

WT-FCTGN: A wavelet-enhanced fully connected time-gated neural network for complex noisy traffic flow modeling

Zhifang Liao(廖志芳), Ke Sun(孙轲), Wenlong Liu(刘文龙), Zhiwu Yu(余志武), Chengguang Liu(刘承光), and Yucheng Song(宋禹成)

Citation: Chin. Phys. B, 2024, 33 (**7**): 078901. DOI: 10.1088/1674-1056/ad3349
Journal homepage: http://cpb.iphy.ac.cn; http://iopscience.iop.org/cpb

What follows is a list of articles you may be interested in

Identifying influential spreaders in complex networks based on density entropy and community structure

Zhan Su(苏湛), Lei Chen(陈磊), Jun Ai(艾均), Yu-Yu Zheng(郑雨语), and Na Bie(别娜)
Chin. Phys. B, 2024, 33 (**5**): 058901. DOI: 10.1088/1674-1056/ad20d6

Source localization in signed networks with effective distance

Zhi-Wei Ma(马志伟), Lei Sun(孙蕾), Zhi-Guo Ding(丁智国), Yi-Zhen Huang(黄宜真), and Zhao-Long Hu(胡兆龙)
Chin. Phys. B, 2024, 33 (**2**): 028902. DOI: 10.1088/1674-1056/ad1482

Identifying multiple influential spreaders in complex networks based on spectral graph theory

Dong-Xu Cui(崔东旭), Jia-Lin He(何嘉林), Zi-Fei Xiao(肖子飞), and Wei-Ping Ren(任卫平)
Chin. Phys. B, 2023, 32 (**9**): 098904. DOI: 10.1088/1674-1056/acac0b

Self-similarity of complex networks under centrality-based node removal strategy

Dan Chen(陈单), Defu Cai(蔡德福), and Housheng Su(苏厚胜)
Chin. Phys. B, 2023, 32 (**9**): 098903. DOI: 10.1088/1674-1056/acd3e2

A cellular automata traffic flow modeling of desired speed variability

K. Bentaleb, K. Jetto, H. Ez-Zahraouy, A. Benyoussef
Chin. Phys. B, 2013, 22 (**1**): 018902. DOI: 10.1088/1674-1056/22/1/018902

WT-FCTGN: A wavelet-enhanced fully connected time-gated neural network for complex noisy traffic flow modeling

Zhifang Liao(廖志芳)[1], Ke Sun(孙轲)[1], Wenlong Liu(刘文龙)[1], Zhiwu Yu(余志武)[2,3], Chengguang Liu(刘承光)[4,†], and Yucheng Song(宋禹成)[1,‡]

[1] *School of Computer Science and Engineering, Central South University, Changsha* 410083, *China*
[2] *National Engineering Research Center of High-speed Railway Construction Technology, Changsha* 410075, *China*
[3] *School of Civil Engineering, Central South University, Changsha* 410075, *China*
[4] *School of Traffic & Transportation Engineering, Central South University, Changsha* 410083, *China*

(Received 8 January 2024; revised manuscript received 26 February 2024; accepted manuscript online 13 March 2024)

Accurate forecasting of traffic flow provides a powerful traffic decision-making basis for an intelligent transportation system. However, the traffic data's complexity and fluctuation, as well as the noise produced during collecting information and summarizing original data of traffic flow, cause large errors in the traffic flow forecasting results. This article suggests a solution to the above mentioned issues and proposes a fully connected time-gated neural network based on wavelet reconstruction (WT-FCTGN). To eliminate the potential noise and strengthen the potential traffic trend in the data, we adopt the methods of wavelet reconstruction and periodic data introduction to preprocess the data. The model introduces fully connected time-series blocks to model all the information including time sequence information and fluctuation information in the flow of traffic, and establishes the time gate block to comprehend the periodic characteristics of the flow of traffic and predict its flow. The performance of the WT-FCTGN model is validated on the public PeMS data set. The experimental results show that the WT-FCTGN model has higher accuracy, and its mean absolute error (MAE), mean absolute percentage error (MAPE) and root mean square error (RMSE) are obviously lower than those of the other algorithms. The robust experimental results prove that the WT-FCTGN model has good anti-noise ability.

Keywords: traffic flow modeling, time-series, wavelet reconstruction

PACS: 89.75.Fb, 45.70.Vn, 05.45.Tp, 95.75.Wx **DOI:** 10.1088/1674-1056/ad3349

1. Introduction

Intelligent traffic systems (ITS) are increasingly valued in intelligent cities, which combine resources from the road and non-road sectors using information technology and communication technology,[1] and provide real-time road infrastructure analysis and more effective road traffic control.[2] Expressways are a crucial component of ITS[3] because they have features including closed scenes, quick speeds and heavy traffic flow.[4] Once congestion occurs, it will cause great economic loss, so the forecasting of short-term flow of expressway traffic becomes fundamental research content in ITS.[5]

Information from various sources, including abrupt sensor failure and interference from aberrant weather, can easily compromise the original data of expressway traffic flow. These interferences introduce significant noise to the collected traffic data, thereby affecting the accuracy of traffic flow forecasting. Denoising the traffic flow data is a crucial step.[6] Commonly employed techniques for denoising traffic data include wavelet transform, empirical mode decomposition,[7] etc, as they can simultaneously consider changes in both frequency and time domain. Many studies use wavelet to handle noisy and non-stationary traffic data. However, relying on the components of high and low frequencies obtained through wavelet decomposition as input features can result in partial information loss.[8] Models based solely on single-wavelet denoising technology struggle to eliminate noise introduced by various factors, such as meteorological conditions and sensor failure. Therefore, it is essential to combine multiple wavelet denoising methods for data preprocessing in forecasting.[9] Since traffic data also exhibits periodicity, such as the variation in flow during morning and evening rush hours, introducing periodic characteristics during the denoising process allows the model to learn more hidden information. Wavelet transform, which considers changes in both frequency and time domain, is widely used for handling erratic and stagnant traffic data. Various studies have employed wavelet transform to analyze and predict traffic data, utilizing different approaches such as stationary wavelet transform (SWT),[10,11] noise and spatio-temporal correlation consideration,[12] and breakdown into high- and low-frequency components.[13,14] The hybrid wavelet basis reconstruction method proposed by Liao *et al.*[9] represents a promising research direction, but it primarily focuses on data denoising and does not consider incorporating additional information, such as periodicity, to enhance the data.

Traditional modeling methods, limited in capturing potential trends and fitting high-dimensional nonlinear data, have prompted a shift towards deep learning techniques for en-

†Corresponding author. E-mail: chengguangliu@csu.edu.cn
‡Corresponding author. E-mail: 234703024@csu.edu.cn

© 2024 Chinese Physical Society and IOP Publishing Ltd

hanced traffic flow forecasting accuracy.[15–21] Recurrent neural network (RNN) models, though inherently adept at predicting continuous data, face challenges with long-term dependencies. Therefore, researchers often opt for advanced RNN models such as long short-term memory (LSTM) and gated recurrent unit (GRU) to anticipate traffic flow.[22–31] Recognizing the impact of spatial information on traffic data, convolutional neural network (CNN) models are frequently employed to extract roadway spatial structures. Hybrid forecasting using both CNN and RNN models is a prevalent approach for improved accuracy.[32–36] The second category combines time series and structured analysis data, processing spatial features with graph neural network (GNN), and generating temporal forecasts with LSTM.[37,38] However, the reusability and scalability of these methods are limited due to their specific requirements for spatial and temporal data. We propose a novel traffic flow forecasting model as a solution to the above mentioned issues, and the model is called the WT-FCTGN (wavelet-enhanced fully connected time-gated neural network). Here are our main contributions:

We provide a preprocessing technique based on periodic data reconstruction and wavelet decomposition. The technique uses wavelet analysis to break down the temporal series data into the components with various frequencies, then combines various soft as well as hard thresholds, layers of decomposition, and wavelet basis functions to obtain the best denoising solution. It then expands one-dimensional data after reconstruction, with cycle time data added, to three-dimensional data through wavelet reconstruction, to strengthen the cycle information of the traffic data.

