

## *Retraction*

# **Retracted: Urban Traffic Flow Prediction with Deep Neural Network**

### **Security and Communication Networks**

Received 5 December 2023; Accepted 5 December 2023; Published 6 December 2023

Copyright © 2023 Security and Communication Networks. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

This article has been retracted by Hindawi, as publisher, following an investigation undertaken by the publisher [1]. This investigation has uncovered evidence of systematic manipulation of the publication and peer-review process. We cannot, therefore, vouch for the reliability or integrity of this article.

Please note that this notice is intended solely to alert readers that the peer-review process of this article has been compromised.

Wiley and Hindawi regret that the usual quality checks did not identify these issues before publication and have since put additional measures in place to safeguard research integrity.

We wish to credit our Research Integrity and Research Publishing teams and anonymous and named external researchers and research integrity experts for contributing to this investigation.

The corresponding author, as the representative of all authors, has been given the opportunity to register their agreement or disagreement to this retraction. We have kept a record of any response received.

### **References**

- [1] J. Yang, "Urban Traffic Flow Prediction with Deep Neural Network," *Security and Communication Networks*, vol. 2022, Article ID 8711873, 10 pages, 2022.

## Research Article

# Urban Traffic Flow Prediction with Deep Neural Network

Jin Yang <sup>1,2</sup>

<sup>1</sup>Wuhan Technical College of Communications, Wuhan 430065, China

<sup>2</sup>Nanjing Chongxin Digital Technology Co. Nanjing, China

Correspondence should be addressed to Jin Yang; [tobbyoung@whtcc.edu.cn](mailto:tobbyoung@whtcc.edu.cn)

Received 20 March 2022; Revised 20 April 2022; Accepted 23 April 2022; Published 1 June 2022

Academic Editor: Muhammad Arif

Copyright © 2022 Jin Yang. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

It is critical to realize accurate collecting, visualization, rule mining, and prediction analysis of the traffic flow operating state in order for the intelligent transportation system to achieve exact management and control of traffic flow. Traffic flow prediction is primarily concerned with traffic data on roadways, which has both temporal and spatial correlations. Aiming at the spatiotemporal characteristics, this paper studies two aspects and designs a traffic flow prediction model with a deep neural network. First, this work proposes a traffic flow spatial feature learning algorithm with the combination of graph convolutional neural network and attention mechanism. Distinct weights are assigned to the degree of mutual impact between different nodes, and node adaptive learning is implemented at the same time, which modifies the standard parameter sharing mode, allowing for improved expressive ability and spatial feature extraction. Secondly, a learning algorithm for temporal characteristics of traffic flow based on the temporal convolutional network is proposed, which ensures that the dimensions of input and output data are consistent through causal convolution. The dilated convolution can flexibly control the receptive field by setting the sampling interval and can also extract temporal features well for long-length spatiotemporal sequence data. Finally, a spatiotemporal graph attention-based traffic flow prediction approach is constructed. To learn features, learn parameters for multiple modes and improve the model effect, this model employs a combination of graph convolutional neural networks and an attention mechanism. It uses a temporal convolutional network to expand the receptive field, better capture temporal features, and finally add residual connections to prevent problems such as overfitting caused by too deep network layers.

## 1. Introduction

With the increase in urban population, it has made the number of motor vehicles also shown a rapid growth trend, and at the same time has brought enormous pressure to urban traffic. According to the 2012 Urban Transportation Report published by the Texas Transportation Institute, the United States spends about \$121 billion annually on traffic congestion control. In China's large cities with a population of more than one million, the annual direct or indirect economic losses caused by traffic congestion have reached 200 billion yuan, accounting for about 5% to 8% of GDP. Since the implementation of the reform and opening-up policy, there have been significant development and changes in urban construction. As the economy continues to grow and the level of urbanization continues to rise, improving the efficiency of transportation is not only the basis for ensuring the quality of urban residents' traffic and promoting the harmonious and stable development of society

but also the cornerstone for implementing the idea of developing a strong transportation country in the new era. Fundamentally speaking, the traffic problem is to explore how people, vehicles, roads, and the environment can be effectively coordinated and unified in the transportation system. The main ways to alleviate traffic problems are as follows:

- (1) Control traffic demand, such as giving priority to the development of traffic modes with less consumption of traffic resources, off-peak travel, prohibiting travel, and using economic levers to control.
- (2) Increase the construction of road infrastructure, mainly including the supporting construction of access roads, elevated roads, and ground public transportation.
- (3) Improve the configuration of intersection facilities in the existing road network, including channelization of intersections, optimization of traffic signal cycle

and green signal ratio, and addition of traffic signs and markings.

- (4) Accelerate the development of intelligent transportation systems and establish a comprehensive transportation management system that integrates information technology, computer technology, and artificial intelligence in the entire urban road network [1–5].

The intelligent transportation system has been applied and used effectively in many developed countries, achieving good results and accumulating a lot of valuable experience. The idea for solving traffic issues has shifted from continuously increasing traffic supply to meet traffic demand to adopting scientific regulation of traffic supply determined by urban traffic planning in order to further strictly control traffic demand, i.e., traditional traffic governance has shifted. For the modernization and intelligent traffic management based on the intelligent transportation system, it will better solve the urban traffic problem [6–10].

