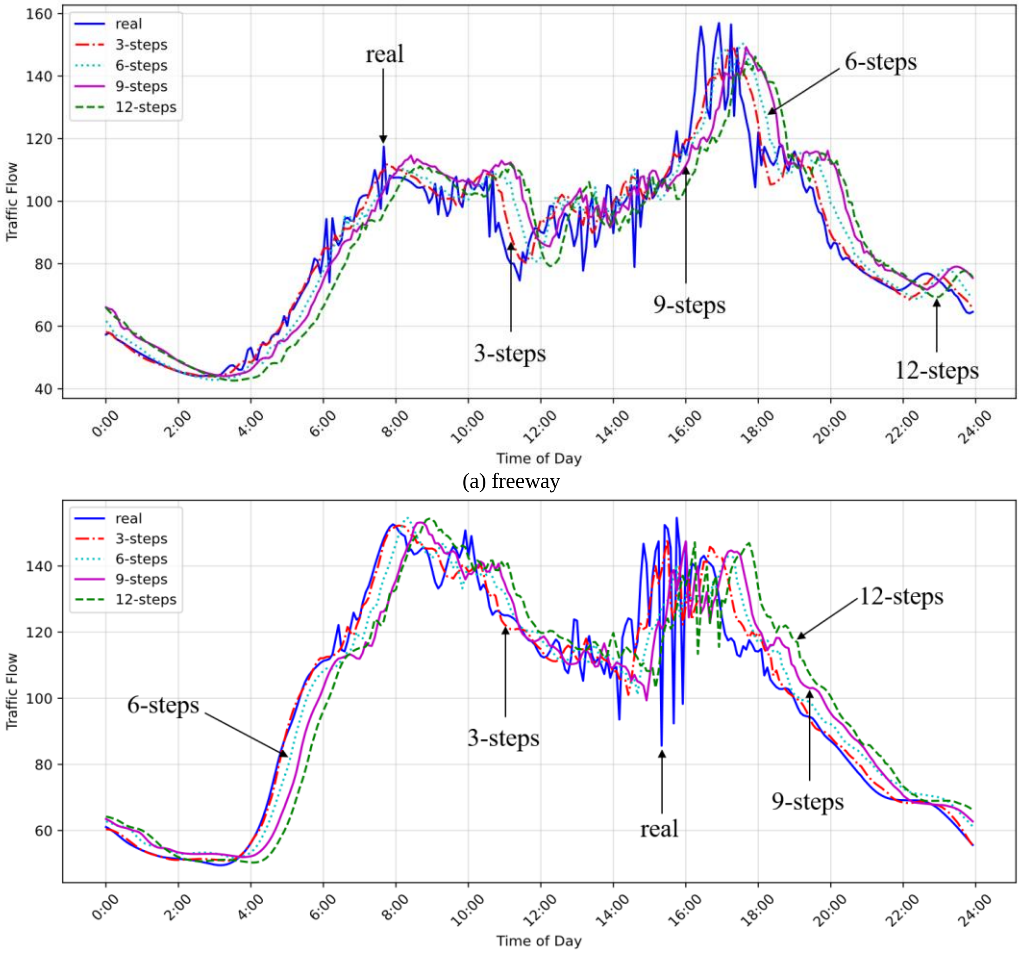
The influence of the layer count in the residual neural network on the forecasting accuracy of the model is depicted in TABLE III. The model's performance is evaluated by varying the layer count in the residual neural network within the set {2, 3, 4, 5}, with the learning rate fixed at 0.001 and batch size at 128. It can be seen that when the layer count of ResNet is 4, the evaluation metrics RMSE, MAE, and MAPE have the smallest values, and R^2 has the largest value when the model performance is the best.

TABLE III
EFFECT OF RESIDUAL NEURAL NETWORK LAYERS ON WRNCL-TCL MODEL

ResNet	RMSE	MAE	MAPE	R^2
2	7.189	4.852	0.052	0.927
3	6.721	4.415	0.047	0.937
4	**6.366**	**4.033**	**0.042**	**0.941**
5	6.907	4.456	0.047	0.934

④ Impact of the amount of residual blocks on model performance

TABLE IV displays the effect between the count of residual blocks and the model's prediction performance. When the learning rate is configured as 0.001, the batch size is specified as 128, and the residual neural network's number of layers set to 4, the model's performance was evaluated for the set {1, 2, 3, 4} of residual blocks. While the stability of the prediction outcomes is lower with 3 residual blocks compared to 2, the model exhibits the lowest prediction error with 3 residual blocks. Therefore, this research



(a) freeway

(b) urban

Fig. 12 Fitting results of the model with different prediction steps establishes the number of residual blocks as 3.

TABLE IV
EFFECT OF THE NUMBER OF RESIDUAL BLOCKS ON THE WRNCL-TCL MODEL

block	RMSE	MAE	MAPE	R^2
1	7.151	4.846	0.052	0.933
2	**6.212**	4.125	0.044	**0.946**
3	6.366	**4.033**	**0.042**	0.941
4	6.468	4.248	0.046	0.940

⑤ The effect of training times on the model

TABLE V illustrates the influence of training duration on the model's prediction performance. By fixing the learning rate at 0.001, the batch size at 128, and the count of ResNet layers and residual blocks at 4 and 3, we evaluate the model's performance across different training iterations in the range {25, 50, 75, 100}. When the count of training times is 50, the values of appraisal indicator MAE and MAPE are the smallest, and the value of R^2 is the largest, and the model has the gfirst-class performance at this time.

TABLE V
EFFECT OF TRAINING TIMES ON WRNCL-TCL MODEL

Epoch	RMSE	MAE	MAPE	R^2
25	**6.243**	4.074	0.042	0.938
50	6.366	**4.033**	**0.042**	**0.941**
75	6.482	4.236	0.045	0.940
100	6.367	4.256	0.044	0.937

D. Experimental results

To validate the forecasted capacity and generalization of the WRNCL-TCL model, the model has been experimented with on the highway and the main city road, respectively. Their errors are presented in TABLE VI.

