Applied Intelligence
https://doi.org/10.1007/s10489-021-02879-1

Deep spatio-temporal neural network based on interactive attention for traffic flow prediction

Hui Zeng[1] · Zhiying Peng[1] · XiaoHui Huang[1] · Yixue Yang[1] · Rong Hu[1]

Received: 30 June 2021 / Accepted: 29 September 2021

Abstract
Traffic flow forecasting is of great significance to urban traffic control and public safety applications. The key challenge of traffic flow forecasting is how to capture the complex correlation of different time levels and learn time dependence. Some external information is closely related to traffic flow, such as accidental traffic accidents, weather, and Point of Interests (PoI) information. This paper proposes a deep learning-based model, called AttDeepSTN+, which is used to predict the inflow and outflow of each area of the entire city. Specifically, AttDeepSTN+ uses the structure of interactive attention and convolution to model the temporal closeness, trend, and periodicity of crowd flow, in the interactive attention layer, learn the importance of closeness to periodicity and trend respectively to model the long-term dependence of time, and then use feature fusion to capture complex correlations at different levels, thereby reducing model prediction accuracy. In addition, PoI information are combined with time factors to express the influence of location attributes on crowd flow, to learn prior knowledge of crowd flow. Finally, a new fusion mechanism is used to fuse the attention layer modules and PoI information and other information together into the module to capture the complex correlation between different levels of features, to predict the final crowd flow in each area, and further improve the prediction accuracy of the model. The New York City crowd flow experiment shows that the model is better than the current state-of-the-art baseline.

Keywords Traffic prediction · Spatio-temporal network · Attention mechanism

1 Introduction

Traffic flow prediction is one of the most important issues in artificial intelligence research. It is of great significance to the application of urban traffic control and public safety. Its purpose is to predict the traffic flow in the future on the premise of given historical data. An accurate traffic prediction model is indispensable for the above-mentioned applications. For example, an accurate prediction model can help drivers optimize driving routes, allocate resources reasonably and reduce urban traffic congestion [12, 13, 21]. Predicting the flow of people in a scenic spot can help the scenic spot better control the total number of tourists, so as to reduce the occurrence of incidents that affect public safety such as trampling [26, 51, 25].Taxi demand forecasting can help taxi companies to reasonably allocate the number of taxis [41, 6, 2].Traffic flow forecasting can help the traffic department to effectively control the traffic [7, 18, 30], thereby reducing the occurrence of traffic congestion and other problems. Due to the availability of large-scale traffic data and the importance of traffic flow forecasting in reality, traffic flow forecasting has attracted more and more attention from researchers in the field of artificial intelligence [23, 50, 8, 32, 3]. The key challenge of traffic flow forecasting is how to model complex temporal and spatial correlations [17, 28, 9]. Secondly, traffic flow is also closely related to some external information, such as accidental traffic accidents, weather, and Point of Interests (PoI) information [40, 11].

In a typical traffic flow prediction problem, given the historical traffic data of an area (for example, the traffic flow of each hour in the previous two months), it is necessary to predict the traffic in a certain time interval in the future (half an hour or an hour) flow. In this paper, we predict two types of crowd movement: inflow and outflow [24, 53]. As shown in Fig. 1, inflow refers to the flow of people flowing into this area from other areas within a certain time interval, outflow refers to the flow of people flowing out of this area

✉ Zhiying Peng
pzhiying0224@163.com

[1] School of Information Engineering Department, East China Jiaotong University, Nanchang 330013, China

Published online: 13 January 2022

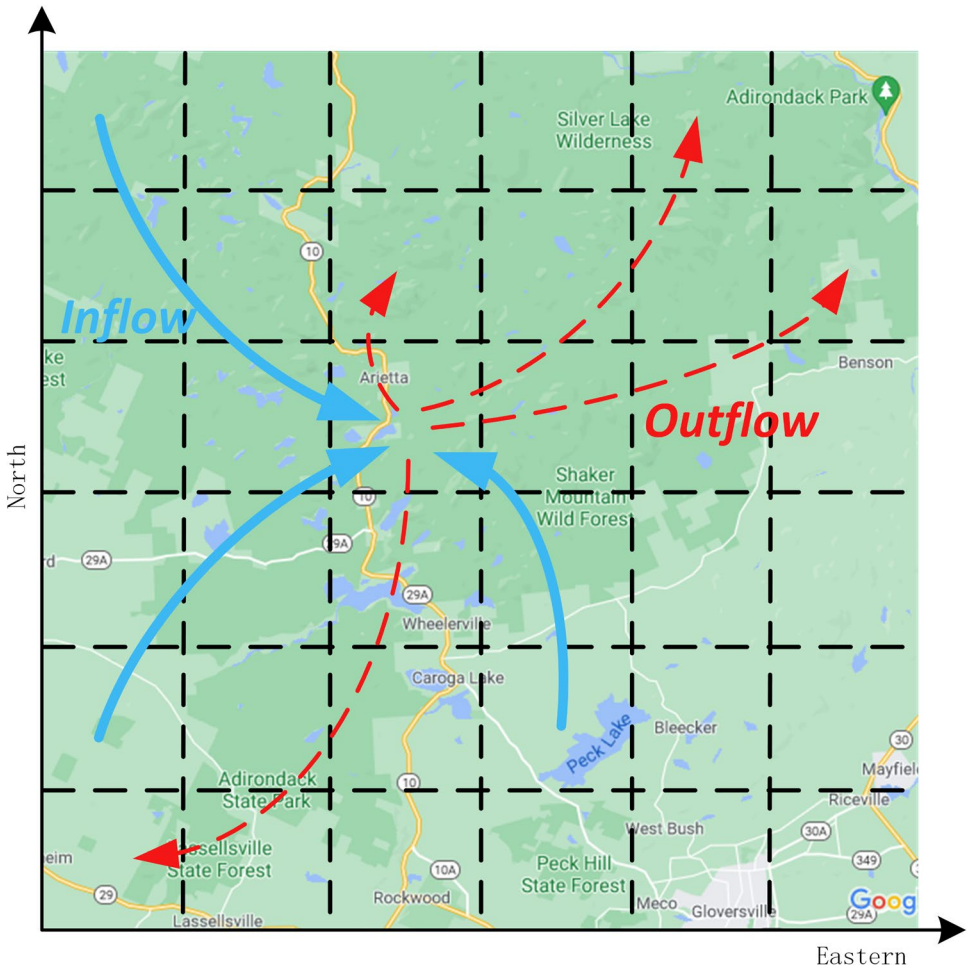


Fig. 1 The movement of people in the region

to other areas within a certain time interval. For decades, many studies have been devoted to the study of traffic flow forecasting. Some classic time series forecasting models [33, 43, 52, 16], such as Auto Regressive Integrated Moving Average (ARIMA) and Kalman filtering, have been widely used in traffic forecasting problems, but they are only suitable for stationary time series. Although research in recent years has begun to incorporate some external information such as space (such as weather, events, etc.), these methods are based on traditional linear time series models or machine learning models and cannot well capture complex nonlinear spatio-temporal relationships.