With the advancement of science and technology and the development of society, the traffic control method based on the intelligent transportation system gradually occupies an important position in modern traffic management and control. When applying intelligent transportation technology to control the running state of traffic flow, the implementation of many important links and functions in the system is based on accurate traffic prediction. It inspired researchers to conduct another round of in-depth research on traffic flow prediction. This task is with the existing historical data, acquisition methods mainly include in-road coil detection, roadside microwave measurement device detection, floating car data, and video recognition. By establishing an appropriate traffic forecast model, the traffic flow operation state in a certain time period in the future can be predicted [11–15]. A GRU model with one fewer gate can be utilized to create predictions, according to reference [16]. As a result, as compared to LSTM [17], it is more straightforward, operates faster, and considers the impact of meteorological variables, resulting in improved prediction accuracy.

To sum up, the traffic flow data does contain a wealth of actual traffic information, but a reasonable and appropriate selection of technical methods can fully excavate the regular features hidden in the data from different perspectives such as time and space and summarize them. The spatial-temporal correlation characteristics and operation laws provide theoretical support for the accurate prediction of traffic flow information.

The following is a summary of the research: Section 2 discusses the related work. Section 3 discusses the method of the proposed work. Section 4 discusses the experiment and discussion; finally, the conclusion brings the paper to a finish in Section 5.

## 2. Related Work

In the long short-term memory network, three gated units, the forgetting gate, the input gate, and the output gate, work together to address the problem of long-term dependency.

Reference [18] proposes a forecasting strategy based on LSTM for nonlinear traffic flow that takes advantage of LSTM's inherent ability to capture long-term associations in time series data. According to reference [19], the LSTM method can be improved by concatenating long sequence time steps with high impact values. Use the attention method to capture these high-impact traffic flow metrics and get strong prediction results. A deep trend model was suggested in reference [20], which comprises an extraction layer that can extract the time-varying trend of traffic flow and a prediction layer that employs an LSTM network to calculate and output the time-variant trend. One LSTM model is similar to this approach. It is better at predicting short-term statistics on traffic flow. There was a KNN-LSTM approach proposed in reference [21]. Two-layer LSTM networks are used to predict traffic for the selected stations individually, using a KNN algorithm to choose the most relevant stations and LSTM to capture temporal aspects of the traffic.

Simple structure and faster training speed led to a new proposal for a better model of LSTM for deep learning, which was promptly applied to a short-term traffic flow prediction study. According to reference [16], a GRU model with one fewer gate can be used to make predictions. As a result, compared to LSTM [17], it is more straightforward, operates more quickly, and takes into account the influence of weather conditions, enhancing prediction accuracy. However, merely utilizing LSTM or GRU models cannot completely assess the temporal and geographical aspects of traffic flow because they can only deal with time series. As a result, the short-term traffic flow prediction method based on spatiotemporal correlation does not have the same level of accuracy.

One of the most important discoveries in traffic flow is convolutional neural networks (CNN). Reference [22] suggested a long-term traffic flow forecast deep generative network based on residual deconvolution. Multichannel residual deconvolutional neural networks are used in the generator, while convolutional neural networks are used in the discriminator to optimize the adversarial training process, with good results. When it comes to short-term traffic flow forecast, the influence of route structure and weather conditions is completely considered in reference [23]. As a convolutional neural network is used in conjunction with a long short-term memory network, the final prediction results are achieved by merging all three features fully. The results of the experiments suggest that the prediction approach can increase the accuracy of the forecast. One of the references [24] recommends the usage of CNN with LSTM to make predictions (CNN-LSTM), uses LSTM and CNN to extract the temporal and periodic aspects of traffic flow, and then merges the three features into a traffic flow prediction result. However, the convolutional neural network modules and the long short-term memory network are used simultaneously for feature extraction. The final traffic flow characteristics cannot be fully incorporated because of this. A convolutional-long short-term memory network (conv-LSTM) was devised in reference [25]. A bidirectional long short-term memory network (bi-LSTM) [26] model is used to extract the periodic aspects of traffic flow and blends RNN

and CNN into one model. Finally, all of the features have been merged to enhance the precision with which traffic flow patterns may be predicted.

The methods of traffic flow prediction are far more than these, and deep belief networks [27], which can perform both supervised and unsupervised learning, have become the research goals of scholars. The deep belief network is mainly composed of neural networks composed of a large number of restricted Boltzmann machines (RBMs), and each RBM has a hidden layer and a display layer. Moreover, these neurons are independent, that is, do not depend on other neurons and can achieve simultaneous parallel computing. Due to the deep neural layer and the slow training speed, the deep belief network is usually used to deal with traffic prediction research with a large amount of data. References [28–30] used deep belief networks for traffic flow prediction.

### 3. Method

Accurate traffic flow forecasting models can have a certain impact on road planning, people's travel, and environmental energy. However, traffic flow prediction faces two challenges in practice: temporal correlation and spatial correlation. In space, the topological structure of the road determines spatial characteristics. The upstream section will affect the downstream section, and the traffic conditions of the adjacent sections will also affect. Just the location is different, the correlation for impact is different. Periodic patterns in traffic flow data are mapped to temporal features over time. For example, traffic is congested during morning and evening rush hours, and during holidays, traffic is worse than usual. The core of traffic flow prediction is to solve the above two problems. Aiming at the above problems, this paper adopts two strategies of graph attention network and node adaptive learning to learn features in solving the issue of spatial dependence. To create a spatial dependency, assign different weights to it. This paper uses a temporal convolutional network to obtain longer dependencies by continuously expanding the receptive field in the temporal dimension. A traffic flow prediction model based on spatiotemporal map features is created using the method described above.

**3.1. Spatiotemporal Correlation of Traffic Data and Prediction Framework.** This work constructs a traffic flow prediction model based on spatiotemporal map features by analyzing the temporal and spatial correlations.

