Research on Big Data-Driven Urban Traffic Flow Prediction Based on Deep Learning

Xiaoan Qin, Anhui Business College of Vocational Technology, China*

ABSTRACT

This paper introduces an innovative approach for the urban traffic flow prediction (TFP) that utilizes big data and deep learning (D-L) to improve accuracy, reducing the incidence of large errors commonplace in traditional methods. By implementing this method, sustainable urban developments are able to be achieved more effectively in the future. First, an Attention-CNN-GRU-ResNet (ACGR) TFP model is built with the D-L network by gridding the urban traffic flow (TF) into a three-dimensional S-T tensor sequence. An attention-based GRU is then introduced to combine spatial and channel attention in the traditional GRU, and the time dependence and spatio-temporal (S-T) heterogeneity of TF in each subset are effectively extracted. Finally, a ResNet module is introduced to capture the S-T dependency, which helps avoid the deep network degradation caused by excessive layers. Results show the proposed method generates the minimum value in RMSE, MAE, and MAPE with 18.32, 10.66, and 5.34, respectively. This research provides a new idea to alleviate data sparsity and consider the difference of input features and offers a novel approach to solve the S-T learning tasks associated with modeling.

KEYWORDS

Attention Mechanism, D-L, GRU, ResNet, Sustainable, TFP, Urban Traffic

INTRODUCTION

With the development of Internet technology and the popularization of location-based services, researchers have discovered valuable knowledge through the analysis and mining of urban S-T big data, which has promoted people's lives, improved urban operation efficiency, reduced resource consumption, and achieved sustainable urban development (Wang et al., 2020; Li. 2021; Cheng et al., 2019). In recent years, significant advances have been made in artificial intelligence-based algorithms, which can make accurate predictions of complex problems with minimal domain knowledge and strong generalization ability, and AI-based algorithms have found a wide range of applications. These characteristics lay a foundation for studying traffic flow (TF) prediction (Li et al., 2021; Shanshan et.al., 2022). As an important part of urban S-T big data mining, TFP plays a key role in developing cities and intelligent transportation. Transportation, a field crucial to daily life, can help promote high-quality social and economic development. To achieve safe, comfortable, convenient, and green transportation is a key link to improving people's sense of security and happiness and ensuring social stability (Liu et al., 2020; Huang 2019; Nguyen et al., 2020).

DOI: 10.4018/IJITSA.323455
*Corresponding Author

The surge in private cars will inevitably lead to traffic congestion. The toxic substances emitted by vehicles in the process of congestion will not only damage people's health, induce various diseases, and affect life and work efficiency but also cause issues such as environmental pollution, resource waste, and economic losses (Zhang et al., 2019; He, 2020; Zhang et al., 2020). Accurate TFP can help analyze road planning, recommend more intelligent travel routes, and reduce traffic accidents. At the same time, if we can realize the urban TFP we can provide a reference for residents to travel, avoid hot traffic areas, ease traffic pressure, reduce traffic control pressure, and improve travel efficiency (Cheng et al., 2021; Cui 2020; Wang et al., 2020).

The parameter method with statistics is a general TFP method, and the Auto-Regressive Integrated Moving Average (ARIMA) model is a typical parameter method. In the 1970s, Ahmed and Cook (1979) first used ARIMA in TFP field to predict short-term TF of the expressway. Later, scholars put forward a variety of improved models, such as KARIMA (Voort et al., 1996), SARIMA (Williams & Hoel, 2019), and STARIMA (Kamarianakis & Prastacos, 2003).

In addition, as an important part of urban S-T big data mining, TFP is key to developing cities and intelligent transportation and has far-reaching practical significance. For example, accurate TFP can help analyze road planning, recommend more intelligent travel routes, and reduce traffic accidents. The existing methods to solve the urban TFP mainly fall into two types. The first is traditional statistical prediction methods, such as Moving Average (MA), ARIMA (Williams & Hoel, 2020), etc., and this method only applies to linear data. However, the traditional statistical methods cannot learn such complex S-T dependence because TF data is complex and nonlinear with highly complex S-T characteristics. The other method is a data-driven prediction model, which makes predictions by learning the change rules of data. This method makes predictions without considering the dynamic characteristics of traffic scenes. Among them, this method has the best performance at present, such as STResNet (Zhang et al., 2017), STDN (Yao et al., 2019), and ConvLSTM (Shi et al., 2021).