We suggest the WT-FCTGN, a traffic state forecasting model based on wavelet augmentation, predicting the expressway traffic flow through a time series block of a fully connected layer and time-gated mechanism, which further improves the learning ability of the model through model stacking.

Research on the actual PeMS data set of California freeways demonstrates that our model is more accurate than the baseline model, performs best in all evaluation indices and has strong anti-interference properties.

The remainder of this article is structured as follows. Section 2 provides a detailed introduction to the wavelet-based wavelet decomposition as well as reconstruction, fully connected time series block and time-gated block, all of which are contained in neural network forecasting of traffic flow. In Section 3, we use the PeMS data set to test our method and compare the forecasting performance with the prevalent deep learning techniques. At the same time, we also use noise to test the robustness of the model. In Section 4, we provide a summary of our findings and suggest areas for additional study. In Section 5, we discuss the application of our approach to physics and to solving transportation problems in China.

2. Methods

Considering that the noise of data of transportation flow causes interference with the model accuracy and spatial traffic data, it is easy to miss data. Therefore, in this study, the time series information in traffic flow is primarily studied, and the WT-FCTGN forecasting model of fast-speed flow of traffic is proposed. This model combines wavelet reconstruction for data denoising and data dimensionality augmentation, as well as integrating the fully connected time series block and time-gated mechanism. The model is primarily broken down into two fundamental stages, namely, WT-FCTGN model forecasting and development of traffic flow characteristic engineering. Figure 1 depicts the forecasting paradigm of our proposed fast-speed flow of traffic.

The first step: feature engineering of traffic flow data. As traffic flow data are a discrete set of information, we select a discrete wavelet to denoise the data while still matching the noise in the data. The low- and high-frequency components are rebuilt once the wavelet decomposed components have been thresholded. We would want to combine the data denoised through various wavelet functions while keeping the original data, expand the statistics from 1D to 2D, and fuse the processed time data with the reconstructed data to form a three-dimensional tensor to maintain the characteristics of the traffic trend, improve its weight and prevent the scenario that some non-noise data could be removed after wavelet denoising.

The second step: model building and forecasting. We take the 3D data obtained from feature engineering as input, combine the fully connected time series block and time-gated mechanism to construct a forecasting model of traffic flow to predict the trend of high-speed flow of traffic in the next 15 minutes. We consider the model's structure, including the parameters for the input layer, the time gate and the model stack layer settings. Different combinations of these model parameters were verified to choose the model with the highest forecasting accuracy.

2.1. Traffic data wavelet enhancement method

We suggest a technique based on wavelet decomposition noise removal and time-period augmentation reconstruction, which breaks down data from time series into components with various frequencies using wavelet analysis and reconstructs them to remove the high-frequency component noise after decomposition. We select the most effective combination as the processing technique by examining the effects of hard and soft threshold, decomposition layer, and wavelet basis function on the effect of denoising. Based on the above wavelet analysis method, we extend the reconstructed one-dimensional data and the cycle time data to three-dimensional data by wavelet reconstruction, and denoise and enhance the original traffic data.

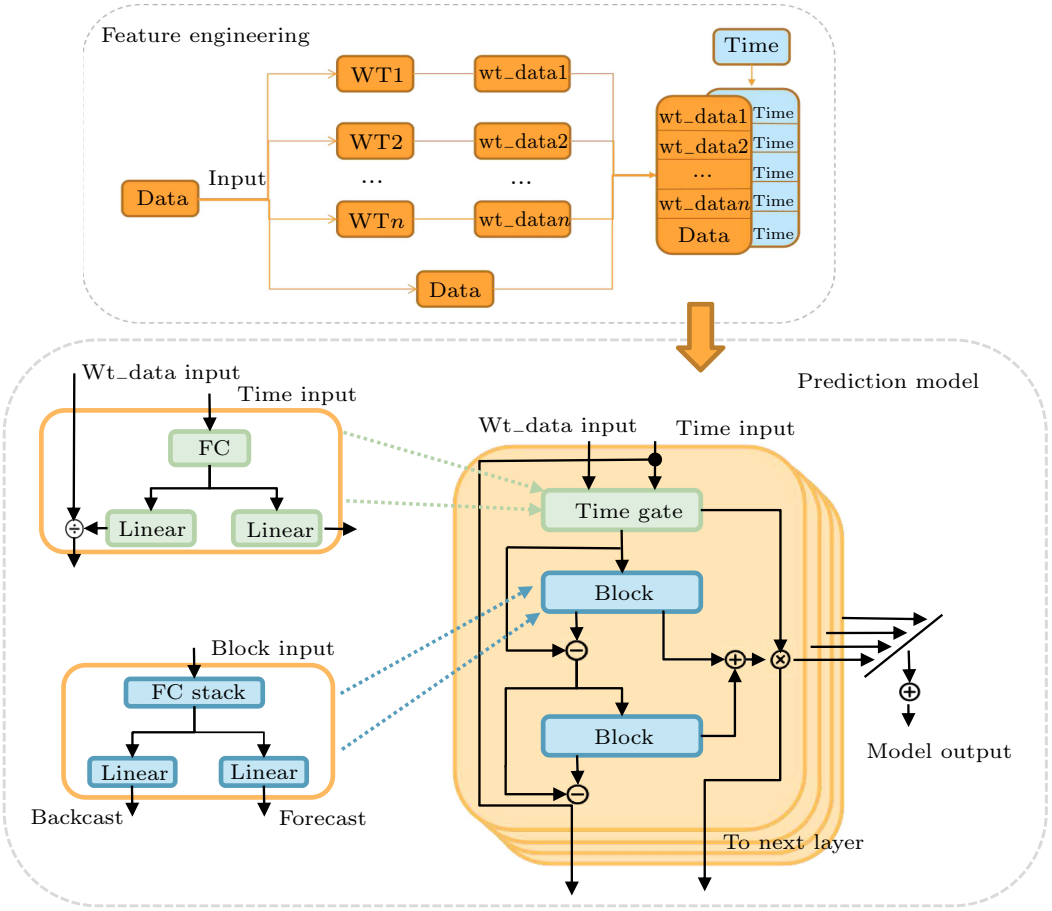


Fig. 1. Overall structure of WT-FCTGN model.

2.1.1. Data noise removal

The flow of traffic is a non-stationary temporal series that exhibits periodic features such as daily morning and evening peaks, and the flow of traffic on weekends of each week is obviously different from that on weekdays (each cycle has similarities). The flow of traffic is also concurrently impacted by numerous factors. It is impacted by the weather, traffic, human factors and other variables. In addition, the noise made during data collection and processing will also have an impact on the data. Since eliminating noise is the major goal of traffic flow forecasting, this study focuses primarily on the effect of noise on traffic state forecasting. Noise is mostly represented by data outliers, and these oddities will cause the forecasting model's performance to decline.

Wavelet transform can detect short-term traffic flow features and traffic data at various scales (such as spikes, sudden variation, and recurring patterns). Signal energy may also concentrate on a few large wavelet coefficients in the wavelet domain due to the wavelet transform's significant data decorrelation. However, the entire wavelet domain contains all the noise energy. As a result, following wavelet decomposition, the magnitude of the signal's wavelet coefficients is greater than that of the noise. The noise information in traffic data can be filtered out using wavelet coefficients with large amplitudes since they are primarily signals while wavelet coefficients having low amplitude are mostly noise. It is demonstrated in Fig. 2 that the components of the traffic flow data (locally) decomposed by a level = 3 wavelet, where cA3 is the low-frequency component and cD1, cD2 and cD3 are the high-frequency components.