With the popularity of deep learning, some deep learning-based models have achieved great success in many challenging tasks. This success has greatly.

stimulated the research of deep learning-based models. For example, some studies model the entire city's traffic flow as a heat map [31, 49] and use convolutional neural networks to deal with nonlinear spatial relationships. In order to model the nonlinear time dependence, some studies have proposed processing based on recurrent neural networks [47, 4, 1, 34]. Some researchers combine convolutional neural networks and LSTM to jointly model time and space dependent methods [5, 44, 39].

Although spatial correlation and temporal dynamics are considered in the deep learning model of traffic prediction, the existing methods still have two shortcomings that make the prediction results inefficient and inaccurate:

- One disadvantage is that although the existing research considers periodicity, *it does not consider the timing correlation between periodicities and the correlation and importance of adjacent time to periodicity.* Traffic flow data has an obvious daily cycle and weekly cycle. The adjacent time has a certain degree of influence on the daily cycle and the weekly cycle. For example, if a region has a large flow of people in a certain time interval today, then the flow of people will also show similarity during the same time interval yesterday. The same time interval last week will also show similarities, which were not considered in the previous study.
- *Another disadvantage is the neural network structure with redundancy and huge parameters.* Existing research has complicated connection methods and lacks

interactivity. In some studies, the author trains the three components independently without interaction, which will also lead to complex and unstable network structure.

In order to solve the above two major challenges, this paper proposes a new deep learning network model, a deep spatio-temporal neural network based on interactive attention for traffic flow prediction (AttDeepSTN+). AttDeepSTN+ is a neural network based on time and space. It proposes an interactive attention mechanism to learn the similarity and influence between temporal closeness, periodicity, and trend. This mechanism uses the attention mechanism to better capture the complex correlation and influence between traffic data. A fusion mechanism is introduced to capture the complex correlation between features at different levels, so that the trainable parameters of the entire network model are reduced, and the structure is more distinct. In this way, our proposed model can be modeled in very complex nonlinear time and space and achieves better performance. Our main contributions are summarized as follows.

- We propose an interactive attention mechanism to learn the correlation between temporal closeness, period, trend and the degree of influence between the three components. This mechanism uses the attention mechanism to better capture the complexity and importance of traffic data.
- We propose a fusion method that uses fusion to capture the complex correlation between features at different levels after the interactive attention layer. Finally, the final multi-feature fusion is used in the model to reduce the trainable parameters of the entire network model and the structure is clear. Improve the accuracy of model prediction.
- We conducted extensive experiments based on the New York City crowd flow experiment dataset. We compared six well-known baselines, including the current state-of-the-art baseline. The experimental results show that our model is better than the current state-of-the-art baseline.

This section mainly introduces some related concepts of traffic forecasting and puts forward some challenges in some previous work. In the second section, we introduce in detail the research of some scholars in this field. In the third section, we introduced some basic concepts of traffic forecasting, and then reviewed DeepSTN+ [27]. In the fourth section, we mainly introduce our AttDeepSTN+ model and related details. In the fifth section, we mainly show our experimental process and experimental results, and analyze the experimental results and the influence of hyperparameters. At the end, we summarized this article and the next steps that can be continued in the future.

2 Related work

In this section, we briefly introduce the algorithm of traffic flow prediction from two perspectives: the traditional time series method and the method based on deep learning.

2.1 Traditional time series method

In the field of time series, Auto-Regressive Integrated Moving Average [46], Kalman filter and their variants [15, 22, 29] have been widely used in traffic forecasting problems, but they are only suitable for stationary time series. In recent studies, scholars have also taken further steps. Some external factors are combined with.

traffic data to improve the accuracy of model predictions [35]. In addition, some studies have also combined spatial information for modeling [42]. However, these traditional models fail to capture the nonlinear spatio-temporal correlation.

2.2 Method based on deep learning

The attention mechanism has shown excellent results in image processing, speech recognition and natural language processing. Its essence is a set of weight coefficients that are learned independently through the network and use dynamic weighting to emphasize what we are interested in. The mechanism of suppressing irrelevant background areas at the same time. In recent years, the attention mechanism has gradually appeared in the deep learning methods of traffic prediction and has achieved great success. In ASTGCN [14], the author proposes a new attention-based spatio-temporal graph convolutional network model to solve the traffic flow prediction problem. In STDN [45], a flow gating mechanism is introduced to learn the dynamic similarity between locations, and a cyclically shifted attention mechanism is designed to deal with long-term periodic time shifts. A self-attention network is used in STAWnet [38] to capture the dynamic spatial dependence between different nodes. Some researchers have also proposed a graphical multi-attention network: GMAN [54] to predict the traffic conditions at different positions in the front time step on the road network map. To this end, this paper proposes an interactive attention mechanism to learn the similarities and influence levels between hourly, daily, and weekly. This mechanism uses the attention mechanism to better capture the complex correlation and influence between traffic data.

With the popularity of deep learning and major successes in natural language processing, computer vision and other fields, some deep learning-based models provide an effective way to capture this complex nonlinear spatiotemporal

correlation. For this reason, some models based on deep learning have emerged. ConvLSTM [36], deep learning hybrid framework [10], STRCNs [19], these models all combine LSTM and convolutional neural networks, Periodic-CRN [55] by combining the pyramid convolution model with periodic representation to simulate the complex periodic nature of crowd flows, these deep learning-based models are based on recurrent neural networks and convolutional neural networks. However, due to the structure of recurrent neural networks it is recursive, which means that it will inevitably lead to time-consuming model training and unable to capture long-term time dependence. Deep-ST [48] is the first model that uses convolutional neural networks to capture complex spatiotemporal correlations. In addition, ST-ResNet [49] proposes a model based on advanced residuals. The difference network replaces the simple convolution operation in the previous model, and in DeepSTN+ [27], a structure is proposed to replace the ordinary convolution, thereby capturing the long-distance spatial dependence, and adding the structure of combining PoI information and time weights and proposes a new fusion mechanism to reduce model parameters and make the structure more stable. These convolutional neural network-based models capture temporal correlation by modeling temporal closeness, periodicity, and trend, but they do not consider the complex correlation between closeness, trend, and periodicity, and more to be precise, it is the relevance and importance of closeness to periodicity and trend. As Fig. 2 shown.