The urban TFP mainly faces the following three new challenges: (1) The data distribution is uneven, and some data are sparse or difficult to obtain. (2) The related urban S-T data learning task is difficult to model. (3) The urban traffic data are in diverse forms, which makes it difficult to express them uniformly with the image matrix. Therefore, new data representation forms need to be introduced, and unique data characteristics have different impacts on the final prediction (Zhang et al., 2019; Du 2020; Zhang & Huang, 2019). How to extract the S-T correlation features of complex road networks, and use D-L and big data technology to achieve accurate short-term TFP, have thus become a hot spot in the industry (Bai et al., 2020; Li, 2021), to provide new technical ideas and effective methods for urban S-T big data analysis. The second part of the article describes the related research, the third part describes the abortion flow prediction model based on deep learning; the fourth part describes the experimental results and analysis; and the fifth part describes the conclusion.

RELATED RESEARCH

Urban TFP is a key component in S-T data mining. With various position sensors used in the era of big data, urban S-T big data have been generated in massive amounts, and the data forms have become more diversified than ever. The great variety in data form and the increasingly complex data correlation pose new research challenges to the existing S-T data mining methods.

Zhang et al. (2021) used D-L to learn the deep S-T features of TF from traffic data and established a combined prediction model of TF GGCN-SA with D-L. The soft attention mechanism was introduced to aggregate the S-T information in different neighborhoods. This method, however, cannot use data to achieve better analysis and prediction of traffic conditions when the data is sparse in the target city. Zhang (2020) proposed a multi-task learning gated recursive unit with residual mapping based on D-L to improve the low accuracy of traffic state prediction of intelligent transportation systems. Accurate prediction was achieved with the introduction of feature engineering. However, it is difficult for this method to effectively correlate multi-modal urban S-T data and perform joint analysis on different analysis tasks. Zhou et al. (2019) proposed a new multi-model integration framework based on D-L by using a stackable automatic encoder to extract the relationship in TF data and fine-tune the architecture, aiming to overcome the weakness that the traditional single TFP model is hardly applicable to different scenarios. This method, however, fails to consider the impact of traffic speed, road occupancy, and other data characteristics on TF. Zhao et al.(2020) studied the applicability of the Time Convolution Network (TCN) in TFP, accurately captured the S-T evolution of TF with TCN, and proposed a D-L framework for short-term city-wide TFP. However, this method reduces the convergence speed while fully using S-T data in the city. Han and Huang (2020) compressed road network data based on correlation analysis and CX decomposition and used the spectral decomposition method to eliminate the impact of TF trend items on accuracy. A new TFP method with D-L was proposed on this basis. However, this method does not consider the spatial features of TF data. Du et al. (2020) realized adaptive learning of S-T and multi-modal traffic data using attention mechanism and gave a hybrid multi-modal D-L method for TFP. However, this method fails to fully consider the correlation between traffic conditions, and the prediction accuracy needs further improvement. Han et al. (2019) proposed a parallel S-T D-L network for highway TFP using CNN to learn TF information characteristics from the time dimension and LSTM to learn TF information characteristics from the spatial dimension. However, this method does not adapt well to the vastly changing traffic situation.

To overcome the weaknesses in the traditional urban TFP methods, namely large errors and inaccurate long-term prediction of the flow of people, this paper proposes a big data-driven urban TFP method based on D-L. The foundation concepts are, first, by gridding urban TF into three-dimensional S-T tensor series, urban TFP is equivalent to multi-dimensional S-T tensor prediction. Second, the ResNet module is introduced to acquire S-T dependencies. Third, the attention-based GRU is introduced to improve performance. Compared with the traditional urban TFP methods, the proposed method is innovative in the following aspects:

1. The translation invariance of CNN structure is used to cut down the model calculation and enhance the prediction efficiency.
2. The introduction of attention-based GRU model, including spatial attention, GRU, and channel attention can capture the time correlation of TF in more detail.
3. To overcome the difficulties caused by excessive layers ResNet module is introduced via identity mapping to calculate residuals, which enhances the model's overall performance.

TFP MODEL BASED ON D-L NETWORK

The TFP Model Framework

The direct prediction used here can predict the future TF volume of T times resolution by inputting historical values into the prediction model. The framework of the proposed urban TFP is shown in Figure 1.

In Figure 1, $\left(x_{t-k}, x_{t-k+1}, x_{t-k+2}, \ldots, x_{t-2}, x_{t-1}, x_t\right)$ represents the historical value and $\left(y_{t+1}, y_{t+2}, y_{t+3}, \ldots, y_{t+T-2}, y_{t+T-1}, y_{t+T}\right)$ represents future TF with T times resolution. Before training the prediction model, any prediction interval T needs to be predefined. The sampling period in the data set can be the same as the predefined prediction interval.