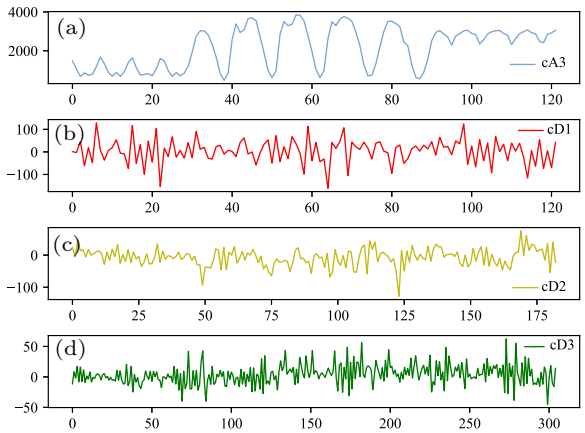


Fig. 2. Schematic diagram of wavelet decomposition of traffic flow data (level = 3).

The specific processing steps for wavelet denoising are as follows. A threshold amount is chosen, and wavelet decomposition is performed on the noisy signal upon every scale. Wavelet coefficients with magnitude greater than the

threshold amount are either totally retained or undergo corresponding contraction, whereas wavelet coefficients with amplitude smaller than the threshold value are set at 0. The processed wavelet coefficients are then used to recreate the denoised signal using the inverted wavelet transform. The higher-frequency components conceal noise. The noise reduction process aims to threshold high-frequency vectors and reconstruct the traffic data flow close to the original data. Therefore, in traffic flow wavelet denoising, wavelet base selection, threshold processing and decomposition layer setting should be carried out.

(1) Selection of wavelet basis function

Various functions of wavelet basis must be chosen for various types of data, and their definitions are

$$\Psi_{a,b} = \frac{1}{\sqrt{|a|}} \Psi \left(\frac{t-b}{a} \right). \tag{1}$$

The wavelet basis function $\Psi(x)$ exhibits a visible fluctuation that deviates from the horizontal axis only at $t = b$, and the function value will quickly decay to zero distance from $t = b$, allowing one to ascertain the corresponding frequency at time b, where a is the scale, b is the translation and t is the time point. In the experiment, the forecasting results of traffic flow data decomposed and reconstructed by various wavelet basis functions are compared, and the final fusion of wavelet basis functions is determined.

(2) Threshold processing method

The threshold function, which reflects the various wavelet coefficient processing schemes above and below the threshold, is the essential component of threshold noise removal. The data curve will be flatter following soft threshold processing during the denoising process. However, some detail will be missed and the signal-to-noise proportion will be poor. Strong thresholding can preserve more information and the local properties of the traffic data curve, yet it is simple to generate a local jitter. We discuss the effects of both hard and soft threshold reconstructions on traffic forecasting for the threshold processing method of the high-frequency wavelet component.

(3) Decomposition layer setting

At various scales, noise that might be present in traffic data must be analyzed and removed. Diverse analytical scales correspond to varied decomposition layers. Different scales are used by each decomposition layer to measure the original signal. When the number of layers increases, the measurement scale improves. Figure 3 depicts the wavelet breakdown and reconstruction of three layers.

As a result, there are N layers of high-frequency elements and one layer of low-frequency elements due to the original data being divided N times. This is known as the number of decomposing layers, or N. In the figure, D_1, D_2 and D_3 represent high-frequency signals containing potential noise, A_3 represents signals with lower frequency, i.e., possible trends of the flow of traffic, and A_1 and A_2 represent the reconstructed data. The impact of various decomposition layers mostly on noise reduction of traffic data was investigated through experiments.

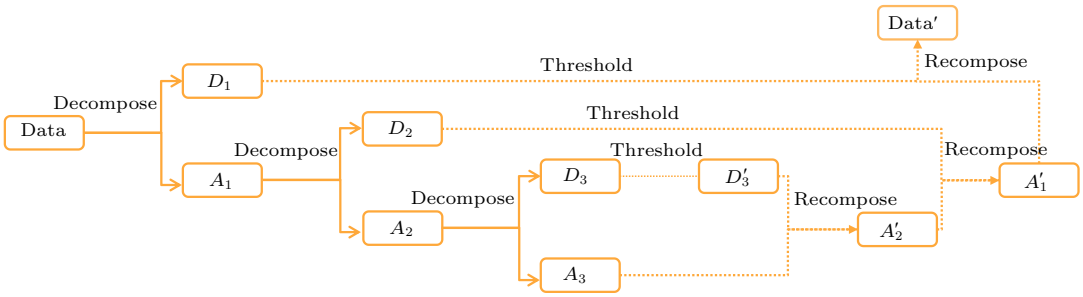


Fig. 3. Wavelet decomposition and reconstruction with three layers.

2.1.2. Introduction of cycle time data

To better mine the temporal message in the flow of traffic, this article introduces cycle temporal data, namely, the timestamp information of traffic flow data, as separate external information, and designs a specific time gate block to process it. In particular, time periodicity affects the information of the factor of traffic flow for the identical route in a single day. For instance, the morning peak traffic flow is very strong, and the traffic approaches zero at around 2:00 a.m., and the periodic time data can better reflect the periodicity in the traffic information than the original traffic data. To further mine the temporal periodicity information, we introduce the traffic flow time stamp and process it as the proportion of the current time in a day, and the processed time data is distributed between 0 and 1.

2.1.3. Data reconstruction

This paper uses multiple wavelet functions for denoising because wavelet analysis may inadvertently remove non-noise volatility at the micro scale. It also aggregates the denoised data and the original data, keeping all the initial data's details while also adding its denoised data, ensuring that the forecasting method can simultaneously learn the original data and the denoised information. The processed time data are combined

with the source data as well as the denoised data.

This study chooses many branches of wavelet basis functions based on the wavelet transform research mentioned above. Three levels of decomposition are chosen, and hard threshold rebuilding methods are chosen for the threshold. Figure 4 depicts the approach for promoting data dimensions suggested in this section. Wavelet decomposes the source data and reconstructs it four times, and the time information after processing is merged into the reconstructed data and original data, respectively, to form a two-dimensional tensor, and four groups of reconstructed data and original data are integrated as the model input. The construction process of the whole 3D time series tensor is shown in Algorithm 1.

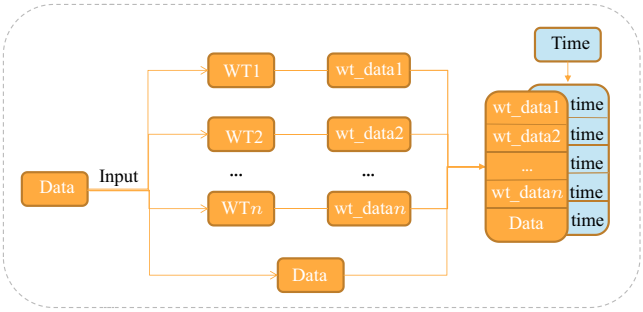


Fig. 4. Schematic diagram of data dimension promotion.

Algorithm 1 3D tensor reconstruction algorithm

Input: Traffic flow data $\boldsymbol{Data}$; Time stamp data $\boldsymbol{Time}$; Decomposition layers $\boldsymbol{L}$; Basis function $\boldsymbol{wti} \in [\boldsymbol{wt1}, \ldots, \boldsymbol{wtn}]$

Output: Reconstructed data $\boldsymbol{wt_data}$

```
wt_data = [ ]
for wt in [wt1, ..., wtn]:
    A_0 = data
    for i in range(L) do
        D_i A_i = decompose(A_{i-1}, wt)
    end for
    X = D_1 + D_2 + D_3 + ... + D_L + A_L
    for i in range(L) do
        if mode == "soft" then
            threshold(D_i, "soft")
        else if mode == "hard" then
            threshold(D_i, "hard")
        end if
    end for
    data' = reconstruct(X, wt)
    data' = reconstruct(data', time)
    wt_data.append(data')
end for
data = reconstruct(data, time)
wt_data.append(data)
return wt_data
```

Based on the above process, a three-dimensional traffic flow data tensor is constructed as an input. Thus, the input tensor of the forecasting model of the flow of traffic can be represented by

$$\boldsymbol{X} = (x_1, x_2, \ldots, x_n). \tag{2}$$

Here $\boldsymbol{X}$ represents the three-dimensional tensor input to the forecasting model, and the tensor dimension is $n \times j \times 2$, where n is the time step, j represents the amount of data after wavelet reconstruction and the original data. We set j as 5, and 2 stands for the data of the flow of traffic and the time stamp, respectively. The input tensor $\boldsymbol{X}$ comprises all the general traffic data from the previous n time points, reflects changes in flow of traffic on the road segment from time $t-n$ to time t, and has strong time correlation, which means better forecasting performance of the model.