In this experiment, we assess the predictive precision of the WRNCL-TCL model for different time intervals (15min, 30min, 45min, and 60 min). According to TABLE VI, the proposed model yields the most accurate results for the two distinct datasets for the 15-minute prediction range. Taking the highway as an example for comparative analysis, its evaluation index RMSE is 2.515, 5.315, and 6.856 lower than those under other views (30min, 45min, and 60min); MAE is 2.137, 4.634, and 6.137 lower than those under other views; MAPE is 0.026, 0.057 and 0.074 lower than those under other views, respectively; the R^2 is 0.048, 0.132 and 0.186 higher than those under other views, respectively. This could be attributed to the fact that more intricate spatial and temporal dependencies influence long-term traffic volume forecast, necessitating the consideration of additional factors. In summary, the WRNCL-TCL model performs better in forecasting short-term predictions than long-term predictions.

TABLE VI
COMPARISON OF ERRORS OF WRNCL-TCL MODEL WITH DIFFERENT PREDICTION STEPS

Time (min)		RMSE	MAE	MAPE	R^2
	15	**6.366**	**4.033**	**0.042**	**0.941**
	30	8.881 (↑2.515)	6.170 (↑2.137)	0.068 (↑0.026)	0.893 (↓0.048)
Freeway	45	11.681 (↑5.315)	8.667 (↑4.634)	0.099 (↑0.057)	0.809 (↓0.132)
	60	13.222 (↑6.856)	10.170 (↑6.137)	0.116 (↑0.074)	0.755 (↓0.186)
	15	**6.848**	**3.728**	**0.033**	**0.955**
	30	9.306 (↑2.458)	6.732 (↑3.004)	0.066 (↑0.033)	0.910 (↓0.045)
Urban	45	12.037 (↑5.189)	8.861 (↑5.133)	0.086 (↑0.053)	0.855 (↓0.1)
	60	14.051 (↑7.203)	10.980 (↑7.252)	0.108 (↑0.075)	0.815 (↓0.14)

To visualize the forecasting performance of the WRNCL-TCL model, one day of data was selected for visualization and analysis in this study. As shown in Fig. 12.

Fig. 12 indicates traffic flow is highly erratic on the highway and main urban thoroughfare. As illustrated in (a), the traffic stream exhibits prominent spikes at 8:00 and 17:00. Similarly, (b) shows that the traffic stream experiences significant peaks at 8:00 and 16:00. When

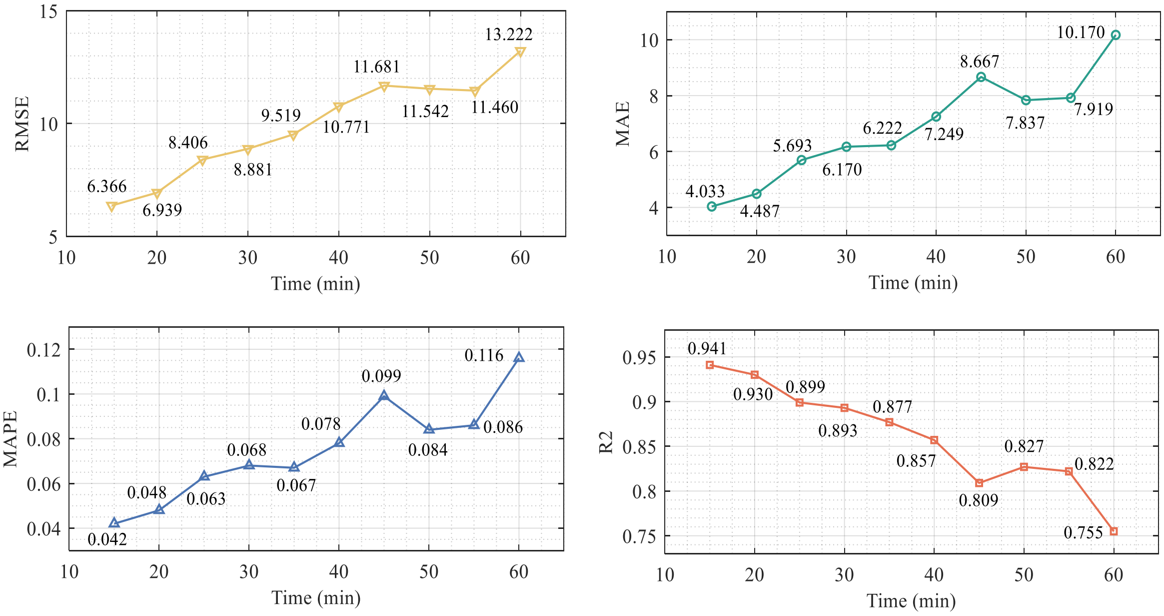
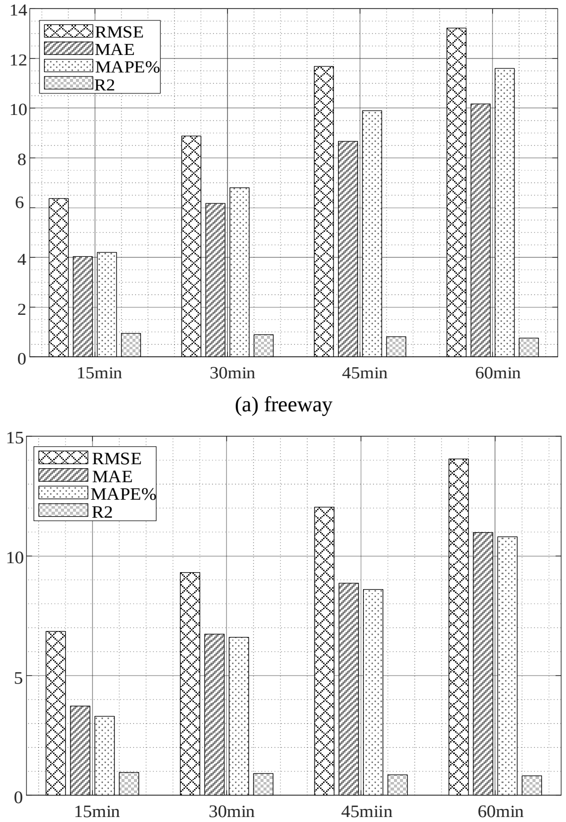


Fig. 13 Prediction error of WRNCL-TCL model at 10 horizons

visualizing the forecasting outcomes of the WRNCL-TCL model from different perspectives, it is evident that the proposed model can better predict traffic stream trends under various viewpoints, whether on a highway or urban arterial road.