Attention-CNN-GRU-ResNet (ACGR) Model

Urban TF can be gridded into 3-dimensional S-T tensor series $A_{t-k\sim t+k}^{p\times q\times w}$, so urban TFP is multi-dimensional S-T tensor prediction. To solve this problem, an Attention-CNN-GRU-ResNet (ACGR) model is proposed with its framework shown in Figure 2.

Figure 1. Overall urban TFP framework

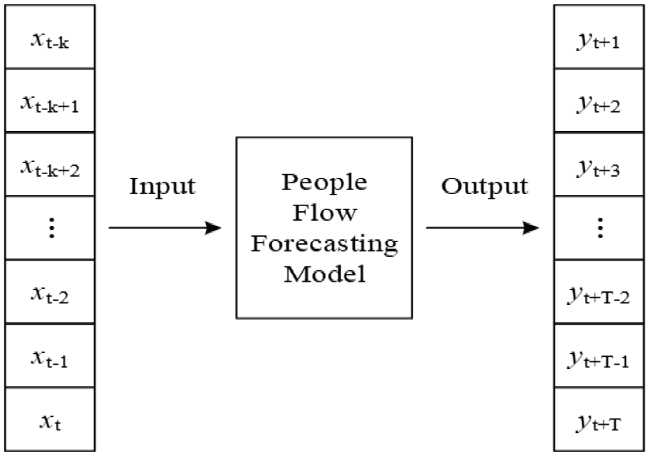
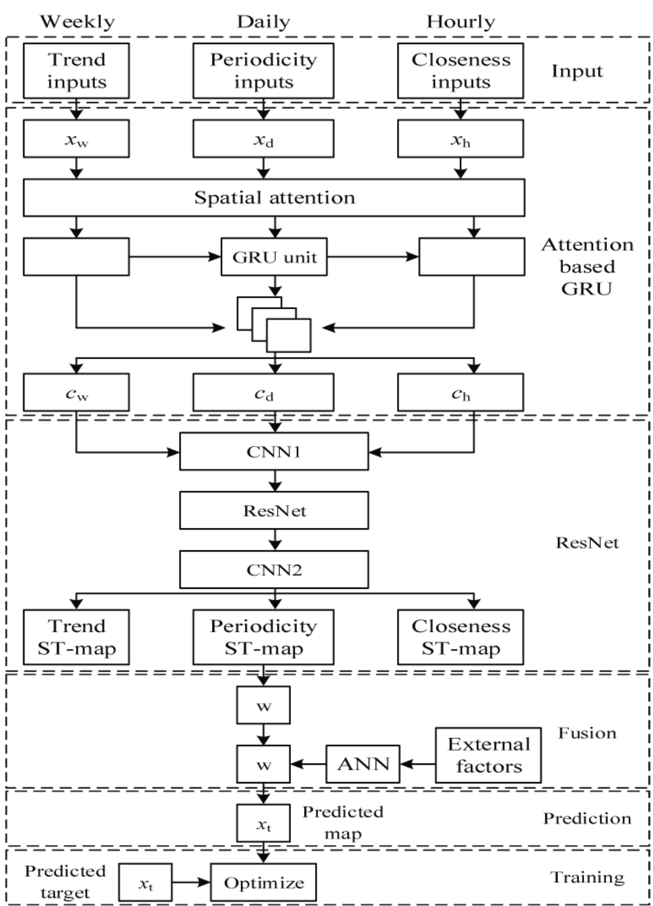


Figure 2. Framework of the ACGR model



The mathematical expression of the three-dimensional S-T tensor prediction problem is shown in formula (1) below:

$$A_t^{p\times q\times w} = G\left(A_{ST}, H\right) \tag{1}$$

In the formula, A_{ST} represents a collection of historical S-T flow tensors. $A_t^{p\times q\times w}$ stands for the prediction target of time t. H is influencing factors. The difficulty of prediction lies in accurately modeling the S-T relationship simultaneously.

As can be seen in Figure 2, the ACGR model is composed of six parts. The input part introduces the time parameter of the adjustment table. The new attention-based GRU can effectively extract information dependence and S-T of TF. ResNet module extracts spatial correlation through convolution operation. In this way, the ACGR model can effectively extract and model the S-T relationship through these six modules.

ACG Model

The attention-based ConvGRU module has three parts: GRU, spatial, and channel attention. Its function is to get the time correlation of TF. The dataset includes hourly, daily, and weekly TF. The ACG model structure of the hourly mode is shown in Figure 3, and the daily mode is like the weekly mode.