**3.1.1. Temporal Correlation.** Changes in traffic flow are measured in months, weeks, or days and reflect periodic changes through time. For example, during morning and evening rush hours, the roads are relatively congested, and traffic flow is considerable on holidays. The prior time period, or even a longer time period, has an impact on traffic flow. The road network traffic changes with time are nonlinear and unstable. If a traffic accident occurs at a certain node of the entire road traffic network, it will inevitably

cause traffic congestion on the surrounding road sections, thereby affecting the traffic speed of the entire road section. In addition to accidents, changes in weather and the occurrence of some unexpected situations will also have a certain impact on traffic speed.

To better quantitatively analyze the temporal correlation of traffic data, this work utilizes the Pearson correlation coefficient to calculate the temporal correlation. This calculation method was proposed by British statistician Carl Pearson, which is defined as for a two-dimensional random variable  $(x, y)$ , and the correlation coefficient  $\rho$  is defined to represent the correlation between  $x$  and  $y$ :

$$\rho = \frac{\text{cov}(x, y)}{\sqrt{\sigma_x} \sqrt{\sigma_y}}. \quad (1)$$

The calculated  $\rho$  value to represent the correlation between variables and the degree of correlation between two variables with different  $\rho$  values are different, as shown in Table 1.

**3.1.2. Spatial Correlation.** In addition to solving the problem of temporal correlation, spatial correlation also needs to be considered. The traffic flow on road forms a spatial topology structure with each sensor as the node, which will inevitably cause the mutual influence between the nodes. This kind of road network structure belongs to the non-Euclidean space, which is different from the Euclidean space where the image is located. Second, the road network is divided into upstream and downstream segments. There will be an effect on downstream traffic from the upstream traffic flow and vice versa for the upstream traffic flow as well. It is impossible to utilize a standard neural network to extract spatial data due to the abovementioned properties of traffic flow. Figure 1 shows the degree of influence of each node on the road surface at different times.

Take point A as an example, nodes around the central node exert different influences on it at various points in time, and this is also true of the influences exerted by different neighbors on the core node. The degree of influence is represented by the depth of the color. The darker color indicates that the traffic flow between the sections has a greater influence, and the light color indicates that the traffic flow between the sections has less influence. For point E, points A and F have the greatest impact in the morning and afternoon, respectively. However, no change has occurred at point D, and the surrounding nodes have little effect on it. The reason for the above phenomenon may be that there are differences in the congestion situation on the upstream as well as downstream roads, and some unexpected situations will also affect the traffic flow on the road. This reflects that traffic flow data has a certain spatial correlation.

**3.1.3. Traffic Flow Forecasting Framework.** This work is primarily concerned with finding a solution to the problem of temporal and spatial correlation of traffic flow data, as defined above and as revealed by the correlation analysis. The overall model structure is shown in Figure 2.

TABLE 1: Description of the correlation coefficient.

$\rho$	Correlation
$ \rho  = 1$	Perfect linear correlation
$0.3 <  \rho  < 1$	Medium linear correlation
$0 <  \rho  < 0.3$	Weak linear correlation
$ \rho  = 0$	No linear correlation

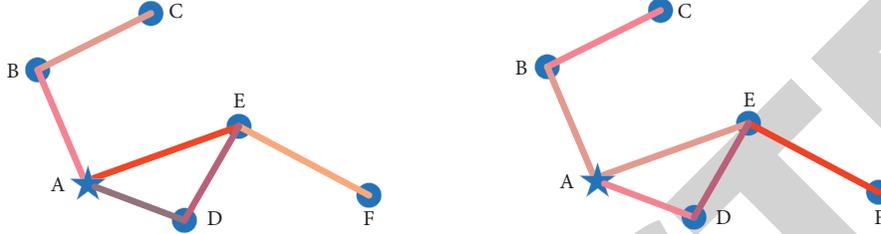


FIGURE 1: Road network spatial correlation diagram.

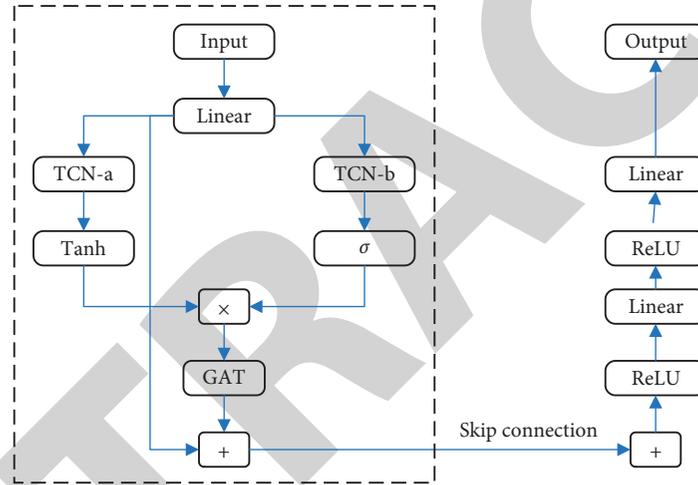


FIGURE 2: The framework of traffic flow prediction with spatiotemporal graph feature.

A spatiotemporal network layer, which deals with spatial and temporal relationships at various temporal levels, is made up primarily of a graph attention network and a temporal convolutional network. The temporal convolutional layer is made up of two temporal convolutional layers. The output layer is directly connected to each of the remaining spatiotemporal layers. On the left, it has a spatial-temporal layer, and on the right, it has an output layer. An initial linear transformation of the input data is performed, followed by temporal convolution and then a neural network for graph attention. Each spatiotemporal layer (as shown in the dashed box) is connected to the output layer through residual connections and linear layers.