2.1.4. Model robustness

In this article, the robustness of the forecasting model of the flow of traffic can be defined as follows.

Definition 1 A model with strong robustness allows errors in the data in the training set, and its forecasting performance is less affected by erroneous data. The model only changes slightly due to tiny training set variances, robustness is defined as

$$f(x+\Delta\delta) \approx f(x), \tag{3}$$

where $f(x)$ is the forecasting model of the flow of traffic, and $\Delta\delta$ represents the amount of variation from the real data caused by noise in the data or adversarial data generated by an attack.

A robust model should give forecasts that are close to the correct data in environments where the amount of variation is below a certain threshold. After defining the robustness, this article tries to give the boundary definition of model robustness, which is defined as follows.

Definition 2 Suppose a sample point of traffic flow data has multiple disturbances added. In the local space of a disturbance, the model can give approximately consistent forecasting results. If it is considered that this part of local space is the robustness space of the model, then the space edge is the model robustness boundary. The model forecasting of the traffic flow in this range is credible.

According to Definition 2, this paper proposes to add Gaussian noise data into the training data as adversarial training to test the robustness of the model and determine its robustness boundary. The process and verification method of adversarial training are described in detail below, and the addition of Gaussian noise is given by

$$\text{noise data} = \text{data} + \text{mean} + \sigma \cdot X, \tag{4}$$

where the noise data is the data after adding noise, the data is the original data, the mean is the mean of Gaussian noise, σ is the variance of Gaussian noise and X is a random value that follows $N(0,1)$ distribution.

Positive and negative symmetric mean values are added to raw data, and multiple groups of variances from small to large

are used to construct noise data, which are used as a comparison data set for training a model. Real data are used as labels, and both a test set and a test label use real data to determine whether the model generated and trained on the noise data set has robustness.

2.2. Network model design

N-BEATS[39] explores the pure deep learning model structure inside the context of forecasting of temporal series, and the N-BEATS model is similar to the "sequence decomposition" method in traditional predictive machine learning time series forecasting, and the output results are interpretable. Therefore, based on the fully connected time series block on the N-BEATS model, this paper proposes the FCTGN model, as demonstrated in Fig. 5. It includes a fully connected temporal series block for learning the global information of the flow of traffic and a time gate systematic block for comprehending the periodic information of the flow of traffic. First, this specific model removes the g function of the N-BEATS fully connected sequence block to obtain the results predicted by the fully connected layer, and adds a time gate mechanism to capture the traffic flow cycle information. According to the idea of sequence decomposition, FCTGN is stackable, with which more potential information in traffic flow can be learned. Below, the specific content of each block of the model will be described separately.

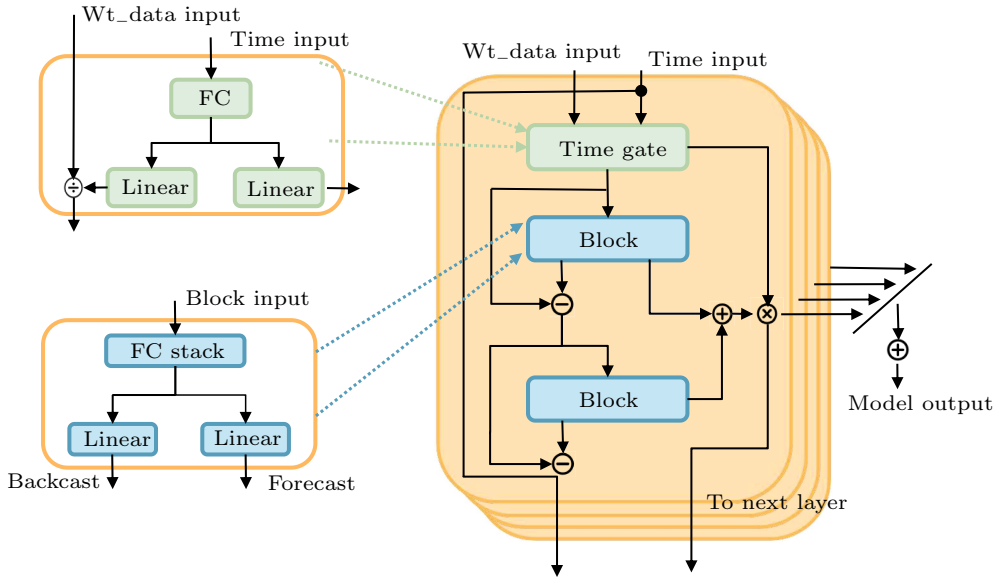


Fig. 5. FCTGN overall model.

2.2.1. Fully connected time series block

The fully connected time series block is used for learning global information modeling in the traffic flow information, including time series period information and fluctuation mutation information in the flow of the traffic. Compared with the traditional LSTM structure, the fully connected time series block does not have a complex gated recurrent unit, and the residual error block formed by the fully connected layer has higher training efficiency. As shown in Fig. 6, the fully connected time series block is composed of r residual blocks, and each residual block has l fully connected layers. The fully connected layers are shown as

$$FC_{r,l}\left(H^{r,l-1}\right) \equiv \mathrm{RELU}\left(A^{r,l}H^{r,l-1}+b^{r,l}\right), \tag{5}$$

where $A^{r,l}$ and $b^{r,l}$ denote the weighting and biases of layer l and layer r residual block, and the input of the fully connected time series block consists of the information input from the previous layer.

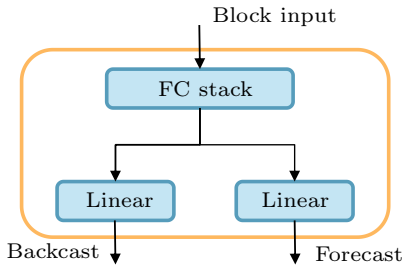


Fig. 6. FCTGN fully connected block.

The fully connected time series block is

$$\begin{aligned} &Z^r = \mathrm{RELU}\left(Z^{r-1}-\hat{Z}^{r-1}\right), \\ &H^{r,1} = FC_{r,1}\left(Z^r\right), \ldots, H^{r,L} = FC_{r,L}\left(H^{r,L-1}\right), \\ &\hat{Y}^r = \left(H^{r,L}\right)^{\mathrm{T}} F^r. \end{aligned} \tag{6}$$

Layer 0 residual block values are $\hat{Z}^0 \equiv 0$, $Z^0 \equiv Z$. Z represents the embedding belonging to the previous layer input information and time gate characteristic, and the initial layer is the direct input information. In the above equation, $F^r \in \mathbb{R}^{d_h \times H}$,

the parameter F is to maintain the same dimension as the predicted value Y. The result of the fully connected time series block is the sum $\hat{Y} = \sum_r \hat{Y}^r$ of the predicted values of each residual block.

2.2.2. Time gate block

The time gate block transforms the properties of the time variable (such as time of day, day of week, etc.) that occur with the node observations into a model. As shown in Fig. 7, this is the first block of each layer, which aims to learn the temporal-periodic features of the flow of traffic. After wavelet processing, the data of the flow of traffic passes through two different linear layers. The final output is one layer, and that of the next block is the other. Finally, the result of the next block will be divided or multiplied by the variable feature to harness the time-period traffic flow information, wherein the variable feature contains the information of the current time in one day. Furthermore, we decouple the temporal effects of input and output by isolated layers of linear projection, since the overall input and output amount of time varies.

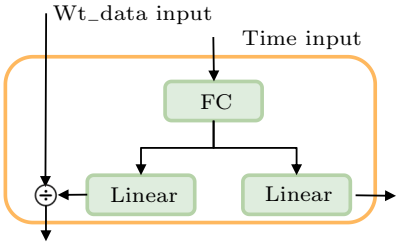


Fig. 7. FCTGN time gate block.