(b) urban

Fig. 14 Error of the model with different prediction steps

Fig. 13 depicts the visualization of forecasting errors for the WRNCL-TCL model across 10 horizons, taking the highway as an example. It is evident that the deviation of the WRNCL-TCL model is the smallest, and its predictive performance is the best in the 15-min prediction range. With the extension of the forecasting range, the model's prediction performance gradually decreases because the traffic stream is more disturbed by external factors as the prediction step length increases. To be able to verify that the WRNCL-TCL model has this phenomenon on different data sets, four different prediction ranges of 15, 30, 45, and 60min are selected in this paper to visualize the predictive errors of the model on two different datasets (freeway and urban) within the above prediction ranges. Fig. 14 displays that the WRNCL-TCL model delivers the most accurate forecasting outcomes within the 15-minute range and is the closest to the actual data for both datasets. However, the model's forecasting accuracy deteriorates gradually as the prediction stride expands.

E. Experimental comparison analysis

1) Comparative analysis with the benchmark model

To validate the forecasted capabilities of the WRNCL-TCL model, several other conventional reference models are chosen for comparison and analysis in this study. The benchmark models considered for evaluation are outlined as follows:

◆HA: Historical Average Model

◆LSTM: Long Short-Term Memory Network

◆GRU: Gated Recurrent Unite

◆AT-Conv-LSTM: A Hybrid Deep Learning Model With Attention-Based Conv-LSTM Networks

This paper compares and analyzes the above benchmark model with the WRNCL-TCL model under different

TABLE VII

COMPARISON OF THE ERRORS OF DIFFERENT PREDICTION MODELS WITH DIFFERENT PREDICTION STEPS

Time		Metric	HA	LSTM	GRU	AT-Conv-LSTM	WRNCL-TCL
freeway	15min	RMSE	16.418(↑10.052)	16.283(↑9.917)	16.027(↑9.661)	15.152(↑8.786)	**6.366**
		MAE	12.662(↑8.629)	12.069(↑8.036)	11.685(↑7.652)	11.088(↑7.055)	**4.033**
		MAPE	0.165(↑0.123)	0.179(↑0.137)	0.162(↑0.12)	0.154(↑0.112)	**0.042**
	30min	RMSE	17.657(↑8.776)	16.640(↑7.759)	16.615(↑7.734)	16.683(↑7.802)	**8.881**
		MAE	13.741(↑7.571)	12.244(↑6.074)	12.189(↑6.019)	11.897(↑5.727)	**6.170**
		MAPE	0.181(↑0.113)	0.180(↑0.112)	0.177(↑0.109)	0.156(↑0.088)	**0.068**
	45min	RMSE	18.960(↑7.279)	17.322(↑5.641)	17.498(↑5.817)	17.136(↑5.455)	**11.681**
		MAE	14.853(↑6.186)	12.750(↑4.083)	12.855(↑4.188)	12.650(↑3.983)	**8.667**
		MAPE	0.197(↑0.098)	0.187(↑0.088)	0.182(↑0.083)	0.207(↑0.108)	**0.099**
	60min	RMSE	20.306(↑7.084)	17.965(↑4.743)	18.162(↑4.94)	17.906(↑4.684)	**13.222**
		MAE	15.982(↑5.812)	13.224(↑3.054)	13.348(↑3.178)	12.951(↑2.781)	**10.170**
		MAPE	0.215(↑0.099)	0.193(↑0.077)	0.193(↑0.077)	0.190(↑0.074)	**0.116**
urban	15min	RMSE	28.428(↑21.58)	16.848(↑10)	16.463(↑9.615)	15.524(↑8.676)	**6.848**
		MAE	20.720(↑16.992)	13.182(↑9.454)	12.120(↑8.392)	11.466(↑7.738)	**3.728**
		MAPE	0.152(↑0.119)	0.175(↑0.142)	0.122(↑0.089)	0.127(↑0.094)	**0.033**
	30min	RMSE	31.453(↑22.147)	17.793(↑8.487)	16.962(↑7.656)	16.860(↑7.554)	**9.306**
		MAE	22.964(↑16.232)	13.689(↑6.957)	13.351(↑6.619)	12.819(↑6.087)	**6.732**
		MAPE	0.172(↑0.106)	0.178(↑0.112)	0.150(↑0.084)	0.148(↑0.082)	**0.066**
	45min	RMSE	34.536(↑22.499)	18.944(↑6.907)	17.532(↑5.495)	17.393(↑5.356)	**12.037**
		MAE	25.217(↑16.356)	14.466(↑5.605)	14.417(↑5.556)	14.185(↑5.324)	**8.861**
		MAPE	0.193(↑0.107)	0.187(↑0.101)	0.168(↑0.082)	0.181(↑0.095)	**0.086**
	60min	RMSE	37.646(↑23.595)	19.611(↑5.56)	18.975(↑4.924)	18.415(↑4.364)	**14.051**
		MAE	27.470(↑16.49)	14.837(↑3.857)	14.969(↑3.989)	14.805(↑3.825)	**10.980**
		MAPE	0.215(↑0.107)	0.191(↑0.083)	0.171(↑0.063)	0.188(↑0.08)	**0.108**