**3.2. Spatial Correlation Modeling with Graph Attention Network.** To enhance the expressiveness of the graph structure, this research uses a combination of the graph convolutional network and the attention mechanism for spatial topology. In order to better extract spatial attributes, the attention mechanism is employed to assign weight settings to individual nodes depending on the various

degrees of mutuality between them. This work employs adaptive learning of node parameters to learn different node parameters for different data in order to optimize the model structure in order to avoid the problem of low performance of some models caused by parameter sharing.

**3.2.1. Spatial Graph Attention Layer.** It is possible to assign weights to neighbor nodes in a graph attention network by using the attention mechanism. To a certain extent, this can enhance the expressiveness of the graph neural network model. The attention mechanism is based on cognitive science's understanding of how the brain processes information. People tend to focus on a small portion of the total knowledge because of their limited ability to digest information. When information is assigned weights, it signifies that it must be processed intensively, and the more weight it has, the more intensive the processing. Using  $S$  as the information source,  $Q$  represents particular prior information and  $V$  is the information retrieved from  $S$  by the attention mechanism under the condition of the given  $Q$  information.  $S$  often contains a wide range of data. Each piece of

information in this article is presented in the form of  $k - v$  pairings. The following is an explanation of what is meant by the term attention mechanism:

$$Att(Q, S) = \sum_i \text{similarity}(Q, K_i) * V_i \quad (2)$$

In this paper, the graph attention layer is defined, and the feature vector corresponding to any node  $v_i$  in the graph is  $h_i$ ,  $h_i \in R^{d(l)}$ , and  $d(l)$  is length feature. The output is  $h'_i$ ,  $h'_i \in R^{d(l+1)}$  of each node. Assuming the central node is  $v_i$ , weight coefficients of  $v_j$  to  $v_i$  is

$$e_{ij} = c(Wh_i, Wh_j). \quad (3)$$

The weight from any node to node  $v_i$  can be calculated. To simplify computation, it is restricted to the first-order neighborhood. It should be noted that, in the graph attention network, each node is also regarded as its own neighbor. As long as  $c: R^{d(l+1)} \times R^{d(l+1)} \rightarrow R$ , output a scalar value to represent correlation:

$$e_{ij} = \text{Leaky ReLU}(w^T [Wh_i \| Wh_j]). \quad (4)$$

The activation function is designed as leaky *ReLU*. To better allocate weights, it is necessary to uniformly normalize the correlations calculated with all neighbor nodes, namely,

$$\alpha_{ij} = \text{softmax}(e_{ij}). \quad (5)$$

With the idea of weighted summation for attention, a new feature  $v_i$  is

$$h'_i = \sigma(\sum \alpha_{ij} Wh_j). \quad (6)$$

**3.2.2. Multihead Attention.** To improve expressiveness, a multihead attention can be used. That is to call the  $N$  group of independent attention mechanisms on the above formula and then stitch the output results together:

$$h'_i = \parallel_{n=1}^N \sigma(\sum \alpha_{ij}^{(n)} W^{(n)} h_j). \quad (7)$$

In order to reduce the dimension of the output feature, the concatenation operation can also be replaced by the averaging operation so that the center node and neighbor nodes can share attention between themselves and multiple related characteristics; further sets of independent attention methods are introduced. In doing so, the system's ability to learn is enhanced.

**3.2.3. Node Adaptive Learning.** In the problem of traffic flow prediction, graph convolution has a certain effect in extracting spatial features. According to the graph attention network adopted in this paper, its operation mechanism can be defined as

$$Y = (I_N + D^{-0.5} A D^{-0.5}) X T + b, \quad (8)$$

where  $A$  is the adjacency matrix,  $D$  is the degree matrix,  $X$  as well as  $Y$  are the input and output of the graph attention network, and  $T$  and  $b$  represent weight and bias, respectively.

From the perspective of a node (e.g., node  $i$ ), the graph convolution operation can be seen as transforming the features of node  $X_i$  into  $Y_i$  and sharing  $T$  and  $b$  among all nodes. While sharing parameters may help learn the most salient patterns across all nodes in many problems and can significantly reduce the number of parameters, there are still some problems. In addition to the close spatial correlation between traffic sequences, different patterns also exist between different traffic sequences due to the dynamic suitability of time series data and various factors that may affect traffic. On the one hand, traffic from two adjacent nodes may also exhibit different patterns during some specific periods due to their specific properties (e.g., POI, weather). On the other hand, a traffic sequence from two disjoint nodes might even show reverse patterns. Therefore, merely capturing shared patterns among all nodes is not sufficient for accurate traffic prediction, and a unique parameter space must be maintained for each node to learn node-specific patterns.

To solve this problem, this paper introduces the idea of matrix decomposition and proposes to use the node adaptive parameter learning module to enhance the graph attention network. Instead of learning  $T$  directly, this method learns two smaller parameter matrices: the node embedding matrix  $T_1$  and the weight matrix  $T_2$ , so that  $T$  can be generated by  $T = T_1 T_2$ . The node-specific parameter patterns to be learned are selected from candidate parameter patterns found in all traffic flow sequences. The same operation can also be applied to  $b$ . Finally, node adaptive learning can be applied to graph convolution:

$$Y = (I_N + D^{-0.5} A D^{-0.5}) X T_1 T_2 + T_1 b_G. \quad (9)$$

**3.3. Temporal Correlation Modeling with Temporal Convolutional Network.** RNN is commonly used structure to modeling time series problems, and it cannot capture long-term dependency information well. The temporal convolutional neural network TCN [31] overcomes the abovementioned defects well.