It can be seen from the figure that the historical flow of different periods has different influences on the traffic flow at different moments in the future, and the corresponding moments have a greater degree of influence. In short, the complex correlation between different feature levels should be considered. For this reason, in this article, we propose a mechanism based on interactive attention for the correlation and importance of these three different levels of features. For modeling, in order to further capture the degree of association between them, we also designed a fusion mechanism to reduce the parameters model and make it more stable.

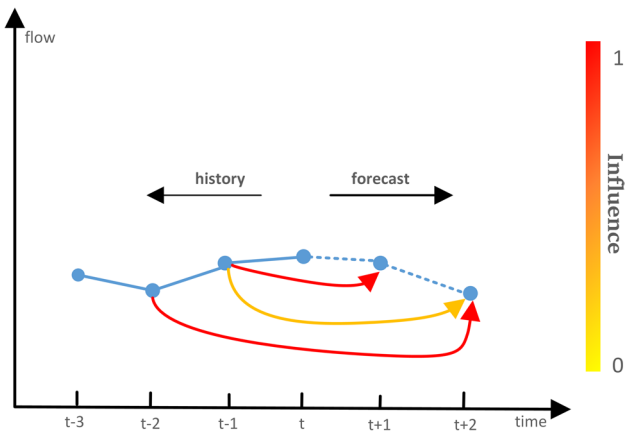


Fig. 2 The impact of time on traffic flow

3 Preliminaries

In this section, we formally introduce the problem of crowd flow prediction and briefly review DeepSTN+ [27] as background knowledge.

3.1 Problem formulation

Definition 1 In order to better represent the area of the city, we divide the entire city into an $H * W$ grid according to the longitude and latitude, and each grid represents each area in the city, and each grid is the size equal.

Definition 2 (Inflow and Outflow): In order to better represent the flow of people in the city, we define the inflow and outflow as follows:

$$x_t^{h,w,in} = \sum_{Tr \in P} \left| \{ i > 1 | g_{i-1} \notin (h,w) \backslash \& g_i \in (h,w) \} \right| \quad (1)$$

$$x_t^{h,w,out} = \sum_{Tr \in P} \left| \{ i \geq 1 | g_{i-1} \in (h,w) \backslash \& g_i \notin (h,w) \} \right| \quad (2)$$

Here P represents the set of trajectories in the t-th time interval, $T_r : g_1 \rightarrow g_2 \rightarrow \cdots g_{Tr}$ represents the trajectory in P, $\boldsymbol{g_i}$ is the geospatial coordinates, $\boldsymbol{g_i} \in (\boldsymbol{h}, \boldsymbol{w})$it means that the geographic spatial coordinates of g_i are in the area (h, w), and vice versa means it is not in the area (h, w).

Crowd flow prediction Given historical data $\{X_i | i=1, 2, \ldots, n - 1\}$, prediction X_n.

3.2 DeepSTN+

The DeepSTN+ [27] model framework includes four parts, closeness, trend, periodicity, and the combination of PoI information and time, and other external information. These four parts are first passed through an ordinary convolution operation. The PoI information and crowd flow information are in the early-stage fusion, and then transfer them to the unit. In order to capture the long-distance spatial dependence, a structure is designed in the unit. In the structure, the author separates a part of the channel, and captures the long-distance dependence through a large-scale convolution kernel to capture the area information at a long distance. Then integrate them together, through another ordinary convolutional layer, and finally merge with the previous input. After coming out of the unit, the units of different levels are fused together by features, and finally through the activation function tanh. The value is mapped to [-1,1]. The external

information contains the distribution information of PoI, including nine PoI information. This nine PoI information are in the form of nine matrices of size H * W, where H and W are the length and width of the entire city, then combine it with time information, assign corresponding weights, and combine with the remaining three components.

4 AttDeepSTN+

Figure 3 shows the framework of our model. As can be seen from the figure, the input is mainly divided into four parts: closeness, period, trend, and information about PoI combined with time. Inflow and outflow are calculated every hour for each area. In order to form a series of crowd flow graphs, these input data are standardized to [-1,1] during data preprocessing. We choose three-time steps for closeness, and each time step contains two channels (inflow and outflow), so there are six channels, and its tensor size is $N * 6 * H * W$, four time steps are selected for periodicity and trend, so there are eight channels, and its tensor size is N * 8 * H * W, where N represents the number of samples, W and H represent the length and width of the entire map respectively. First, through a convolution operation, make their channel numbers the same, and then send the closeness, periodicity, and trend to the interactive attention layer after convolution. In the interactive attention layer, we also use feature fusion to further capture the complex correlation

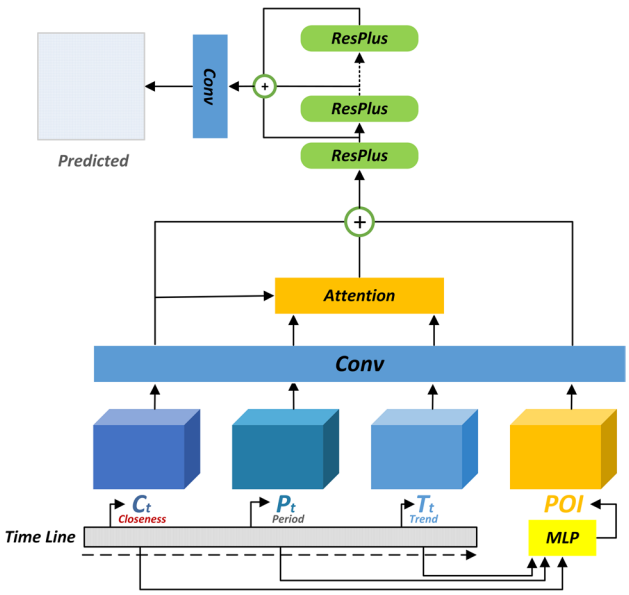


Fig. 3 AttDeepSTN+ structure, where MLP represents multi-layer perceptron, C_t, P_t and T_t respectively represent the time steps of closeness, period, and trend selection, POI represents the Point of Interests information, and Conv represents it is the convolution module, Attention represents the interactive attention layer, and ResPlus is the ResPlus unit

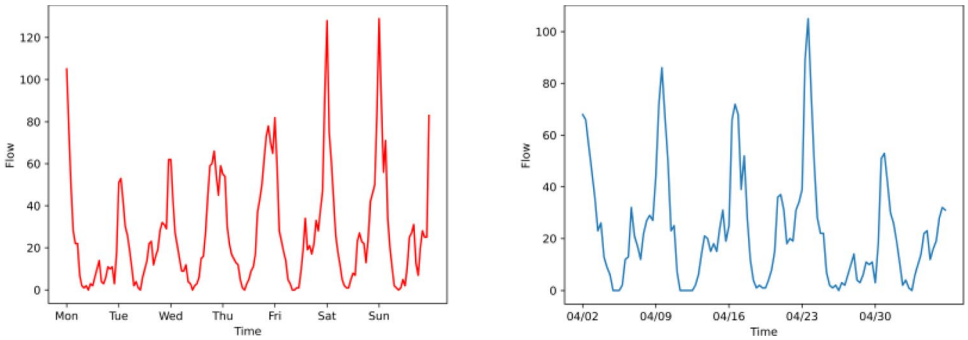
between features at different levels. The components and closeness, PoI information and time weighted information after the interactive attention layer are combined to capture the complex correlation between features at different levels through a feature fusion mechanism, and then sent to the structure, and finally through feature fusion, the complex correlation between different levels is fully extracted, thereby improving the prediction accuracy of the model. Finally, through a convolution operation and then it is mapped to [-1,1] through the tanh function and the final output tensor size is 2 * H * W The details of the interactive attention layer and the feature fusion mechanism will be introduced below.