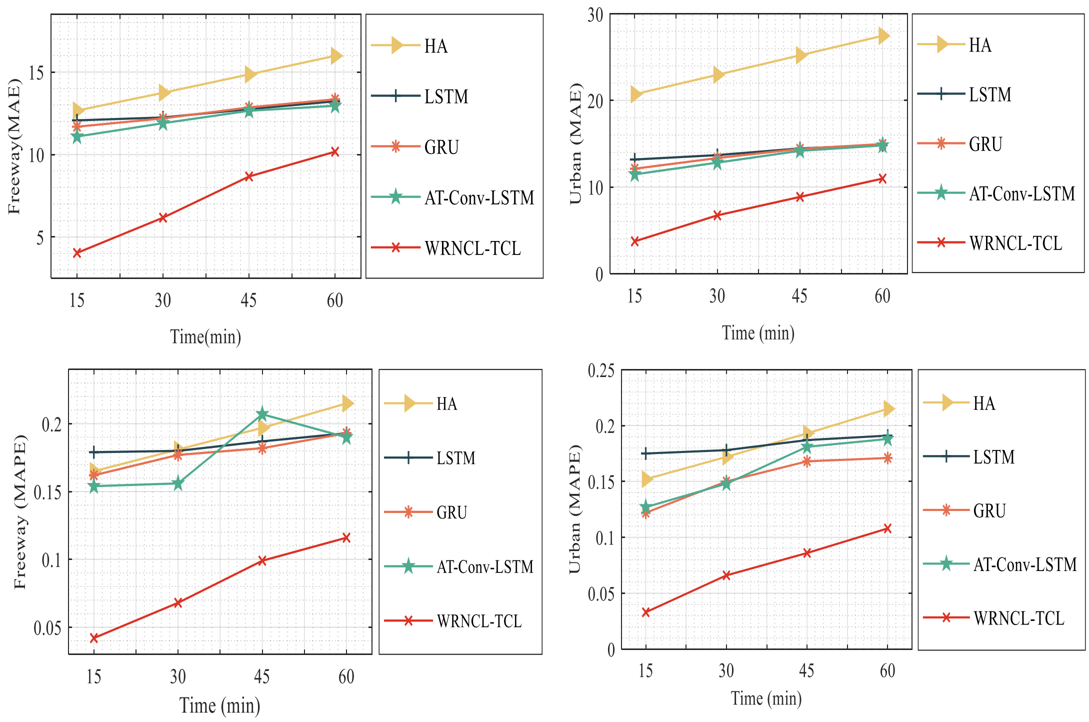


Fig. 15 Error visualization of different models with different prediction steps

horizons. Although the HA model is simpler and faster to compute, TABLE VII reveals that the model exhibits the poorest forecasting accuracy, which is because the model is not able to handle complex traffic conditions. Following that, with the emergence of deep learning, certain researchers utilized models like LSTM and GRU for short-term traffic volume forecast. Although the predictive capability of LSTM and GRU models surpasses that of HA models, it still falls short compared to AT-Conv-LSTM models. This is mainly because LSTM and GRU models do not comprehensively capture traffic flow's spatial characteristics. AT-Conv-LSTM models, on the other hand, take into consideration the spatiotemporal dependencies and periodicity of the traffic stream, thus exhibiting superior prediction accuracy. However, compared with the WRNCL-TCL model, the predictive performance of AT-Conv-LSTM is relatively inferior due to its failure to account for the potential deterioration of the model as the network depth increases and the noise contamination of the raw data.

In order to see more intuitively how the WRNCL-TCL model compares with the benchmark model, the MAE and MAPE metrics of the above model on the two data sets are visualized in this paper. Fig. 15 illustrates that the proposed model in the paper exhibits the most accurate forecasting performance.

2) Comparative analysis with ablation experiments

To validate the forecasted performance of the presented model in this research, a contrast experiment is carried out between the WRNCL-TCL and ablation models, namely RNCL-TCL, WCL-TCL, and WRNCL-LSTM. The prediction effectiveness of each ablation model is presented in TABLE VIII for various time intervals.

◆RNCL-TCL: The impact of the noise in the raw data on the short-time traffic volume forecast is not considered. The

TABLE VIII
ERROR COMPARISON OF ABLATION MODEL UNDER DIFFERENT PREDICTION STEPS

Time (min)		Freeway			Urban		
		RMSE	MAE	MAPE	RMSE	MAE	MAPE
15	RNCL-TCL	14.609	11.994	0.153	13.683	11.467	0.126
	WCL-TCL	6.976	4.454	0.048	6.885	3.981	0.036
	WRNCL-LSTM	6.555	4.129	0.043	7.419	4.855	0.048
	WRNCL-TCL	6.366	4.033	0.042	6.848	3.728	0.033
30	RNCL-TCL	14.172	11.574	0.147	15.399	12.462	0.138
	WCL-TCL	8.869	6.180	0.068	9.412	6.578	0.063
	WRNCL-LSTM	8.997	6.244	0.068	9.803	6.874	0.067
	WRNCL-TCL	8.881	6.170	0.068	9.306	6.732	0.066
45	RNCL-TCL	15.037	12.103	0.153	16.664	13.022	0.141
	WCL-TCL	11.355	8.387	0.094	11.699	8.784	0.086
	WRNCL-LSTM	11.464	8.321	0.092	12.056	9.134	0.090
	WRNCL-TCL	11.681	8.667	0.099	12.037	8.861	0.086
60	RNCL-TCL	15.961	12.887	0.160	17.766	13.242	0.140
	WCL-TCL	13.451	10.285	0.117	13.935	10.837	0.107
	WRNCL-LSTM	13.478	10.224	0.115	14.161	11.133	0.111
	WRNCL-TCL	13.222	10.170	0.116	14.051	10.980	0.108

CNN-LSTM architecture captures spatial and temporal traffic flow characteristics while integrating residual neural units into the CNN model to mitigate potential network degradation problems. The TCN-LSTM is leveraged to acquire the periodic peculiarity of the traffic stream.