The following three ideas form the foundation for stacking FCTGN layers. The sum of the forecasts from the preceding layer is the first thing the subsequent layer accepts as input. Second, every FCTGN layer has a unique time gate system. Thus, each layer can set the freedom of the information flow of the next layer based on the processing that has been done by the layer before. For instance, if the layer ahead of any other layers of traffic flow learns the characteristics of the first five minutes, the next layer of FCTGN will pay more attention to the information of other moments. The mean of the layer forecasts represents the final model result.

3. Experimental output and analysis

3.1. Setup of the data set

This study makes extensive use of a large-scale traffic database gathered in California, the United States, by the performance evaluation system (PeMS) of the California Transportation Management Agency (Caltrans). Every five minutes, a set of data is summarized from the data set that is analyzed and autonomously gathered by the inductor loop sensor for a sensor collection period of 30 seconds. The PeMS data set has numerous highways since there are separate linkages between the various highway directions. We adopted the method of Liao *et al.*,[9] and selected the cluster center of high-grade traffic with the most obvious fluctuation and the most characteristic, the direction N of California Highway 880 as the research object. The predicting intervals of freeways should be set at 15 minutes, according to published works, and finally we choose to conduct feature engineering validity, model structure validity, model comparison and robustness experiments on actual traffic flow data from January to March 2019, which has a total of 8640 time points. In the experiment, the training data sets for the forecasting model's training as well as validation are January and February, while the testing data sets for the final model predicting result are set as March.

3.2. Evaluation index and experimental conditions

To quantitatively assess the effectiveness of the traffic flow forecasts, this article uses mean absolute error (MAE), mean absolute percentage error (MAPE) and root mean square error (RMSE) as forecasting indexes, and the indicators are defined as

$$\mathrm{MAE} = \frac{1}{n}\sum_{i=1}^{n}\left|f_i - \hat{f}_i\right|, \tag{7}$$

$$\mathrm{MAPE} = \frac{1}{n}\sum_{i=1}^{n}\frac{\left|f_i - \hat{f}_i\right|}{f_i}, \tag{8}$$

$$\mathrm{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(f_i - \hat{f}_i\right)^2}, \tag{9}$$

where n is the ultimate number of samples for verification, f_i is the actual flow of traffic in the set of data, the model predicts the flow of traffic as $\hat{f}_i$, and f_i and $\hat{f}_i$ are expressed in units of vehicles/15 min. The anticipated data of traffic flow is more likely to reflect reality if the MAE, MAPE and RMSE are smaller, and otherwise if the predicted findings are very different from reality.

3.3. Feature engineering validity test

The goal of the experiment in this part is to choose the optimal wavelet reconstructing variable and demonstrate its efficacy. The most appropriate wavelet reconstruction technique is investigated using three different parts of the experiment: wavelet basis function selection, wavelet threshold rebuilding technique and decomposition layers. The FCTGN model framework is used to conduct the study of wavelet basis function and the succeeding research project of contrasting both hard and soft thresholds and decomposition layers in the phase of deciding on a single-wavelet basis function since the traffic flow statistics are separate 1D data. Since the source traffic flow data set is independent, discrete wavelet is chosen for comparison purposes using the wavelet basis function. Specified input wavelet and discrete wavelet group are two categories of discrete wavelets. Several functions of the identical

type are included in the discrete wavelet group, including sym group (including sym2, sym3, etc).

The first function in the function group of the wavelet family and a singular function of wavelet basis are selected for data noise reduction comparison experiments. The noise reduction effect of the 1D function of wavelet basis is already demonstrated in Table 1. Coif and sym categories of wavelet functions have greater and relatively close forecasting accuracy when compared to the other seven types of basis functions that can handle discrete data since these two major types of wavelet functions are more adept at handling data with symmetric attributes. This is because the underlying tendency of traffic flow involves regularity and has a symmetric structure in the broad sense. Sym also has the highest forecasting performance. As a result, according to the MAPE standard, coif and sym both have positive impacts.

Table 1. Noise reduction effect of one-dimensional wavelet basis function.

Wavelet basis function	MAE	MAPE (%)
haar	20.94	2.57
db1	20.91	2.62
coif1	20.92	2.56
bior1.1	21.03	2.57
rbio1.1	20.97	2.60
dmey	20.92	2.58
sym2	20.86	2.55

According to the above experimental outputs and research by Tang,[40] this section selects the first four functions in the three wavelet categories of coif, sym and db, as the combination, and the combination of four single-wavelet functions. Therefore, this section proposes four wavelet function combinations for comparative testing. It is demonstrated in Table 2 that the experimental results are clear. The forecasting effect of various functions of that same family (such as 'db1', 'db2', 'db3', 'db 4') and the forecasting effect of functions using different families are compared. The outcome demonstrates that the final mode has the greatest effect, which is consistent with the fact that the db waveform proposed by Tang *et al.*[40] performs well in terms of fitting the change mode of data of the flow of traffic.

Table 2. Combination effect of wavelet basis functions.

Combination of wavelet basis functions	MAE	MAPE (%)
'sym2', 'sym3', 'sym4', 'sym5'	20.90	2.56
'coif1', 'coif2', 'coif3', 'coif4'	20.80	2.55
'coif1', 'db1', 'sym2', 'dmey'	20.85	2.58
'db1', 'db2', 'db3', 'db4'	20.67	2.52

It is vital to establish the quantity of decomposition layers through trials since the layer count is a crucial kernel element that, to some extent, defines the intricacy of feature engineering. Some researchers use only three layers for comparison in the wavelet reconstruction stage, and draw the conclusion that three layers have the best effect. Because the number of layers of wavelet decomposition has a great influence on the ultimate forecasting result, this section compares the influence of 2–5 wavelet decomposition layers on the noise reduction effect.

Table 3 illustrates the experimental outcome. Under the mixed wavelet combination and db wavelet combination, the three-layer wavelet reconstruction method in FCTGN has the best effect, followed by two-layer and five-layer reconstructions. Considering that the greater the number of layers, the more complex the structure of the wavelet transform block, and the worse the real-time performance, the three-layer wavelet reconstruction method is selected in the following experiments.

Table 3. Comparison of forecasting effects of model stacking layers.

Model	Layers	MAE	MAPE (%)
FCTGN(mix)	level_2	20.82	2.58
	level_3	20.74	2.54
	level_4	21.13	2.57
	level_5	21.01	2.60
FCTGN(db)	level_2	20.81	2.54
	level_3	20.67	2.52
	level_4	21.03	2.58
	level_5	20.87	2.55

Table 4. Comparison of hard and soft thresholds.

Model	Threshold	Decomposition layers	MAE	MAPE (%)
FCTGN (db)	soft	2	20.95	2.67
		3	22.95	2.56
	hard	2	20.81	2.54
		3	**20.67**	**2.52**

The right threshold processing technique is another crucial impacting aspect in the wavelet de-noising procedure. As a result, the findings of this section's comparison between the soft threshold and the hard threshold are displayed in Table 4. The findings demonstrate that the hard threshold processing is superior to the soft threshold processing for analyzing information on flow of traffic. This is due to the hard threshold processing retaining the peak characteristics of the data while the soft threshold processing smooths them out, making it easier to extract the important difference in daily flow of traffic at specific times.

The key distinction between our method and others is that we analyze the impact of the wavelet basis function and adaptive threshold method on the rebuilding and instead use the rebuilt data as the source of the forecasting model. To deal with distinct noise in data on traffic flows four times, we first apply four different wavelet basis functions. Then, we mix the four groups of data that were recovered using various wavelet basis functions with the source data to obtain the final data. It is akin to giving the fundamental tendency in the source data more weight since every measurement includes it, emphasizing the significance of the fundamental tendency in the data on traffic flows.