In TCN, the output at  $t$  is dependent on the current layer as well as the previous layer. The mechanism adopted to keep the data dimension in TCN unchanged is the structure of a one-dimensional fully connected neural network, and the input and output of each hidden layer maintain the same time step. This allows the output layer to be passed on at the same length as the input layer, which can be done using zero-value padding.

If a longer dependency relationship needs to be obtained, the number of layers needs to be continuously deepened, which will cause the number of network layers to deepen, and the receptive field obtained in this way is limited. To solve this issue, the key lies in how to expand a larger receptive field with fewer parameters. For this reason, the dilated convolution is introduced into the temporal convolutional network model:

$$F(9) = \sum_{i=0}^{k-1} f(i) x_{s-d \cdot i}, \quad (10)$$

where  $d$  is the dilated factor, which represents the expansion rate of the atrous convolution.  $d$  increases exponentially with layers, and its role is to deal with the problem of long dependencies. Each layer computes convolutions at intervals of  $d - 1$  position. The different layers are set to 2, 4, and 8 in sequence from bottom to top. To make the output length the same as the input, the input sequence data can be filled. The schematic diagram of dilated convolution is illustrated in Figure 3.

## 4. Experiment and Discussion

In this chapter, we defined the dataset, experimental set and evaluation metric, comparison with other methods, evaluation of temporal correlation and spatial correlation, and evaluation on multihead attention in detail.

*4.1. Dataset.* This paper conducts experiments on the model on two self-made urban traffic datasets (UTA and UTB) to verify the effectiveness of the proposed model. UTA records the traffic information of city A, including the four-month statistical data from January 1, 2018, to April 30, 2018, collected by 212 sensors. UTB records the traffic information of city B, including the four-month statistics from January 1, 2018, to April 30, 2018, collected by 325 sensors. The specific data set is illustrated in Table 2.

This work uses Z-score to normalize the data. This method, also known as the standard score, is calculated as the difference between a number and the mean divided by the standard deviation. In statistics, the concept is the number of symbols that a data value is above the mean of the observed values by the standard deviation:

$$Z = \frac{(X - \bar{X})}{S}, \quad (11)$$

where  $X$  the traffic flow data from the sensor is,  $\bar{X}$  is the mean of the data values, and  $S$  is the standard deviation of the data values. The normalized data is calculated according to the above formula and used for subsequent experiments.

*4.2. Experimental Set and Evaluation Metric.* The experimental setup of this experiment has the following aspects:

- (1) In the graph attention neural network, the value of  $N$  in the multi-head attention mechanism is set to 4
- (2) The expansion factor in the temporal convolutional network is set to 1, 2, 1, 2, 1, 2, 1, 2, and the convolution kernel size is set to 2
- (3) The model is trained using the ADAM optimizer with an initial learning rate of 0.001, dropout is set to 0.5, batch size is set to 64, and ReLU is used as the activation function of the neural network

This work utilizes mean absolute error (MAE) and root mean square error (RMSE) as evaluation metrics. The value range of the price indicator is  $[0, +\infty)$ . When the predicted

value exactly matches the true value, MAE or RMSE is equal to 0, which is a perfect model. The larger the error, the larger the value.

*4.3. Comparison with Other Methods.* To verify the validity, this topic selected the current mainstream and classic traffic flow prediction models for comparative experiments. The main methods of comparison are STGCN (spatiotemporal graph convolutional neural network, which combines graph convolutional layers and convolutional neural networks), DCRNN (diffusion convolution cyclic neural network, the combination of diffuse convolution and cyclic neural network, the prediction effect is ideal), ASTGCN (the convolutional neural network is combined with graph convolution to form a spatiotemporal block structure, which extracts adjacent, daily and weekly dependent features respectively), WaveNet (a sequence generation model using dilated convolutions), and GraphWaveNet (a method of combining graph convolution with dilated convolution).

The experimental results on the traffic flow prediction problem based on graph attention traffic flow prediction results and comparison methodologies are shown in Tables 3–5. The study of these results reveals that the method in this paper's prediction outcomes are relatively good.

Compared with other models, our method achieves the best results on these two datasets. Compared with methods that consider time-dependent representation, the model proposed in this paper not only considers the temporal feature factors but also combines temporal and spatial correlations and has better performance. For graph-based convolution methods, taking WaveNet as an example, the architecture proposed in this paper does not adopt the traditional parameter sharing mode and pays more attention to the weight of each node of the graph. This makes it possible to improve the expressive ability of the model through adaptive learning between different nodes.

*4.4. Evaluation of Temporal Correlation and Spatial Correlation.* As mentioned earlier, this work fuses temporal and spatial correlations. To verify the effectiveness of this strategy, this paper conducts comparative experiments to compare the prediction performance when using a single temporal correlation, a single spatial correlation, and when the two correlations are fused. The experimental results are illustrated in Figure 4.

Obviously, only using a single temporal correlation feature or a single spatial correlation feature does not achieve the best prediction performance. When the two features are combined, the use of the spatiotemporal correlation feature can achieve the lowest MAE and RMSE. Therefore, this can prove the validity and correctness of using spatiotemporal features in this work.

*4.5. Evaluation on Multihead Attention.* As mentioned above, this work uses a multihead attention mechanism when building an urban traffic flow prediction network.