4.1 Interactive attention

The attention mechanism has shown excellent results in image processing, speech recognition and natural language processing. Its essence is a set of weight coefficients that are learned independently through the network and used dynamic weighting to emphasize what we are interested in. The mechanism of suppressing irrelevant background areas at the same time. In recent years, the attention mechanism has gradually appeared in the deep learning methods to traffic prediction, and breakthroughs have been made. Traffic flow data has an obvious daily cycle and weekly cycle. The adjacent time has a certain degree of influence on the daily cycle and the weekly cycle. For example, if a region has a large flow of people in a certain time interval today, then the flow of people will also show similarity during the same time interval yesterday. The same time interval last week will also show similarities, which were not considered in the previous study. In order to solve the above problems, the complex correlation between the three different levels should be clearly modeled. This paper proposes an interactive attention mechanism. The attention in this paper is based on a variant of channel attention. As can be seen from the figure below, the traffic data has obvious daily and weekly cycles, and there are complex correlations and similarities among the characteristics of different levels of closeness, cycles, and trends, which were not considered in previous models.

Figure 4(a) and (b) respectively show the traffic flow data between different dates and different weeks, which is the periodicity and trend proposed in this article. The points on the broken line are the closeness of this article. We can find the closeness and there is a certain similarity between periodicity and trend. For this reason, we propose an interactive attention layer to capture the complex correlation between these three different levels.

Figure 5 shows the interactive attention layer we designed. In this layer, we regard closeness as context, combine it with periodicity and trend respectively, and dynamically learn the weight each of them during model training, which indicates



(a) Daily periodicity of time

(b) Weekly periodicity of time

Fig. 4 Daily periodicity and Weekly periodicity of time. (**a**) Daily periodicity of time; (**b**) Weekly periodicity of time

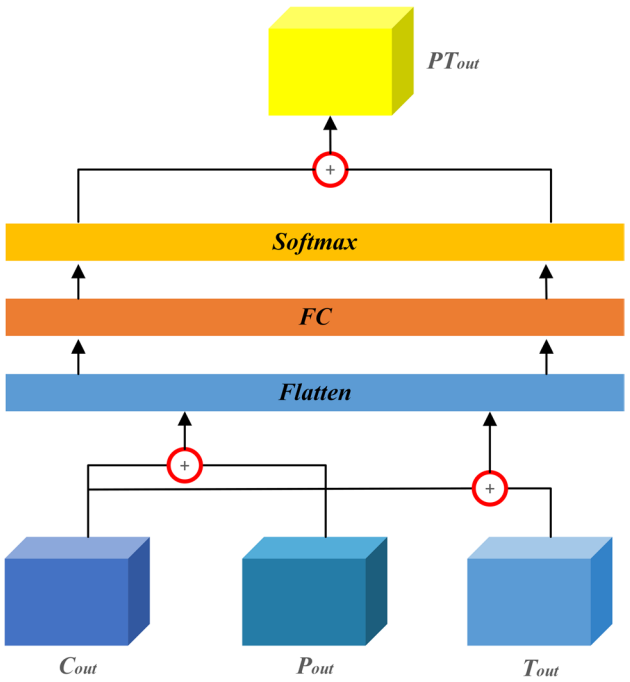


Fig. 5 Interactive attention layer structure, where C_{out}, P_{out}, T_{out}, respectively represent proximity, periodicity, and trend are obtained through convolution operation, Flatten refers to the stretching operation, FC refers to the fully connected layer, and softmax refers to the softmax activation function operating

the corresponding importance and correlation between the two components, and then through the SoftMax function operation, the purpose is to make its value normalized to [0,1], and the sum is 1, and finally the fusion mechanism is used the components are respectively multiplied and combined with the corresponding weights obtained. The formula of this method is as follows:

$$CP_{score} = W_{cp}(Flatten[C_{out}, P_{out}]) \tag{3}$$

$$CT_{score} = W_{ct}(Flatten[C_{out}, T_{out}]) \tag{4}$$

$$\alpha_p = \frac{exp(CP_{score})}{exp(CP_{score}) + exp(CT_{score})} \tag{5}$$

$$\alpha_t = \frac{exp(CT_{score})}{exp(CP_{score}) + exp(CT_{score})} \tag{6}$$

$$PT_{out} = \alpha_p * P_{out} + \alpha_t * T_{out} \tag{7}$$

Among them, W_{cp} and W_{ct} are the parameters that can be learned during the full connection process, α_p and α_t are the learned weights, they respectively represent the importance and correlation between periodicity, trend and closeness. These parameters can all be learned during training. The periodicity and trend are weighted and merged according to the degree of importance, in order to merge the different characteristic levels with each other, so as to be able to capture the complex correlation between them.

4.2 Multi feature fusion

When the three components enter the attention layer, we replace the simple linear combination, because closeness, periodicity, and trend should have more complex interactions. In order to model these interactions, we introduced a more effective fusion method is to fuse the features obtained through the attention layer, and then fuse them with the weighting and closeness of PoI information and time. Compared with the fusion of the DeepSTN+ [27] model, the number of parameters is reduced and further captured the complex correlation between different feature levels is further improved, the prediction accuracy is further improved, the complex structure of the model is greatly simplified, and the network is optimized. In the end, in order to consider further feature fusion, we also set a layer of final fusion method at the end, the purpose is to further capture the feature correlation. The entire fusion method can be expressed as follows:

$$\hat{X} = f_{mul}\left(f_{Res}\left(f_{conv}\left(f_{att}\left(X^{c} + X^{p} + X^{t}\right) + X^{poi}\right)\right)\right) \quad (8)$$

Here the function f_{att} represents the fusion in the interactive attention layer, and the function f_{conv} represents the convolution operation, in order to combine them, the number of channels is converted to the same number for fusion, thus entering the unit. The function f_{Res}means that pass the ResPlus unit, and the function f_{mul} means Use the final multiple fusion mechanism at the tail. X^{c}, X^{p}, X^{t} respectively represent the initial closeness, periodicity, and trend input, X^{poi} represents the information combining PoI information and time.