3.4. Model structure validity test

We conduct studies to confirm the efficiency of the time gate method and that WT-FCTGN is stackable. We select the data processed by wavelet combination db for model training and forecasting comparison. We select 1–5 layers of model stacking. Table 5 illustrates the experimental results of MAPE and MAE of WT-FCTGN with different stacking layers. It is clearly shown that the model performance is gradually improved with the number of layers increasing from one to four of the model stack, and the model can mine more information in the traffic flow with the increasing number of layers. However, this model cannot be stacked all the time, and the performance of the model is degraded after four layers. We think that this is because the model overlearns some special features in the data and ignores the generality, so the WT-FCTGN model is better when there are four stacked layers.

Table 5. WT-FCTGN model stacking experiments.

Wavelet basis function	MAE	MAPE (%)	Level & mode	Stack
db_fam	20.94	2.60	3 & hard	1
db_fam	20.81	2.58	3 & hard	2
db_fam	20.78	2.58	3 & hard	3
db_fam	20.65	2.52	3 & hard	4
db_fam	20.83	2.58	3 & hard	5

To test the time-gate system's performance, we remove the time gate block of WT-FCTGN while training data and removing the time data including only wavelet de-noised data. We compare the processed model with WT-FCTGN to test the time-gate block's performance. Table 6 presents the outcomes. As can be seen, the MAE and MAPE of WT-FCTGN are increased by 1.29% and 0.24% when compared to FCTGN without a gate mechanism, demonstrating the effectiveness of our time gate mechanism in enhancing prediction accuracy and enabling the model to collect more time information in flow of traffic. Simultaneously, we compare the FCTGN without a gate system with the WT-FCTGN without a gate mechanism after wavelet processing, and it can be seen that the MAPE and MAE models are improved by 0.38% and 0.02%, which again proves the effectiveness of wavelet denoising.

Table 6. Comparison between the model with time gate removed and the original model.

Model	MAE	MAPE (%)	Wavelet basis function
FCTGN (without time gate)	21.94	2.76	no
WT-FCTGN (without time gate)	21.56	2.74	db_fam
WT-FCTGN	20.65	2.52	db_fam

3.5. Performance comparison test on public data sets

In this section, traffic statistics from PeMS are utilized to assess the efficacy of the suggested WT-FCTGN model by contrasting the developed framework with many other mainline forecasting deep architecture techniques. This comparison serves to validate the predictive ability of the present model on public data sets and the cost of model deployment, which is reflected in the period of model iteration in the experiments. All experiments utilize a single graphics card Nvidia 2080Ti 11 GB — for training to maximize the efficiency of the experimental time. Tensorflow as well as Keras frameworks are the foundation on which all neural network techniques are developed, trained and verified.

We introduce the following model as a baseline on the data set to compare the prediction performance with WT-FCTGN:

RNN: Recurrent neural network with sequence data as input, recursive in the evolutionary direction of the sequence and all nodes (recurrent units) connected in a chain.

LSTM: The structural principle of LSTM is the use of specialized memory cells with gating mechanisms to regulate the flow of information, allowing it to selectively remember or forget information over long sequences.

GRU: GRU (gated recurrent unit) is the use of a simplified gating mechanism that combines the functions of input and forget gates in LSTM, resulting in a more streamlined architecture.

WT-Models: LSTM\RNN\GRU neural network model enhanced with wavelet noise reduction.

N-Beats: N-BEATS (neural basis expansion analysis for interpretable time series forecasting) is a deep learning architecture for time series forecasting that utilizes a stack of fully connected layers to model and predict future values in a time series.

EMD-FCTGN: Fully connected time-gated neural network model enhanced with empirical modal decomposition.

RevIN-FCTGN: Fully connected time-gated neural network model enhanced with reversible instance normalization, RevIN consists of two different steps, normalization and non-normalization. The former normalizes the input to fix its distribution based on the mean and variance, while the latter returns the output to the original distribution.

The specific experimental results of the model forecasting efficacy in this section at various temporal steps are shown in Fig. 8. The experiment proves that the effect is the best when the time step is one day. Finally, this section contrasts the prediction ability of the two proposed models in this section with the current mainstream traffic flow forecasting model on the public data set PeMS. The training process of WT-FCTGN with wavelet combination of 'dbfam' and four stacking layers is shown in Fig. 9, and the curve comparison between the model predicted values and real labels is shown in Fig. 10, and the specific experimental outcomes are illustrated in Table 7. This section not only compares GRU, LSTM and RNN models, but at the same time obtains WT-LSTM, WT-GRU and WT-RNN models through processing using wavelet of the above three models. Liao *et al.*[9] proposed WT-1DCNN and

WT-2DCNN models based on wavelet denoising. The performance between the above model and the proposed WT-FCTGN is compared.

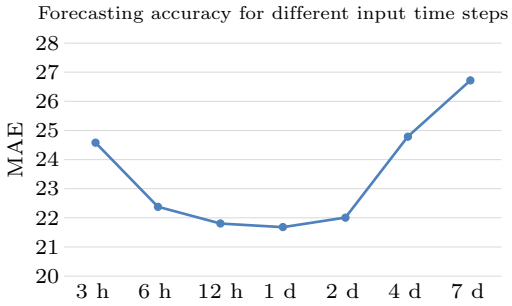


Fig. 8. Forecasting accuracy for different input time steps.

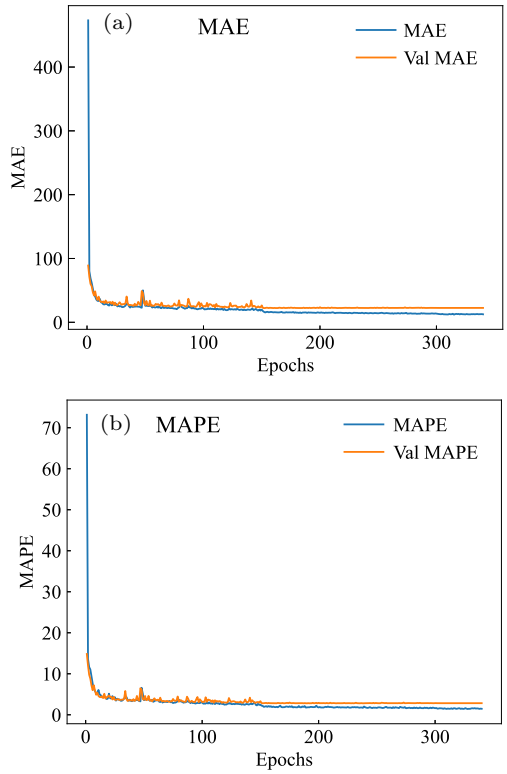


Fig. 9. Change process of indicators of the training set and validation set in model training.

As shown in Table 7, FCTGN decreased by 7.09, 8.38, 8.31 and 0.88 on MAE compared with RNN, LSTM, GRU and N-beats, respectively, MAPE decreased by 0.96%, 1.14%, 0.98% and 0.31%, and RMSE decreased by 5.30, 6.71, 5.82 and 1.44, respectively; WT-FCTGN decreased by 7.87, 7.61, 8.01, 2.67 and 1.81 on MAE compared with WT-RNN, WT-LSTM, WT-GRU, WT-1DCNN and WT-2DCNN, respectively, MAPE decreased by 1.06%, 1.06%, 1.00%, 0.44% and 0.28%, and RMSE decreased by 4.53, 6.13, 5.14, 3.36 and 2.23, respectively. From the above results, it can be seen that FCTGN exhibits more than 20% improvement in all the metrics compared to the traditional recurrent neural network, and shows better performance compared to N-Beats. We experimented with all mainstream recurrent neural networks using wavelet enhancement and found that the performance of these recurrent neural network models using wavelet enhancement does not decrease but increases in some of the metrics. This is because they do not have a structure that specializes in wavelet data, and they are unable to learn effectively from wavelet reconstructed data. Compared with the neural network models WT-1DCNN and WT-2DCNN, which are specialized in learning wavelet data, FCTGN performs better, indicating that our model structure can better learn the information after wavelet enhancement. Meanwhile, for other data enhancement and noise reduction methods, WT-FCTGN compared to EMD-FCTGN and RevIN-FCTGN decreased by 4.47 and 1.06 in MAE metrics and 0.63% and 0.18% in MAPE metrics, and the FCTGN model proposed here has achieved greater ability over mainstream models without wavelet denoising.