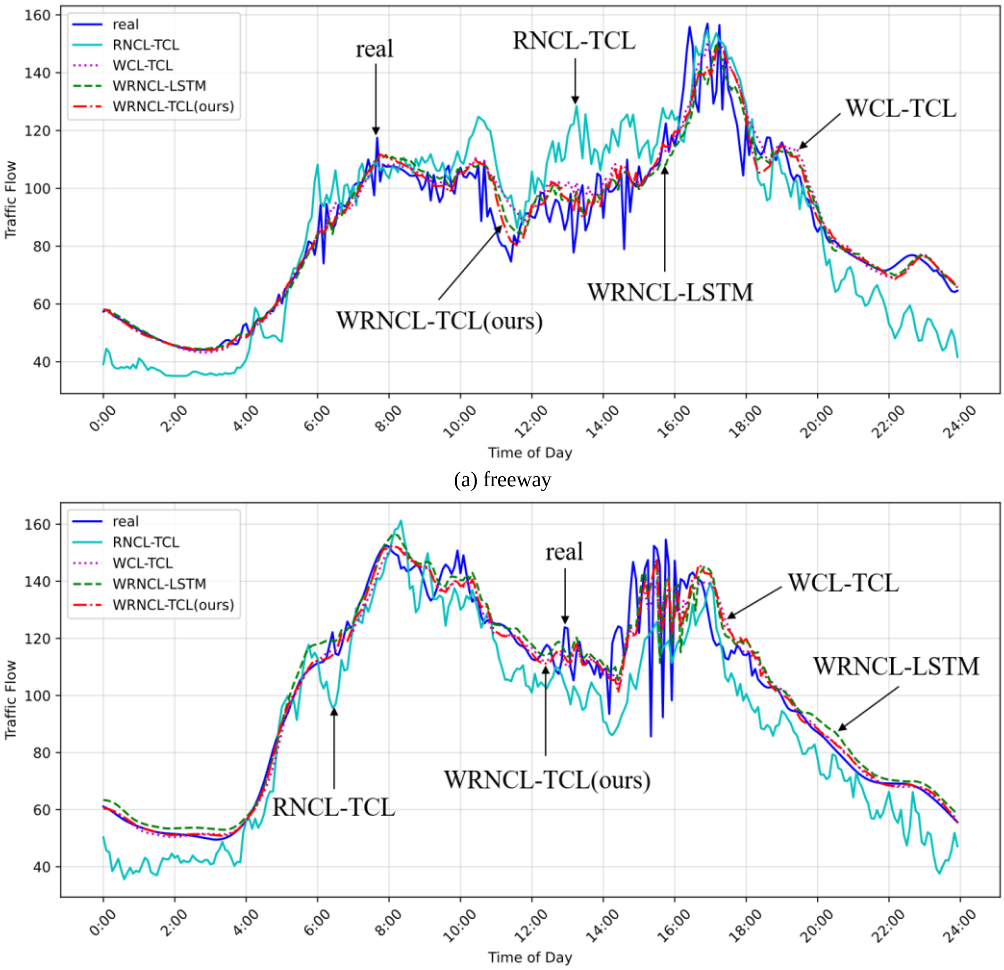
◆WCL-TCL: The possible degradation of the network is not considered. The initial traffic stream data undergoes denoising using the wavelet threshold technique before employing the CNN-LSTM model to acquire the spatiotemporal peculiarity of the traffic stream. Additionally, TCN-LSTM is applied to obtain the periodic peculiarity of the traffic volume.

◆WRNCL-LSTM: The time lag problem when a single LSTM is utilized for traffic volume forecast is not addressed. The wavelet thresholding technique is applied to denoise the initial traffic stream data to overcome this. Following that, the RNCL model is applied to obtain the spatio-temporal attributes of the traffic stream. Simultaneously, the LSTM is applied to acquire the cyclical characteristics of the traffic stream.

TABLE VIII demonstrates that the proposed model delivers the most accurate forecasting results for the 15-minute prediction range. However, from the 30-minute mark and beyond, the model's forecasting accuracy declines, primarily due to the influence of more intricate spatiotemporal dependencies that arise in long-term traffic volume forecasts. Therefore, in addition to accounting for the spatiotemporal characteristics and periodicity of the traffic stream, it is crucial to consider other factors that impact the traffic stream when forecasting. Using the expressway as a case study, the forecasting outcomes of the WRNCL-TCL model and each ablation model in the 15-minute prediction range are analyzed. It can be seen that compared with RNCL-TCL, WCL-TCL, and WRNCL-TCL, the RMSE of this paper model decreases by 8.243, 0.61, and 0.189, respectively; MAE decreases by 7.961, 0.421, and 0.096, respectively; and MAPE decreases by 0.111, 0.006, and 0.001, respectively. This is due to the utilization of wavelet threshold theory for data noise removal in this paper's model, ResNet addresses the potential degradation issue that can arise with increased network depth, and the combined application of TCN and LSTM models resolves the time delay problem present in the standalone LSTM model.

To visualize the forecasting effects of the WRNCL-TCL model and the ablation models RNCL-TCL, WCL-TCL, and WRNCL-TCL more visually, one day of data was selected in this paper to visualize and analyze the models in this paper and each ablation model in the 15-minute prediction range. As depicted in Fig. 16.

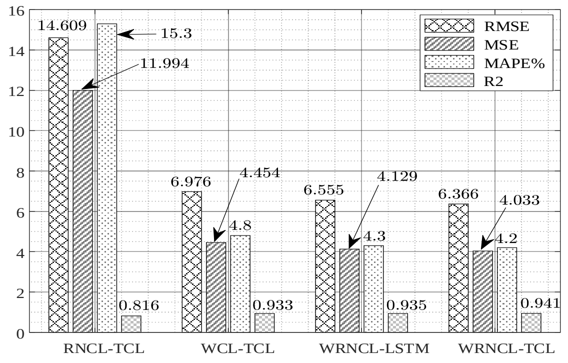
From Fig. 16, the fit between the forecasted curve of the RNCL-TCL model and the practical curve is the worst because the collected raw data are contaminated by noise, which can interfere greatly with the model's prediction. Therefore, it is imperative to apply denoising techniques when conducting traffic flow prediction. Although the



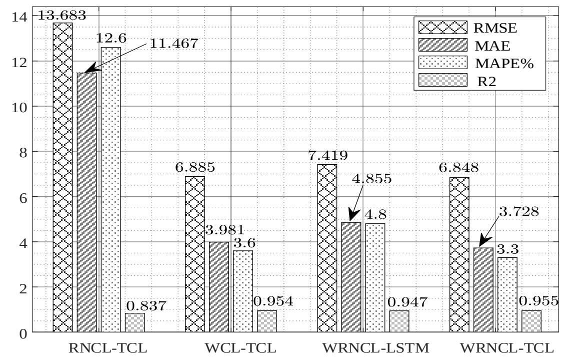
(a) freeway

(b) urban

Fig. 16 Fitting results of different ablation models in the 15-minute prediction range