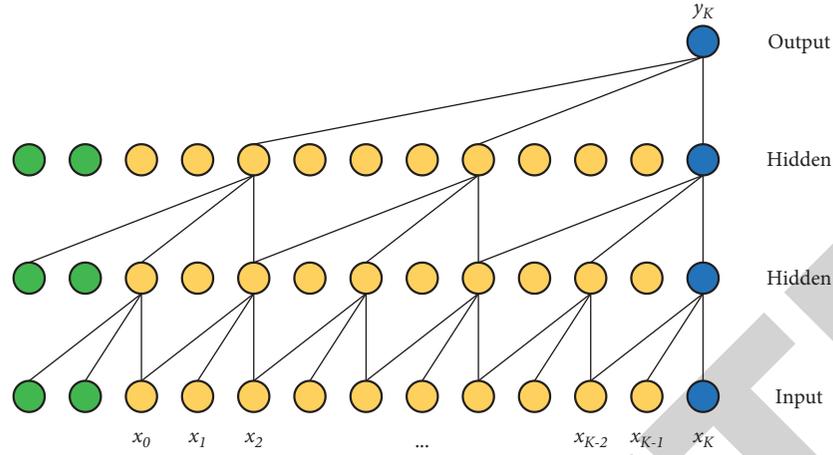


FIGURE 3: The diagram of dilated convolution.

TABLE 2: The detailed information about UTA and UTB.

Dataset	Training set	Test set	Total
UTA	30981	12683	43664
UTB	34108	15925	50033

TABLE 3: Comparison of prediction results of traffic flow in the next 15 minutes.

Method	UTA (15 min)		UTB (15 min)	
	MAE	RMSE	MAE	RMSE
STGCN	3.98	6.74	2.38	3.99
DCRNN	3.89	6.38	2.65	3.96
ASTGCN	3.72	6.26	2.51	4.03
WaveNet	4.01	6.90	2.41	4.11
GraphWaveNet	3.86	7.32	2.32	3.83
Ours	3.69	6.11	2.28	3.70

TABLE 4: Comparison of prediction results of traffic flow in the next 30 minutes.

Method	UTA (30 min)		UTB (30 min)	
	MAE	RMSE	MAE	RMSE
STGCN	4.51	8.15	2.88	5.29
DCRNN	4.17	7.56	2.82	4.99
ASTGCN	4.15	8.23	2.85	4.88
WaveNet	4.35	8.12	2.62	4.91
GraphWaveNet	4.48	7.25	2.67	4.73
Ours	4.13	7.19	2.55	4.67

TABLE 5: Comparison of prediction results of traffic flow in the next 60 minutes.

Method	UTA (60 min)		UTB (60 min)	
	MAE	RMSE	MAE	RMSE
STGCN	5.62	9.82	3.52	6.71
DCRNN	4.66	9.91	3.12	5.77
ASTGCN	4.71	8.66	3.25	6.41
WaveNet	5.47	9.82	3.36	6.46
GraphWaveNet	4.54	8.67	2.99	5.56
Ours	4.43	8.52	2.91	5.41

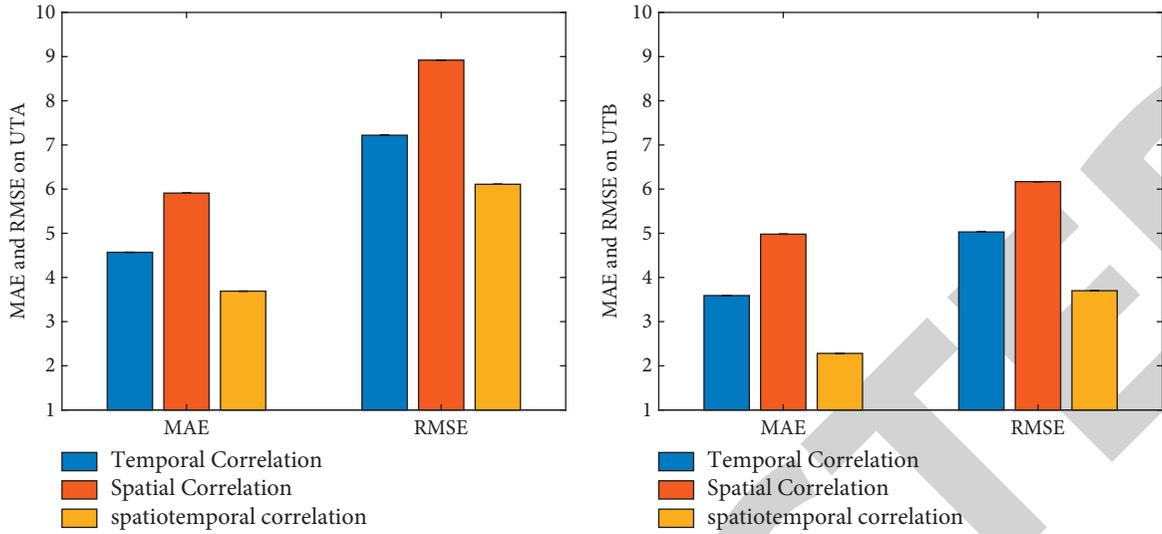


FIGURE 4: Evaluation on temporal correlation and spatial correlation.