Table 7. Comparison results of mainstream model forecasting.

Models	MAE	MAPE (%)	RMSE
RNN[21]	27.83	3.49	35.84
LSTM[23]	29.12	3.67	37.25
GRU[22]	29.05	3.51	36.36
WT-RNN	28.52	3.58	34.85
WT-LSTM	28.26	3.58	36.45
WT-GRU	28.66	3.52	35.46
WT-1DCNN[9]	23.32	2.96	33.68
WT-2DCNN[9]	22.46	2.80	32.55
N-Beats	21.53	2.73	31.98
EMD-FCTGN	25.12	3.15	34.67
RevIN-FCTGN[41]	21.71	2.70	31.23
FCTGN	20.74	2.53	30.54
WT-FCTGN(Layer4)	20.65	2.52	30.32

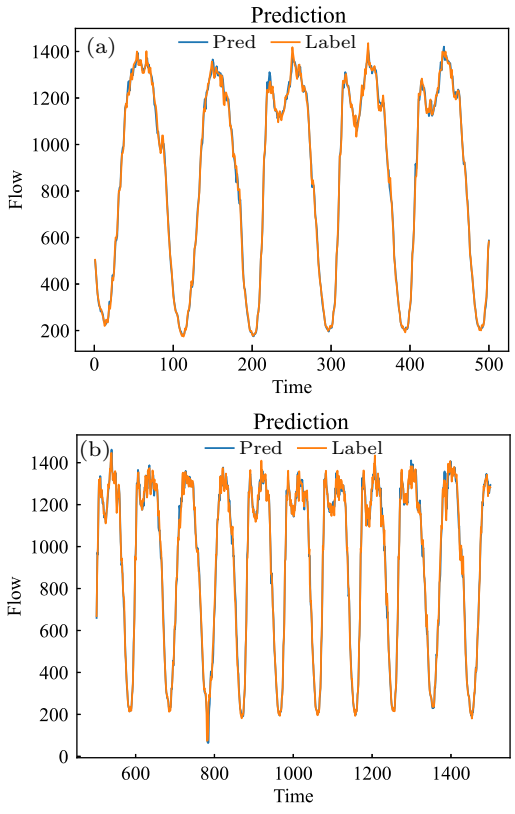


Fig. 10. Model forecasting results.

Furthermore, the model forecasting capability is obviously optimized after wavelet denoising. It is shown that WT-FCTGN exhibits significantly poorer ability than other mainstream models in dealing with traffic flow forecasting tasks.

Meanwhile, WT-GRU, WT-LSTM and WT-RNN models with wavelet rebuilding are added as comparison models in this section. WT-RNN and WT-GRU prediction accuracy is lower in the experiment than the initial formulation, and the effective information in the original data will be lost if the wavelet function alone is used. WT-FCTGN based on 3D time series reconstruction is the best, which demonstrates that the theory behind extending the data dimension by utilizing various wavelet functions is sound.

3.6. Robustness experiment

In this section, Gaussian noise is added to the data of the set for training to simulate data noise interference, where the mean of the noise is -4, -3, -2, -1, 0, 1, 2, 3, 4, and the variance is 1, 2, 3, 4, 5, 6, 7, 8, 9, respectively. A total of 81 groups of noise are compared to test the robustness of the proposed model in this article.

Figures 11(a) and 11(b) show the MAE change in the WT-FCTGN model in the mixed noise environment, and it is shown that the largest amount of the WT-FCTGN model on the MAE is only 21.5. Figures 11(c) and 11(d) show the change in MAPE of the WT-FCTGN model under interference, and the maximum value of MAPE of the WT-FCTGN model is 2.68%. The deviation range of MAE is about 0.7, the deviation range of MAPE is about 0.12, and the overall deviation is less than 5%, which shows that the WT-FCTGN model has good robustness in the face of noisy data, and can cope with more varied scenarios and resist attacks. The above chart shows that when noise variation increases, the WT-FCTGN model's reliability starts to decline, but is not sensitive to the change in mean value. How accurate the model will mainly be influenced by variance, which demonstrates that the translation in the data cannot cause too much impact on the WT-FCTGN model with the time gate mechanism.

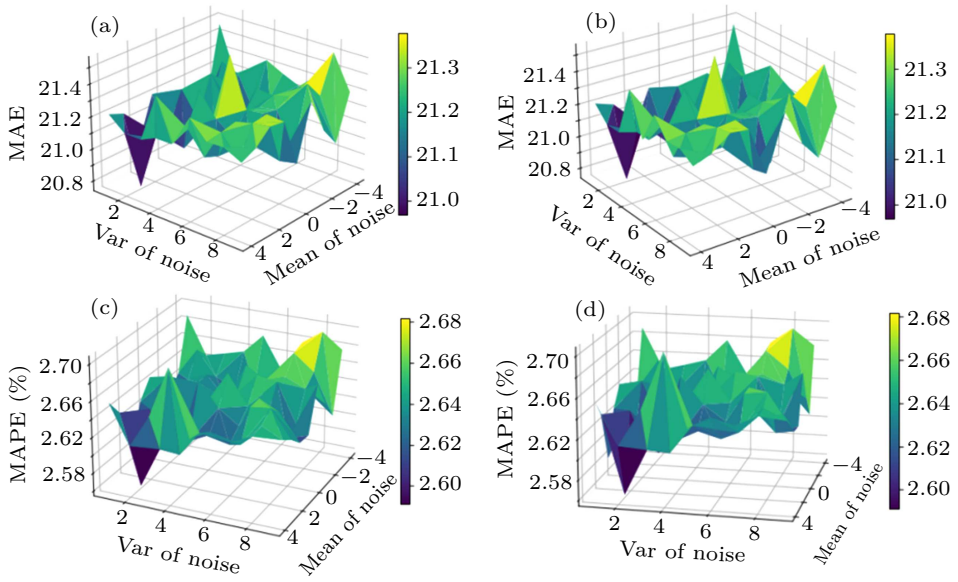


Fig. 11. Model MAE and MAPE of WT-FCTGN in different noise environments.

4. Discussion

Wavelet reconstruction is a signal processing technique that utilizes wavelet transform to eliminate noise from signals. Wavelet transform is a time-frequency analysis method that decomposes signals into components of different frequencies. First, wavelet transform possesses excellent time-frequency localization characteristics, implying that it can provide high-resolution information simultaneously in both time and frequency domains. This makes wavelet transform highly suitable for processing non-stationary signals, i.e., signals with statistical properties that vary over time. In practical applications, many signals exhibit non-stationarity, such as speech and vibration signals. For physical applications, various signal processing procedures are involved, such as spectral signals, seismic wave signals, etc. The WT-FCTGN model proposed in this article integrates wavelet reconstruction as a signal processing technique and artificial intelligence neural network modeling. This model can serve as a reference for modeling other signals in physical applications.

Furthermore, to demonstrate its effectiveness in addressing traffic flow prediction issues in China, we utilized the publicly available TaxiBJ data set in the transportation domain. This data set comprises Beijing taxi data, where the city is divided into a 32×32 grid, with collected data on taxi inflow and outflow in each grid. The comparison between the WT-FCTGN model and the baseline model on the 2016 TaxiBJ

data set for Beijing cab areas (grid location: (16,16)) is shown in Table 8.

As shown in Table 8, it is evident that the WT-FCTGN model consistently outperforms the baseline across the three evaluation metrics on the Beijing taxi data set. Particularly noteworthy is the selection of data augmentation and denoising methods, where wavelet reconstruction demonstrates superior suitability for traffic flow prediction in comparison to empirical mode decomposition (EMD) and reverse instance normalization (RevIN). This not only validates the effectiveness of the proposed approach in this article but also indicates its applicability as a reference for addressing traffic issues in China.

Table 8. Comparative results of mainstream predictive models on the TaxiBJ data set.