(a) freeway

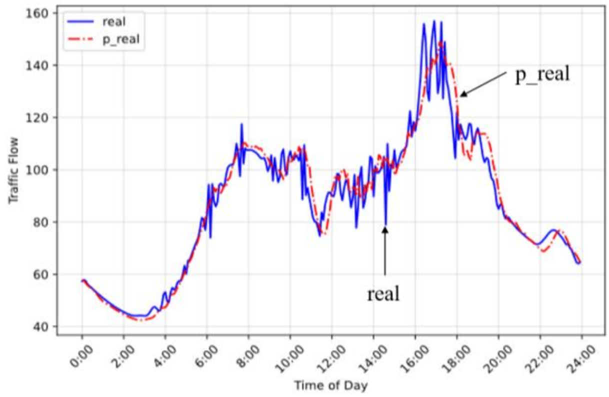


(b) urban

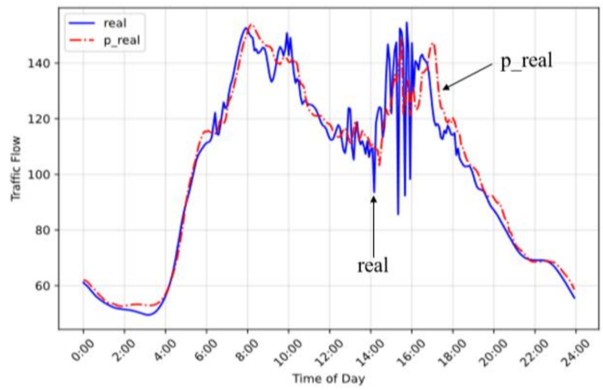
Fig. 17 Comparison of the errors of different ablation models in the 15-minute prediction range

WCL-TCL model and WRNCL-LSTM model can accurately forecast the traffic flow trend, their accuracy is inferior compared to the model introduced in this research. The R^2 evaluation metric is primarily employed to assess the level of model fitting. From Fig. 17(a), it can be observed that R^2 of the model introduced in this study is increased by 0.8% and 0.6% compared to models WCL-TCL and WRNCL-LSTM, respectively. Similarly, from Fig. 17(b), it is evident that the R^2 of the model in this research is enhanced by 0.1% and 0.8% compared to models WCL-TCL and WRNCL-LSTM, respectively.

Given the evident periodic characteristics of traffic flow, it is influenced not just by daily periodicity but also by weekly periodicity. In this paper, the model considers the effects of its daily and weekly periodicity on the prediction model's accuracy when performing traffic volume forecasts. This can be observed in Fig. 18.



(a) freeway



(b) urban

Fig. 19 Traffic flow fitting results considering daily periodicity

The traffic volume is affected not only by the spatio-temporal characteristics but also by the periodic

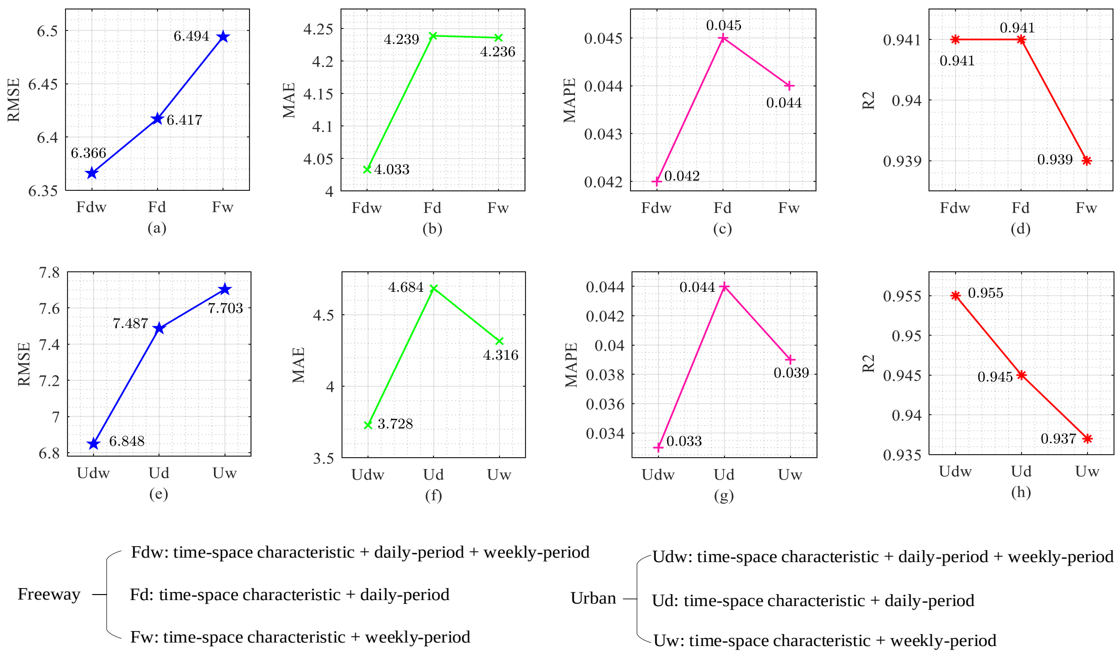
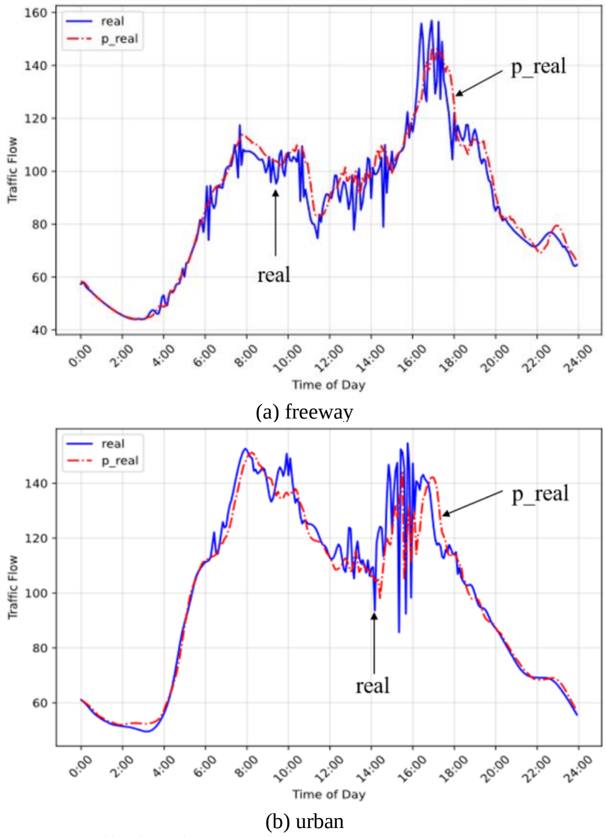