5 Experiment

In this chapter, we mainly introduce the used data sets, baselines and evaluation standards, and analysis of experimental results.

5.1 data set

BikeNYC This data set comes from the New York Bike System. It records data from April 1st to September 30th, 2014. It mainly includes travel time, start time and end time, start station and end station. This data map size is 21 * 12, and the time interval is one hour. Among these data, we chose the data of the last 14 days as our test set, and the rest as the training set. Based on this data set, we also added some subdata sets of PoI information distribution, mainly including some location function, location information, as shown in the Table 1.

5.2 Baselines

HA This model predicts the inflow and outflow at the corresponding time in the future based on the average value of the inflow and outflow at all corresponding moments in the history.

VAR [15] The Vector Auto-Regressive model predicts the corresponding crowd flow by capturing the paired relationship between all streams. It is a more advanced spatio-temporal model, but due to its large number of parameters, it leads to huge computational costs. The confidence level we choose is 95 %, epochs are 1000.

ARIMA [46] Auto-Regressive Integrated Moving Average is a combination of autoregressive, moving average and difference process. It is a well-known model used to predict future time series. AR is autoregressive, MA is moving average, p and q are the corresponding orders, and d is the number of differences made when the time series becomes stationary. Where $d = 1$, $p = 2$, $q = 1$.

ConvLSTM [36] ConvLSTM is a neural network composed of convolution and LSTM. It captures spatial information through convolution.

and time information through LSTM. The patch size to 4×4 so that each 64×64 frame is represented by a $16 \times 16 \times 16$ tensor. The 1-layer network contains one ConvLSTM layer with 256 hidden states, the 2-layer network has two ConvLSTM layers with 128 hidden states each, and the 3-layer network has 128, 64, and 64 hidden states respectively in the three ConvLSTM layers. All the input-to-state and state-to-state kernels are of size 5×5. The batch size is 32. However, due to the recursive structure, training is relatively time-consuming.

ST-ResNet [49] It is a spatio-temporal traffic prediction model based on convolutional neural network. An advanced residual network is proposed to replace the simple convolution operation in the previous model. The convolutions of Conv1 and all residual units use 64 filters of size 3×3, and Conv2 uses a convolution with 2 filters of size 3×3.

DeepSTN+ [27] This model is also a spatio-temporal flow prediction model based on convolutional neural networks, which constructs a new long-term spatial dependence and uses information combined with PoI information and time to represent the impact of location attributes on traffic flow, the number of channels of convolution and the structure ConvPlus (a convolution) in DeepSTN+ is 64, all channels in ConvPlus are 64. Separated channels in ConvPlus are 8, the number of ResPlus units as 2 in DeepSTN+ model representing the current state-of-the-art baseline.

Table 1 categories of PoIs for BikeNYC

DataSet	Point of Interests (PoI)
BikeNYC	Food, Residence, ShopServic, CollegeUniversity, NightlifeSpot, TravelTransport, ArtEntertainment, ProfessionalOtherPlace, OutdoorsRecreation

5.3 Metrics

In this article, we use Root Mean Square Error (RMSE) and Mean Absolute Error (MAE) as evaluation indicators:

$$RMSE = \sqrt{\frac{1}{T}\sum_{i=1}^{T}\left(X_i - \hat{X}_i\right)^2} \tag{9}$$

$$MAE = \frac{1}{T}\sum_{i=1}^{T}\left|X_i - \hat{X}_i\right| \tag{10}$$

Among this formula, X_i represents the real traffic in the i-th time interval, and $\hat{X}_i$ represents the i-th time interval. The predicted traffic in a time interval, T represents the total number of samples on the test set.

5.4 Convergence research

We first conducted a convergence study when training the model. It can be seen from Fig. 6(a) and (b) that during the training and verification process, loss, rmse, and mae all decrease with the increase in the number of trainings, and gradually converge, which shows the effectiveness of our model and stability.

5.5 Experiment and hyperparameter setting

We divide the entire city into a 21 * 12 rectangular grid, each grid represents an area in the city, the length of each time interval is set to 1 h, and the inflow sum is calculated once every hour. The outflow is used to form a series of crowd flow graphs. We use min-max normalization to convert the traffic flow into [-1,1]. The experiment uses the last 14 days of data as the test set, and the remaining days as the training set.

We set the hyperparameters according to the performance of the verification set. It can be seen from Table 2, our batch number is set to 32, the optimizer chooses Adam [20], the learning rate is set to 0.0002, and we set 350 epochs. We choose 3-time steps for closeness, and 4-time steps for periodicity and trend. Each time step contains 2 channels (inflow and outflow). The number of convolution kernels we choose is 64. The number of our convolution kernels in the hidden layer is 64. We also used BN and dropout [37]. We have selected 9 PoI information. Our computer configuration: graphics card is RTX3060, RAM (random access memory) is 16G, processor is AMD (Advanced Micro Devices,) Ryzen 7 5800 H with Radeon Graphics 3.20 GHz.

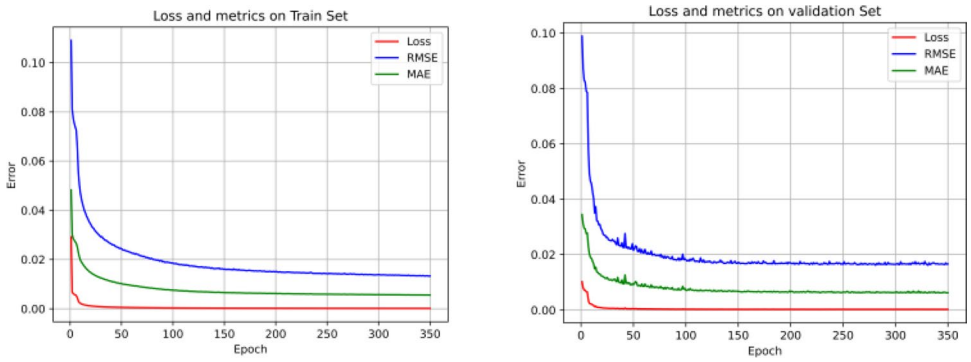
Table 2 Settings of parameters

Parameter	Number
Map size	21 * 12
Time interval	one hour
Number of convolution kernels	64
Categories of PoIs	9
Number of ResPlus units	2
Separated channels in ConvPlus	8
Epochs	350
Legth of closeness	3
Legth of period	4
Legth of trend	4

Our entire algorithm can be expressed as follows: in order to obtain better predicting results, we first normalized the data of past historical moments and the PoI distribution to make it into four parts: closeness, period, trend and PoI, and then learn the parameters through our Att-DeepSTN+ model and continue to perform backpropagation and optimizer adjustments until the model is overfitted. Finally, what we output is a forecast of the inflow and outflow of the entire city at a certain time.