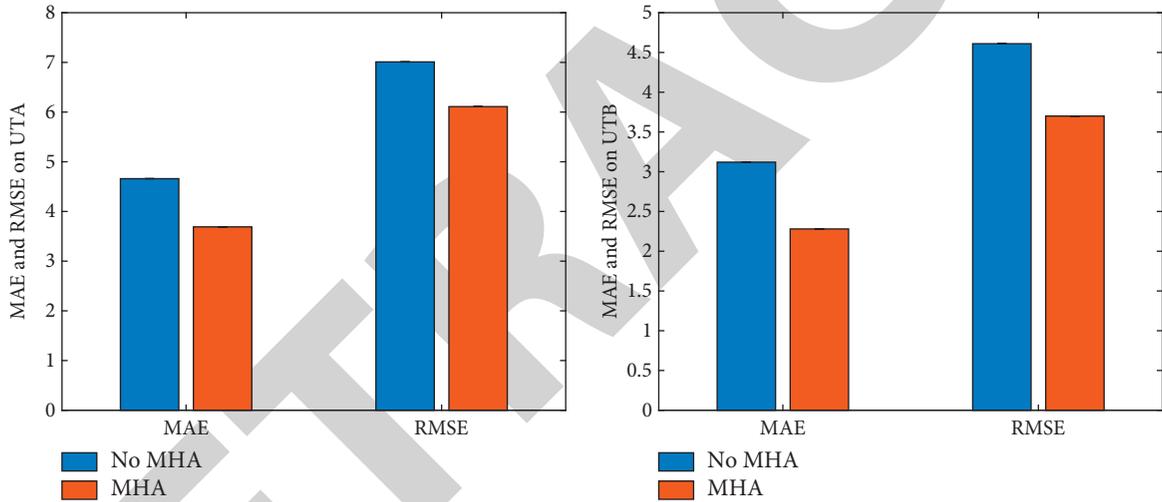


FIGURE 5: Evaluation on multihead attention.

Comparative studies are carried out in this work to validate the effectiveness and correctness of this technique. This paper compares the prediction performance without this attention mechanism and the prediction performance with this attention mechanism, and the experimental results are illustrated in Figure 5. MHA is the multihead attention mechanism.

As the multihead attention mechanism is implemented, the network's prediction performance is clearly improved when compared to when this technique is not used. This is because applying this technique can lead to the network learning more discriminative features, proving the strategy's validity and feasibility in this study.

## 5. Conclusion

With the rapid growth of smart cities, smart transportation has become a prominent topic, with the challenge of urban traffic flow forecasting at the center. This is also directly tied

to people's daily lives, as it has an ongoing impact on people's travel, road planning, environmental, and energy conservation. The rise of graph convolutional neural networks in the past two years has provided a new idea for urban traffic flow prediction. With above research background as well as related theoretical basis, this work proposes a traffic flow prediction strategy with spatiotemporal graph features. The main results achieved are as follows: (1) a graph attention neural network and node adaptive learning are proposed spatially. To extract spatial topology more efficiently, assign various weights to different nodes and use nodes to learn different parameters adaptively. (2) It is proposed to flexibly control the receptive field to extract temporal features through temporal convolutional network. In the model building, the residual connection is introduced to avoid the occurrence of overfitting and the over-complexity of the model. (3) Apply the model described above to two real-world datasets. The results of the experimental test on the original data set reveal that our model has

the lowest value of two evaluation indicators when compared to the usual prediction approach. This indicates that the predictive ability of the model is relatively good. There are still the following areas for improvement in this paper: (1) this paper does not consider other external factors that affect traffic conditions, such as weather, traffic accidents, etc. This will lead to a certain degree of impact on accuracy. (2) Although the proposed technique takes into account static dependencies, the relationship between locations is dynamic and can alter depending on a variety of conditions. This is something that should be taken into account in future studies.

## Data Availability

The datasets used during the current study are available from the corresponding author on reasonable request.

## Conflicts of Interest

The author declares no conflicts of interest.