Fig. 18 Error comparison of WRNCL-TCL model with different period characteristics

characteristics. Fig. 18 illustrates that every evaluation metric of the WRNCL-TCL model is the best when the spatio-temporal characteristics are considered along with the daily and weekly peculiarity of traffic stream. Taking the urban trunk road as an example, it is evident that when the spatiotemporal characteristics and daily-periodic as well as weekly-periodic characteristics of traffic stream are considered, the evaluation indexes RMSE are 0.639 and 0.855 lower; MAE are 0.956 and 0.588 lower; MAPE are 1.1% and 0.6% lower; and R^2 are 1% and 1.8%.

The cyclical nature of traffic flow has a significant impact on model prediction. Typically, traffic stream is higher in the morning and afternoon, and traffic congestion occurs because people need to travel to and from work. In contrast, traffic volume is relatively low in midday and evening. For weekly cyclicity, traffic flows are higher on weekdays and lower on weekends. This cyclicality has a great impact on traffic congestion and traffic management, so it is essential to think about the effect of cyclic patterns on traffic stream when conducting traffic stream predictions. In order to be able to further understand the effect of periodicity on model performance, we selected one day of data. We visualized and analyzed the prediction results of traffic flow considering daily periodicity and weekly periodicity over a 15-minute prediction range. This is illustrated in Fig. 19 and Fig. 20.



(a) freeway

(b) urban

Fig. 20 Traffic flow fitting results considering weekly periodicity

Based on the observations from Fig. 19 and Fig. 20, it is evident that effective traffic stream trend prediction is achievable by taking into account the daily or weekly periodicity of traffic stream. The predicted curves also exhibit better fitting to the actual curves. However, Fig. 18 demonstrates that the R^2 is better than both, considering only the daily or weekly periodicity when considering the traffic stream's daily or weekly periodicity.

IV. Conclusion

This study proposes a novel integrated forecasting model (WRNCL-TCL) for short-term traffic volume forecast that accounts for both spatiotemporal characteristics and cyclic patterns. The initial traffic volume data is preprocessed using the wavelet thresholding technique. The CNN and LSTM architectures are employed to obtain the spatiotemporal attributes of the traffic volume. Simultaneously, the CNN model is enhanced with a residual neural network to mitigate the issue of network degradation. Additionally, the TCN and LSTM models are employed to acquire the cyclic patterns of the traffic stream. The WRNCL-TCL model is evaluated against other methods, and the results show that it outperforms them in all evaluation metrics. However, although the WRNCL-TCL model considers the effect of spatiotemporal and periodic features on traffic stream, the experimental findings indicate that the forecasting results are relatively unsatisfactory when utilizing the model presented in this paper for long-range traffic volume prediction. This is due to the greater reliance on long-term traffic volume prediction on spatiotemporal characteristics and additional factors. Consequently, in forthcoming research, it would be beneficial to consider the effect of traffic accidents, POI, and other exterior variables on traffic volume prediction to enhance the predictive capability of this model.

References

[1] W. Du, Q. Zhang, Y. Chen, and Z. Ye, "An urban short-term traffic flow prediction model based on wavelet neural network with improved whale optimization algorithm," *Sustainable Cities and Society*, vol. 69, p. 102858, 2021/06/01/ 2021.

[2] T. Alghamdi, K. Elgazzar, M. Bayoumi, T. Sharaf, and S. Shah, "Forecasting Traffic Congestion Using ARIMA Modeling," in *2019 15th International Wireless Communications & Mobile Computing Conference (IWCMC)*, 2019, pp. 1227-1232.

[3] H. Jiang, C. Ye, X. Deng, H. Hu, and H. Zhou, "Deep Learning for Short-term Traffic Conditions Prediction," in *2020 International Conference on Service Science (ICSS)*, 2020, pp. 70-75.

[4] J. Tang, X. Chen, Z. Hu, F. Zong, C. Han, and L. Li, "Traffic flow prediction based on combination of support vector machine and data denoising schemes," *Physica A: Statistical Mechanics and its Applications*, vol. 534, p. 120642, 2019/11/15/ 2019.

[5] A. Thaduri, V. Polepally, and S. Vodithala, "Traffic Accident Prediction based on CNN Model," in *2021 5th International Conference on Intelligent Computing and Control Systems (ICICCS)*, 2021, pp. 1590-1594.

[6] D. Yu, Y. Liu, and X. Yu, "A Data Grouping CNN Algorithm for Short-Term Traffic Flow Forecasting," in *Web Technologies and Applications*, Cham, 2016: Springer International Publishing, pp. 92-103.

[7] J. Mo, Z. Gong, and J. Chen, "Attentive differential convolutional neural networks for crowd flow prediction," *Know.-Based Syst.*, vol. 258, no. C, p. 13, 2022.

[8] 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, pp. 20708-20722, 2019.

[9] T. Nguyen, G. Nguyen, and B. M. Nguyen, "EO-CNN: An Enhanced CNN Model Trained by Equilibrium Optimization for Traffic Transportation Prediction," *Procedia Computer Science*, vol. 176, pp. 800-809, 2020/01/01/ 2020.

[10] Yue Pang, Wei Zhao, Yannan Zhang, and Hongke Xu, "Deep learning based LSTM for traffic flow prediction," *Microcontrollers & Embedded Systems*, vol. 19, no. 03, pp. 72-75, 2019. (in Chinese)

[11] B. Yang, S. Sun, J. Li, X. Lin, and Y. Tian, "Traffic flow prediction using LSTM with feature enhancement," *Neurocomputing*, vol. 332, pp. 320-327, 2019/03/07/ 2019.