5.6 Experimental results and analysis

Table 3 shows the performance of our model and baseline on the BikeNYC dataset. Among them, +mul represents the use of multiple fusion mechanisms at the end of the model. The notation Δ indicates the reduction of error compared



(a) training

(b) verification

Fig. 6 Change curve of loss, rmse and mae during training and verification

Table 3 Comparison among different baselines and variants of AttDeepSTN+ on BikeNYC

Model	RMSE	Δ	MAE
HA	7.885	31.7 %	2.823
VAR	10.097	68.7 %	5.49
ARIMA	10.894	82.1 %	3.246
ConvLSTM	6.412	7.15 %	2.543
ST-Resnet	6.475	8.21 %	2.395
DeepSTN	6.213	3.82 %	2.388
DeepSTN+	5.984	0	2.292
AttDeepSTN(ours)	**5.836**	**-2.47 %**	**2.253**
AttDeepSTN+mul(ours)	**5.795**	**-3.16 %**	**2.241**
5-fold cross-validation	**5.84)**±(0.1	**-2.41 %**	**2.25**±(0.03)

with DeepSTN+. From Table 3, we can see that our AttDeepSTN+ model has achieved the best performance on both the RMSE and MAE evaluation indicators. It can be seen from the table that some traditional machine learning and time series models cannot achieve ideal prediction results, which indicates that traditional machine learning methods cannot well capture complex nonlinear spatio-temporal correlations. In contrast, some are based on models of deep learning methods can achieve better prediction results than traditional models. Among them, ConvLSTM, ST-ResNet, DeepSTN+ and our model AttDeepSTN+ are all prediction models based on time correlation and spatial correlation. AttDeepSTN+ can achieve better predictions than DeepSTN+, and the results show that the interactive attention layer we designed and the fusion mechanism in it can effectively capture the complex correlation between the non-hierarchical features, indicating the superiority of the model. The addition of multiple feature fusion at the tail reduces the model's prediction error, which further improves the model's prediction accuracy, and can better capture the dynamic changes of traffic data. Compared with DeepSTN+, after adding the interactive attention layer, our model has been improved by 2.47 %, it shows that our interactive attention layer can effectively capture different levels of complex correlation, and after the final fusion, our model has been improved by 3.16 %. In order to further demonstrate the stability and effectiveness of our model, we have added a five-fold cross-validation procedure. It can be seen from the table that the average value of the five-fold cross-validation is lower than the current optimal baseline DeepSTN+, it obtained an enhancement of 2.41 % on average, which indicates that our model is better than DeepSTN+. Based on the real data set, the model obtained an enhancement of 3.16 %, showing the superiority of our model.

In order to further show that the components we designed are effective, we have conducted numerous variant experiments for this purpose. The experimental results are shown in the Table 4:

In Table 4 DeepSTN+ represents the original comparison model, AttDeepSTN+ represents the model with only the interactive attention layer added to the original, DeepSTN+mul represents only the fusion of multiple tail features on the original basis, and AttDeepSTN-con represents there is no fusion in the interactive attention layer. It can be seen from the table that the effectiveness of the interactive attention and feature fusion mechanism we designed. First, we compared the original model. Based on the original model, we added an interactive attention layer, and the error was reduced, indicating the effectiveness of the attention mechanism. Based on the original model, we added the fusion of multiple features at the tail. The reduced error indicates the effectiveness of multiple feature fusion at the tail. The error becomes larger when no feature fusion is performed in the interactive attention layer than when the feature fusion is performed, which indicates the effectiveness of feature fusion.

5.7 Hyperparameter effect

As shown in Fig. 7(a) and (b), the x-axis is the number of feature levels, and the y-axis is the evaluation index. We studied the influence of different number of feature levels for tail multiple fusions on the experimental results. We found that the two evaluation indicators RMSE and MAE have the same changing trend as the feature level changes. When the number of different feature levels increases from two to three, RMSE and MAE changed significantly, but as the number of feature levels continued to increase, the changes were not significant. The experimental error also increases with the increase in the number of feature levels. Among them, we find that the fusion of two feature levels is the best. This may be due to the redundancy of the network structure due to too many parameters, which affects the experimental results.

6 Conclusions

Our new model based on the interactive attention mechanism is used for traffic flow prediction. It uses the attention mechanism to capture the complex correlations between

Table 4 Variant experiment results based on AttDeepSTN+

Model	RMSE	MAE
DeepSTN+	5.984	2.292
AttDeepSTN	5.836	2.253
AttDeepSTN+mul	5.795	2.241
AttDeepSTN-con	5.965	2.266

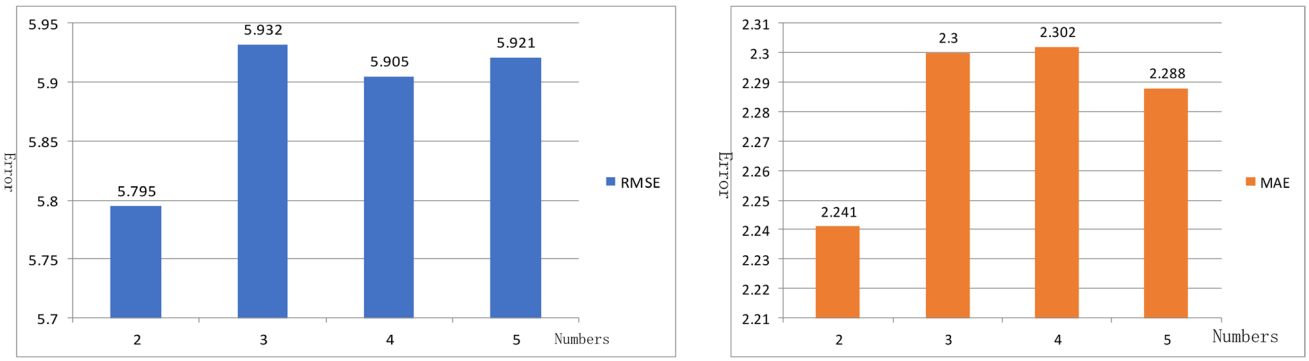


Fig. 7 The effect of tail multiple fusions with different number of feature levels on experimental results

features at different levels, plus an appropriate fusion mechanism to make the model prediction more accurate. In New York City crowd flow experiments show that the model is better than the current state-of-the-art baseline, confirming that our model is more suitable for traffic flow prediction. In the future, we will further subdivide the interactive attention layer. On the other hand, due to the many factors affecting crowd flow, we will further consider other external factors and dynamically adjust the size of the convolution kernel to capture the correlation between regions.