Models	MAE	MAPE (%)	RMSE
RNN	14.37	34.52	16.23
LSTM	15.17	35.24	16.86
GRU	15.04	34.76	16.63
WT-1DCNN	10.23	30.52	13.24
WT-2DCNN	10.04	29.41	12.82
N-Beats	9.85	28.73	12.65
EMD-FCTGN	11.49	31.22	14.36
RevIN-FCTGN	9.52	28.15	12.33
FCTGN	8.67	27.73	12.03
WT-FCTGN(Layer4)	8.52	27.32	11.76

5. Conclusion

In this article, a WT-FCTGN model made for predictive modeling of traffic flow is proposed by utilizing a fully connected time-gated modular neural network based on wavelet denoising combination. This research employs the discrete wavelet noise removal technique to enhance the quality of the data by removing various types of traffic noise and reconstructing it. The proposed model can find more detailed aspects of flow of traffic in the forecasting block than earlier deep learning approaches that used RNNs to collect time series data, and we compare the variables of several models while fine-tuning to boost forecast accuracy. Finally, we conduct tests on PeMS data sets to assess the performance of the suggested approach and compare it to other approaches such as LSTM, GRU, WT-RNN, WT-LSTM, WT-GRU, WT-1DCNN and WT-2DCNN (the latter five methods combined with wavelet denoising). The outcomes demonstrate that the suggested strategy performs better than other widely used strategies.

At the same time, we also found some interesting phenomena in the robustness experiment. For example, when the mean value of noise is 4 and the variance is 2, the accuracy of the model is improved. We consider that adding random blank blocks to the training image during image completion will improve the ability of image completion of the model, which provides some inspiration for our future traffic data of feature engineering. For future work, we plan to do more in-depth research on related work. For example, this study utilizes the wavelet denoising approach of feature engineering to denoise data and enhance the data features, but this requires prior knowledge and manual data preprocessing to guide the deep learning model to train. At the same time, the data feature work of the flow of traffic can be designed as a network module and integrated into the model training procedure of deep learning types to obtain more features and improve the overall efficiency of forecasting work.

Acknowledgements

The Science and Technology Research and Development Program Project of China Railway Group Ltd provided funding for this study (Project Nos. 2020-Special-02 and 2021-Special-08). The authors gratefully acknowledge the technical and financial assistance provided by China Railway Group.

References

[1] Guo K, Hu Y, Qian Z, Sun Y, Gao J and Yin B 2020 *IEEE Trans. Intell. Transp. Syst.* **23** 1009
[2] Wei P, Cao Y and Sun D 2013 *Transp. Res. Part B Methodol.* **53** 1
[3] Jiang W and Luo J 2022 *Expert Syst. Appl.* **207** 117921
[4] Zheng C, Fan X, Wang C and Qi J 2020 *Proceedings of the AAAI conference on artificial intelligence* **34** 1234
[5] Yin X, Wu G, Wei J, Shen Y, Qi H and Yin B 2021 *IEEE Trans. Intell. Transp. Syst.* **23** 4927
[6] Liu Y, Liu Z and Jia R 2019 *Transp. Res. Part C Emerg. Technol.* **101** 18
[7] Nai W, Liu L, Wang S and Dong D 2017 *Algorithms* **10** 139
[8] Mousavizadeh Kashi S O and Akbarzadeh M 2019 *J. Intell. Transp. Syst.* **23** 60
[9] Liu Y, Song Y, Zhang Y and Liao Z 2022 *Phys. Stat. Mech. Its Appl.* **603** 127817
[10] Wang Y, Zhang D, Liu Y, Dai B and Lee L H 2019 *Transp. Res. Part C Emerg. Technol.* **99** 144
[11] Wu Y, Tan H, Qin L, Ran B and Jiang Z 2018 *Transp. Res. Part C Emerg. Technol.* **90** 166
[12] Peng Y and Xiang W 2020 *Phys. Stat. Mech. Its Appl.* **549** 123913
[13] Dai G, Ma C and Xu X 2019 *IEEE Access* **7** 143025
[14] Zhang H, Wang X, Cao J, Tang M and Guo Y 2018 *Appl. Intell.* **48** 3827
[15] Vlahogianni E I, Golias J C and Karlaftis M G 2004 *Transp. Rev.* **24** 533
[16] Van Lint J W C and Van Hinsbergen C 2012 *Artif. Intell. Appl. Crit. Transp. Issues* **22** 22
[17] Dudek G 2016 *Electr. Power Syst. Res.* **130** 139
[18] Van Der Voort M, Dougherty M and Watson S 1996 *Transp. Res. Part C Emerg. Technol.* **4** 307
[19] Sun H, Zhang C and Ran B 2004 *The 7th International IEEE Conference on Intelligent Transportation Systems (IEEE Cat. No. 04TH8749)* (IEEE) p. 410
[20] Okutani I and Stephanedes Y J 1984 *Transp. Res. Part B Methodol.* **18** 1
[21] Ojeda L L, Kibangou A Y and De Wit C C 2013 2013 *American Control Conference* (IEEE) p. 4724
[22] Silver D, Huang A, Maddison C J, Guez A, Sifre L, Van Den Driessche G, Schrittwieser J, Antonoglou I, Panneershelvam V and Lanctot M 2016 *Nature* **529** 484
[23] Park D and Rilett L R 1999 *Comput. Civ. Infrastruct. Eng.* **14** 357
[24] Huang W, Song G, Hong H and Xie K 2014 *IEEE Trans. Intell. Transp. Syst.* **15** 2191
[25] Siłka J, Wieczorek M and Woźniak M 2022 *Neural Comput. Appl.* **34** 13305
[26] Dey R and Salem F M 2017 2017 *IEEE 60th international midwest symposium on circuits and systems (MWSCAS)* (IEEE) p. 1597

[27] Cho K, Van Merriënboer B, Gulçehre Ç, Bahdanau D, Bougares F, Schwenk H and Bengio Y 2014 *Proceedings of the* 2014 *Conference on Empirical Methods in Natural Language Processing (EMNLP)* p. 1724

[28] Sherstinsky A 2020 *Phys. Nonlinear Phenom.* **404** 132306

[29] Hora S K, Poongodan R, de Prado R P, Wozniak M and Divakarachari P B 2021 *Appl. Sci.* **11** 11263

[30] Graves A and Graves A 2012 *Supervised Seq. Label. Recurr. Neural Netw.* pp. 37–45

[31] Fu R, Zhang Z and Li L 2016 2016 31*st Youth Academic Annual Conference of Chinese Association of Automation (YAC)* (IEEE) p. 324

[32] Jin Y, Xu W, Wang P and Yan J 2018 *2018 5th International Conference on Information, Cybernetics, and Computational Social Systems (ICCSS)* (IEEE) p. 241

[33] Zhang J, Zheng Y and Qi D 2017 *AAAI Conf. Artif. Intell.* **31** 10735

[34] Liu Y, Zheng H, Feng X and Chen Z 2017 2017 9*th International Conference on Wireless Communications and Signal Processing (WCSP)* (IEEE) pp. 1–6

[35] Ke J, Zheng H, Yang H and Chen X M 2017 *Transp. Res. Part C Emerg. Technol.* **85** 591

[36] Yu H, Wu Z, Wang S, Wang Y and Ma X 2017 *Sensors* **17** 1501

[37] Li Y, Yu R, Shahabi C and Liu Y 2017 ArXiv Prepr. ArXiv170701926

[38] Diao Z, Wang X, Zhang D, Liu Y, Xie K and He S 2019 *Proceedings of the AAAI conference on artificial intelligence* **33** 890

[39] Oreshkin B N, Carpov D, Chapados N and Bengio Y 2019 ArXiv Prepr. ArXiv190510437

[40] Tang J, Chen X, Hu Z, Zong F, Han C and Li L 2019 *Phys. Stat. Mech. Its Appl.* **534** 120642

[41] Kim T, Kim J, Tae Y, Park C, Choi J H and Choo J 2021 *International Conference on Learning Representations*