[12] L. Mou, P. Zhao, H. Xie, and Y. Chen, "T-LSTM: A Long Short-Term

Memory Neural Network Enhanced by Temporal Information for Traffic Flow Prediction," *IEEE Access,* vol. 7, pp. 98053-98060, 2019.
[13] H. Huang, T. Wang, J. Liu, and S. Xie, "Predicting Urban Rail Traffic Passenger Flow Based on LSTM," in *2019 IEEE 3rd Information Technology, Networking, Electronic and Automation Control Conference (ITNEC)*, 2019, pp. 616-620.
[14] Y. Xiao and Y. Yin, "Hybrid LSTM Neural Network for Short-Term Traffic Flow Prediction," *Information,* vol. 10, p. 105, 03/08 2019.
[15] S. Lu, Q. Zhang, G. Chen, and D. Seng, "A combined method for short-term traffic flow prediction based on recurrent neural network," *Alexandria Engineering Journal,* vol. 60, no. 1, pp. 87-94, 2021/02/01/ 2021.
[16] M. Cao, V. O. K. Li, and V. W. S. Chan, "A CNN-LSTM Model for Traffic Speed Prediction," in *2020 IEEE 91st Vehicular Technology Conference (VTC2020-Spring)*, 2020, pp. 1-5.
[17] Lei Li, Qingmiao Zhang, Junhui Zhao, and Yiwen Nie, "Hourly short-term traffic flow prediction based on improved CNN-LSTM combination model," *Journal of Applied Sciences,* vol. 39, no. 02, pp. 185-198, 2021. (in Chinese)
[18] X. Luo, D. Li, Y. Yang, and S. Zhang, "Spatiotemporal Traffic Flow Prediction with KNN and LSTM," *Journal of Advanced Transportation,* vol. 2019, p. 4145353, 2019/02/27 2019.
[19] H. Zheng, F. Lin, X. Feng, and Y. Chen, "A Hybrid Deep Learning Model With Attention-Based Conv-LSTM Networks for Short-Term Traffic Flow Prediction," *IEEE Transactions on Intelligent Transportation Systems,* vol. 22, no. 11, pp. 6910-6920, 2021.
[20] Yueyi Pu, Wenhan Wang, Qiang Zhu, and Pengpeng Chen, "An algorithm for urban short-term traffic flow prediction based on CNN-ResNet-LSTM model," *Journal of Beijing University of Posts and Telecommunications,* vol. 43, no. 05, pp. 9-14, 2020. (in Chinese)
[21] R. He, Y. Liu, Y. Xiao, X. Lu, and S. Zhang, "Deep spatio-temporal 3D densenet with multiscale ConvLSTM-Resnet network for citywide traffic flow forecasting," *Knowledge-Based Systems,* vol. 250, p. 109054, 08/01 2022.
[22] Y. Liu, Y. Song, Y. Zhang, and Z. Liao, "WT-2DCNN: A convolutional neural network traffic flow prediction model based on wavelet reconstruction," *Physica A: Statistical Mechanics and its Applications,* vol. 603, p. 127817, 2022/10/01/ 2022.
[23] S. Zhu, Y. Zhao, Y. Zhang, Q. Li, W. Wang, and S. Yang, "Short-Term Traffic Flow Prediction With Wavelet and Multi-Dimensional Taylor Network Model," *IEEE Transactions on Intelligent Transportation Systems,* vol. 22, no. 5, pp. 3203-3208, 2021.
[24] Yating Cheng, Shengli Zhao, and Yuanli Gu, "Short-time traffic flow prediction model considering noise effects and validation," *Technology & Economy in Areas of Communications,* vol. 25, no. 02, pp. 11-16+23, 2023.(in Chinese)
[25] R. Liu, Y. Wang, H. Zhou, and Z. Qian, "Short-Term Passenger Flow Prediction Based on Wavelet Transform and Kernel Extreme Learning Machine," *IEEE Access,* vol. 7, pp. 158025-158034, 2019.
[26] Z. Hou, Z. Du, G. Yang, and Z. Yang, "Short-Term Passenger Flow Prediction of Urban Rail Transit Based on a Combined Deep Learning Model," *Applied Sciences,* vol. 12, p. 7597, 07/28 2022.
[27] J. Bi, H. Yuan, K. Xu, H. Ma, and M. Zhou, "Large-scale Network Traffic Prediction With LSTM and Temporal Convolutional Networks," in *2022 International Conference on Robotics and Automation (ICRA)*, 2022, pp. 3865-3870.
[28] H. Li, J. Sun, and X. Liao, "A Novel Short-term Load Forecasting Model by TCN-LSTM Structure with Attention Mechanism," in *2022 4th International Conference on Machine Learning, Big Data and Business Intelligence (MLBDBI)*, 2022, pp. 178-182.
[29] Zhiming Gui, Zhuangzhuang Li, and Limin Guo, "Short-time traffic flow prediction based on ACGRU model," *Computer Engineering and Applications,* vol. 56, no. 21, pp. 260-265, 2020. (in Chinese)
[30] H. Zheng, F. Lin, X. Feng, and Y. Chen, "A Hybrid Deep Learning Model With Attention-Based Conv-LSTM Networks for Short-Term Traffic Flow Prediction," *Trans. Intell. Transport. Sys*., vol. 22, no. 11, pp. 6910–6920, 2021.

Qingrong Wang was born in 1977, female, professor, master's supervisor, and teacher at the Department of Computer Science and Technology. Her main research direction is big data and the application of data mining in intelligent transportation.

Xiaohong Chen was born in 1995, female, master student. Her research interests are machine learning and intelligent transportation.

Changfeng Zhu was born in 1972, male, professor, doctoral supervisor, Deputy Dean of School of Traffic and Transportation Lanzhou Jiaotong University. His research interest covers rail transit organization, rail transit system optimization and decision-making.

Kai Zhang was born in 1997, male, master student. His main research interests are machine learning and intelligent transportation.

Runtian He was born in 1993, male, doctoral student. His main research interests are Machine Learning and Intelligent Transporation System.

Jinhao Fang was born in 1992, male, doctoral student. His main research interests are transportation planning, three-dimensional transportation.