Acknowledgements This research was supported by the National Natural Science Foundation of China(No.62062033) and the Science and Technology Research Project of the Education Department of Jiangxi Province (200604) and the Natural Science Foundation of Jiangxi Province under Grant No.20192ACBL21006 and the Key Research & Development Plan of Jiangxi Province No.20203BBE53034.

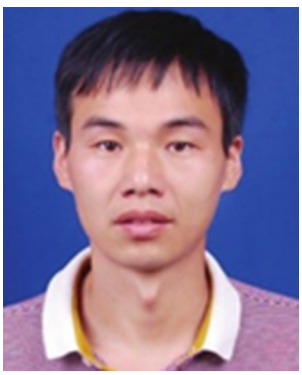
References

1. Abbasimehr H, Shabani M, Yousefi M (2020) An optimized model using lstm network for demand forecasting. Comput Ind Eng 143:106435
2. Chang Z, Zhang Y, Chen W (2019) Electricity price prediction based on hybrid model of adam optimized lstm neural network and wavelet transform. Energy 187:115804
3. Chang YS, Chiao HT, Abimannan S, Huang Y, Tsai YT, Lin KM (2020) An lstm-based aggregated model for air pollution forecasting. Atmos Pollut Res 11:1451–1463
4. Chen Y (2020) Voltage's prediction algorithm based on lstm recurrent neural network. Optik 220:164869
5. Chu KF, Lam AYS, Li V (2020) Deep multi-scale convolutional lstm network for travel demand and origin-destination predictions. IEEE Trans Intell Transp Syst 21:3219–3232
6. Cui Z, Ke R, Wang Y (2018) Deep bidirectional and unidirectional lstm recurrent neural network for network-wide traffic speed prediction. ArXiv abs/1801.02143
7. Cui Z, Henrickson KC, Ke R, Wang Y (2020) Traffic graph convolutional recurrent neural network: A deep learning framework for network-scale traffic learning and forecasting. IEEE Trans Intell Transp Syst 21:4883–4894
8. Ding Y, Zhu Y, Feng J, Zhang P, Cheng Z (2020) Interpretable spatio-temporal attention lstm model for flood forecasting. Neurocomputing 403:348–359
9. Kuang, L., Zheng, J., Li, K., & Gao, H. (2021). Intelligent Traffic Signal Control Based on Reinforcement Learning with State Reduction for Smart Cities. ACM Transactions on Internet Technology (TOIT), 21, 1 - 24
10. Du S, Li T, Gong X, Yang Y, Horng S (2017) Traffic flow forecasting based on hybrid deep learning framework. 2017 12th International Conference on Intelligent Systems and Knowledge Engineering (ISKE) pp 1-6
11. Du B, Hu X, Sun L, Liu J, Qiao Y, Lv W (2021) Traffic demand prediction based on dynamic transition convolutional neural network. IEEE Trans Intell Transp Syst 22:1237–1247
12. Gao, H., Huang, W., & Yang, X. (2019). Applying Probabilistic Model Checking to Path Planning in an Intelligent Transportation System Using Mobility Trajectories and Their Statistical Data. Intelligent Automation and Soft Computing. 25, 547-559
13. Gao H, Liu C, Li Y, Yang X (2021) V2vr: Reliable hybrid-network-oriented v2v data transmission and routing considering rsus and connectivity probability. IEEE Trans Intell Transp Syst 22:3533–3546
14. Guo S, Lin Y, Feng N (2019) Attention based spatial-temporal graph convolutional networks for traffic flow forecasting. Proceedings of the AAAI Conference on Artificial Intelligence 33:922-929
15. Geurts, M.D., Box, G.E., & Jenkins, G.M. (1976). Time Series Analysis: Forecasting and Control. Journal of Marketing Research, 14, 269
16. Hoque J, Erhardt GD, Schmitt D, Chen M, Wachs M (2021) Estimating the uncertainty of traffic forecasts from their historical accuracy. Transp Res Part A-Policy Pract 147:339–349
17. Hu J, Li B (2020) A deep learning framework based on spatio-temporal attention mechanism for traffic prediction. 2020 IEEE 22nd International Conference on High Performance Computing and Communications; IEEE 18th International Conference on Smart City; IEEE 6th International Conference on Data Science and Systems (HPCC/SmartCity/DSS), pp 750-757
18. Ji J, Wang J, Jiang Z, Ma J, Zhang H (2020) Interpretable spatio-otemporal deep learning model for traffic flow prediction based on potential energy fields. 2020 IEEE International Conference on Data Mining (ICDM), pp 1076-1081
19. Jin W, Lin Y, Wu Z, Wan H (2018) Spatio-temporal recurrent convolutional networks for citywide short-term crowd flows prediction. In: ICCDA, 2018