## References

- [1] F. Zhang, T.-Y. Wu, Y. Wang et al., "Application of quantum genetic optimization of LVQ neural network in smart city traffic network prediction," *IEEE Access*, vol. 8, pp. 104555–104564, 2020.
- [2] S. Carrasco, P. Medina, J. Rogan, and J. A. Valdivia, "Simulations suggest that navigation software may not be as efficient as expected for city traffic," *Chaos: An Interdisciplinary Journal of Nonlinear Science*, vol. 31, no. 3, Article ID 033103, 2021.
- [3] D. D. Nguyen, J. Rohacs, and D. Rohacs, "Autonomous flight trajectory control system for drones in smart city traffic management," *ISPRS International Journal of Geo-Information*, vol. 10, no. 5, p. 338, 2021.
- [4] S. Sun, H. Wu, and L. Xiang, "City-wide traffic flow forecasting using a deep convolutional neural network," *Sensors*, vol. 20, no. 2, p. 421, 2020.
- [5] I. F. Echendu, F. Okeke, and R. C. Nnaemeka-Okeke, "An investigation of challenges in the existing pattern of intra-city traffic in Enugu metropolis," *TeMA-Journal of Land Use, Mobility and Environment*, vol. 13, no. 2, pp. 191–208, 2020.
- [6] C. Chen, B. Liu, S. Wan, P. Qiao, and Q. Pei, "An edge traffic flow detection scheme based on deep learning in an intelligent transportation system," *IEEE Transactions on Intelligent Transportation Systems*, vol. 22, no. 3, pp. 1840–1852, 2021.
- [7] S. Kaffash, A. T. Nguyen, and J. Zhu, "Big data algorithms and applications in intelligent transportation system: a review and bibliometric analysis," *International Journal of Production Economics*, vol. 231, Article ID 107868, 2021.
- [8] S. Saharan, S. Bawa, and N. Kumar, "Dynamic pricing techniques for Intelligent Transportation System in smart cities: a systematic review," *Computer Communications*, vol. 150, pp. 603–625, 2020.
- [9] X. Li, J. Tan, A. Liu, P. Vijayakumar, N. Kumar, and M. Alazab, "A novel UAV-enabled data collection scheme for intelligent transportation system through UAV speed control," *IEEE Transactions on Intelligent Transportation Systems*, vol. 22, no. 4, pp. 2100–2110, 2021.
- [10] H. Zhang and X. Lu, "Vehicle communication network in intelligent transportation system based on Internet of Things," *Computer Communications*, vol. 160, pp. 799–806, 2020.
- [11] Z. Zhang, A. Zhang, C. Sun, S. Xiang, J. Guan, and X. Huang, "Research on air traffic flow forecast based on ELM non-iterative algorithm," *Mobile Networks and Applications*, vol. 26, no. 1, pp. 425–439, 2021.
- [12] R. Wang, "Application research of short-term traffic flow forecast based on bat algorithm support vector machine," *IOP Conference Series: Materials Science and Engineering*, vol. 740, no. 1, Article ID 012168, 2020.
- [13] J. Pei, K. Zhong, J. Li, J. Xu, and X. Wang, "Evaluating a cluster-neural network model for city innovation capability," *Neural Computing and Applications*, pp. 1–13, 2021.
- [14] Y. Zhang, S. Yang, and D. Xin, "Short-term traffic flow forecast based on improved wavelet packet and long short-term memory combination model," *Journal of Transportation Systems Engineering & Information Technology*, vol. 20, no. 2, pp. 204–210, 2020.
- [15] M. A. C. Alves and R. L. F. Cordeiro, "Effective and unburdensome forecast of highway traffic flow with adaptive computing," *Knowledge-Based Systems*, vol. 212, Article ID 106603, 2021.
- [16] D. Zhang and M. R. Kabuka, "Combining weather condition data to predict traffic flow: a GRU – based deep learning approach," *IET Intelligent Transport Systems*, vol. 12, no. 7, pp. 578–585, 2018.
- [17] R. Fu, Z. Zhang, and L. Li, "Using LSTM and GRU Neural Network Methods for Traffic Flow prediction," in *Proceedings of the Youth Academic Annual Conference of Chinese Association of Automation*, pp. 324–328, Wuhan, China, November 2016.
- [18] H. Shao and B. H. Soong, "Traffic Flow Prediction with Long Short-Term Memory Networks (LSTMs)," in *Proceedings of the IEEE region 10 conference*, pp. 2986–2989, Singapore, November 2016.
- [19] 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.
- [20] X. Dai, R. Fu, and Y. Lin, "Deeptrend: A Deep Hierarchical Neural Network for Traffic Flow prediction," 2017, <https://arxiv.org/abs/1707.03213#:~:text=Inspired%20by%20the%20detrrending%20method,extraction%20layer%20and%20prediction%20layer.>
- [21] X. Luo, D. Li, Y. Yang, and S. Zhang, "Spatiotemporal traffic flow prediction with KNN and LSTM," *Journal of Advanced Transportation*, vol. 2019, pp. 1–10, Article ID 4145353, 2019.
- [22] D. Zang, Y. Fang, Z. Wei, K. Tang, and J. Cheng, "Traffic flow data prediction using residual deconvolution based deep generative network," *IEEE Access*, vol. 7, pp. 71311–71322, 2019.
- [23] Z. Zheng, Y. Yang, J. Liu, H.-N. Dai, and Y. Zhang, "Deep and embedded learning approach for traffic flow prediction in urban informatics," *IEEE Transactions on Intelligent Transportation Systems*, vol. 20, no. 10, pp. 3927–3939, 2019.
- [24] Y. Wu and H. Tan, "Short-term traffic flow forecasting with spatial-temporal correlation in a hybrid deep learning framework," 2016, <https://arxiv.org/abs/1612.01022>.
- [25] T. Chen, R. Xu, Y. He, and X. Wang, "Improving sentiment analysis via sentence type classification using BiLSTM-CRF and CNN," *Expert Systems with Applications*, vol. 72, pp. 221–230, 2017.
- [26] Y. Liu, H. Zheng, and X. Feng, "Short-term traffic flow prediction with Conv-LSTM," in *Proceedings of the*

- International Conference on Wireless Communications and Signal Processing*, pp. 1–6, Nanjing, China, October 2017.
- [27] G. E. Hinton and R. R. Salakhutdinov, “Reducing the dimensionality of data with neural networks,” *Science*, vol. 313, no. 5786, pp. 504–507, 2006.
- [28] H. Tan, X. Xuan, and Y. Wu, “A comparison of traffic flow prediction methods based on DBN,” in *Proceedings of the 16th COTA International Conference of Transportation Professionals*, pp. 273–283, Shanghai, China, June 2016.
- [29] Z. Yaying and H. Guan, “Traffic flow prediction model based on deep belief network and genetic algorithm,” *IET Intelligent Transport Systems*, vol. 12, no. 6, pp. 533–541, 2018.
- [30] L. Zhao, Y. Zhou, H. Lu, and H. Fujita, “Parallel computing method of deep belief networks and its application to traffic flow prediction,” *Knowledge-Based Systems*, vol. 163, pp. 972–987, 2019.
- [31] S Bai, J. Z Kolter, and V. Koltun, “Terllis Networks for sequence modelling,” 2018, <https://arxiv.org/abs/1810.06682>.

RETRACTED