20. Kingma, D. P., & Ba, J. (2014). Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980
21. Kuang, L., Hua, C., Wu, J., Yin, Y., & Gao, H. (2020). Traffic Volume Prediction Based on Multi-Sources GPS Trajectory Data by Temporal Convolutional Network. Mobile Networks and Applications, 25,1-13
22. Li X, Pan G, Wu Z, Qi G, Li S, Zhang D, Zhang W, Wang Z (2011) Prediction of urban human mobility using large-scale taxi traces and its applications. Front Comp Sci 6:111–121
23. Li Y, Yu R, Shahabi C, Liu Y (2018) Diffusion convolutional recurrent neural network: Data-driven traffic forecasting. Learning, arXiv
24. Li T, Zhang J, Bao K, Liang Y, Li Y, Zheng Y (2020) Autost: Efficient neural architecture search for spatio-temporal prediction. Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining
25. Li P, Wang X, Gao H, Xu X, Iqbal M, Dahal K (2021) A dynamic and scalable user-centric route planning algorithm based on polychromatic sets theory. IEEE Transactions on Intelligent Transportation Systems, pp 1–11
26. Liebig T, Piatkowski N, Bockermann C (2017) Dynamic route planning with real-time traffic predictions. Inf Syst 64:258–265
27. Lin, Z., Feng, J., Lu, Z., Li, Y., & Jin, D. (2019). Deepstn+: context-aware spatial-temporal neural network for crowd flow prediction in metropolis. Proceedings of the AAAI Conference on Artificial Intelligence, 33, 1020–1027
28. Lin K, Xu X, Gao H (2021) Tscrnn: A novel classification scheme of encrypted traffic based on flow spatiotemporal features for efficient management of iiot. Comput Netw 190:107974
29. Lippi M, Bertini M, Frasconi P (2013) Short-term traffic flow forecasting: An experimental comparison of time-series analysis and supervised learning. IEEE Trans Intell Transp Syst 14:871–882
30. Luo H, Huang M, Zhou Z (2019) A dual-tree complex wavelet enhanced convolutional lstm neural network for structural health monitoring of automotive suspension. Measurement 137:14–27
31. Ma, X., Zhuang, D., He, Z., Ma, J., & Wang, Y. (2017). Learning traffic as images: a deep convolutional neural network for large-scale transportation network speed prediction. Sensors, 17(4), 818.
32. Moreno SR, da Silva RG, Mariani V, Coelho L (2020) Multi-step wind speed forecasting based on hybrid multi-stage decomposition model and long short-term memory neural network. Energy Convers Manag 213:112869
33. Ribeiro, M.H., & Coelho, L.D. (2020). Ensemble approach based on bagging, boosting and stacking for short-term prediction in agribusiness time series. Appl. Soft Comput., 105837,86.
34. Ribeiro GT, Mariani V, Coelho L (2019) Enhanced ensemble structures using wavelet neural networks applied to short-term load forecasting. Eng Appl Artif Intell 82:272–281
35. Rong L, Cheng H, Wang J (2017) Taxi call prediction for online taxicab platforms. In: APWeb/WAIM Workshops
36. Shi X, Chen Z, Wang H, Yeung D, Wong W, Woo W (2015) Convolutional lstm network: A machine learning approach for precipitation nowcasting. In: NIPS
37. Srivastava N, Hinton GE, Krizhevsky A, Sutskever I, Salakhutdinov R (2014) Dropout: a simple way to prevent neural networks from overfitting. J Mach Learn Res 15:1929–1958
38. Tian C, Chan WK (2021) Spatial-temporal attention wavenet: A deep learning framework for traffic prediction considering spatial-temporal dependencies. IET Intell Transp Syst 15:549–561
39. Wang F, Xuan Z, Zhen Z, Li K, Wang T, Shi M (2020) A day-ahead pv power forecasting method based on lstm-rnn model and time correlation modification under partial daily pattern prediction framework. Energy Convers Manag 212:112766
40. Wang J, Zhu W, Sun Y, Tian C (2020b) An effective dynamic spatiotemporal framework with multi-source information for traffic prediction. ArXiv abs/2005.05128
41. Wu W, Xia Y, Jin W (2021) Predicting bus passenger flow and prioritizing influential factors using multi-source data: Scaled stacking gradient boosting decision trees. IEEE Trans Intell Transp Syst 22:2510–2523
42. XiaoMing S, Qi H, Shen Y, Wu G, Yin B (2020) A spatial-temporal attention approach for traffic prediction. IEEE Transactions on Intelligent Transportation Systems pp 1–10
43. Yang H, Li X, Qiang W, Zhao Y, Zhang W, Tang C (2021) A network traffic forecasting method based on sa optimized arima-bp neural network. Comput Netw 193:108102
44. Yao H, Wu F, Ke J, Tang X, Jia Y, Lu S, Gong P, Ye J, Li Z (2018) Deep multi-view spatial-temporal network for taxi demand prediction. In: AAAI
45. Yao H, Tang X, Wei H, Zheng G, Li Z (2019) Revisiting spatial-temporal similarity: A deep learning framework for traffic prediction. In: AAAI
46. Young P, Shellswell S (1972) Time series analysis, forecasting and control. IEEE Trans Autom Control 17:281–283
47. Yu R, Li Y, Shahabi C, Demiryurek U, Liu Y (2017) Deep learning: A generic approach for extreme condition traffic forecasting. In: SDM
48. Zhang J, Zheng Y, Qi D, Li R, Yi X (2016) Dnn-based prediction model for spatio-temporal data. Proceedings of the 24th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems
49. Zhang J, Zheng Y, Qi D (2017) Deep spatio-temporal residual networks for citywide crowd flows prediction. In: AAAI
50. Zhang J, Zheng Y, Sun J, Qi D (2020) Flow prediction in spatio-temporal networks based on multitask deep learning. IEEE Trans Knowl Data Eng 32:468–478
51. Zhang J, Chen F, Guo Y (2020) Multi-graph convolutional network for shortterm passenger flow forecasting in urban rail transit. Physics and Society, arXiv
52. Zhang S, Chen Y, Zhang W (2021) Spatiotemporal fuzzy-graph convolutional network model with dynamic feature encoding for traffic forecasting. Knowl-Based Syst 231:107403
53. Zhao L, Song Y, Zhang C, Liu Y, Wang P, Lin T, Deng M, Li H (2020) T- gcn: A temporal graph convolutional network for traffic prediction. IEEE Trans Intell Transp Syst 21:3848–3858
54. Zheng C, Fan X, Wang C, Qi J (2020) Gman: A graph multi-attention network for traffic prediction. In: AAAI
55. Zonoozi A, Kim J, Li X, Cong G (2018) Periodic-crn: A convolutional recurrent model for crowd density prediction with recurring periodic patterns. In: IJCAI

Hui Zeng native of Ganzhou, Jiangxi, associate professor,master supervisor. Mainly engaged in the research of database technology andmachine learning. In recent years, he has presided over the completion of 3provincial-level scientific research projects, the backbone has participated inthe completion of a number of National Natural Fund projects, provincial-levelscientific research projects, and has presided over the completion of a numberof enterprise unit horizontal research projects; the research and developmentsystem won the third prize of Jiangxi Science and Technology Progress Award 1item, 1 second prize of Nanchang Science and Technology Progress Award;published many academic papers in EI search and domestic core journals as thefirst author; owns more than 30 software copyrights.

Zhiying Peng graduated from Jiangxi University of Science andTechnology with a major in Information and Computing Science. He is currently apostgraduate in Computer Science and Technology at the School of InformationEngineering, East China Jiaotong University. The main research direction istraffic flow prediction based on deep learning. In 2020, he won the secondprize of Jiangxi Province Mathematical Modeling Competition.



Xiaohui Huang received the BEng degree and the master's degree fromJiangxi Normal University, Nanchang,P.R.China, in 2005 and 2008, respectively,and the PhD degree in the Harbin Institute of Technology, P.R.China, in 2014.Since Dec 2014, he has been at the School of Information EngineeringDepartment, East China Jiaotong University, P.R. China, where he is currentlyan associate professor of Computer Science. His research interests are in theareas of machine learning, deep learning, and clustering algorithm.

Yixue Yang graduated from the School of Information Technology ofJingdezhen Ceramic University, majoring in digital media technology, and iscurrently a postgraduate of the School of Software, East China JiaotongUniversity. The main research direction is natural language processing. In2020, he won the second prize of Jiangxi Province Mathematical ModelingCompetition.

Rong Hu from Nanchang, Jiangxi, studying at the School ofInformation Engineering, East China Jiaotong University. The main researchdirections are machine learning, deep learning, and reinforcement learning. Wonthe third prize in the Central China Division of the China PostgraduateElectronic Design Competition.