The ConvGRU module based on hour attention receives tensors $A_F^{f\times p\times q\times 2}$, including the H hour TF approaching the target time t. In time $t-k$, the feature tensor is expressed as $A_{t-k}^{p\times q\times 2}$. Figure 3 extracts spatial highlighted information by implementing a pooling operation. The application of aggregation operations has proven effective in highlighting information areas. Therefore, two spatial context descriptors can be obtained by using average pooling and maximum pooling operations: $M_{t-k(av)}^{p\times q}$ and $M_{t-k(\max)}^{p\times q}$, representing average pooling features and maximum pooling. The average set feature $M_{t-k(av)}^{p\times q}$, and the maximum set feature $M_{t-k(\max)}^{p\times q}$ are connected to the spatial feature tensor $M_{t-k(s)}^{p\times q\times 2}$. Then, a CNN with an s-shaped activation function generates spatial attention $N_{t-k(s)}^{p\times q\times 2}$. By multiplying the weight matrix and the feature graph. Finally, the spatial attention feature map $A_{t-k(s)}^{p\times q\times 2}$ of time $t-k$ is obtained.

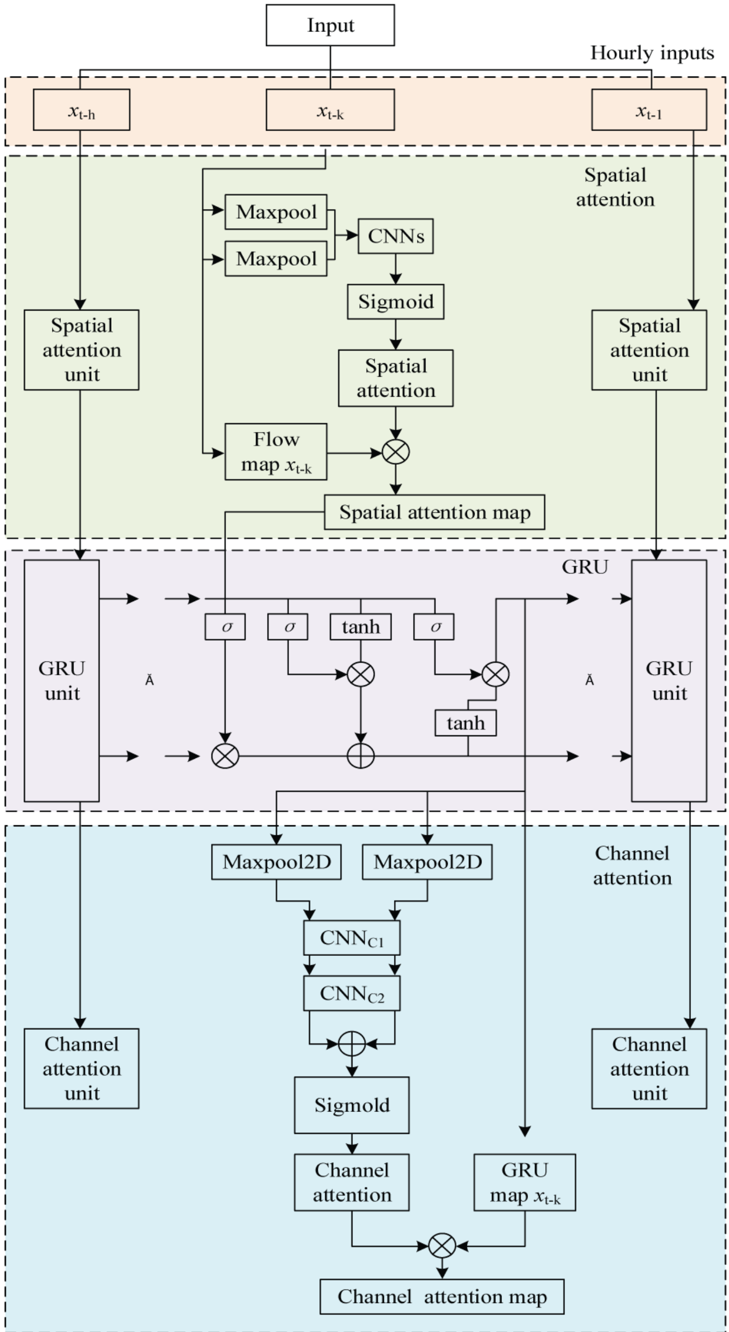
GRU Model

Because the traffic flow data is temporal and vulnerable to the front and back time slices, the RNN network can analyze the sequence data suitable for traffic flow research. The defects of gradient attenuation or explosion in the model of RNN, however, make it difficult to obtain the time dependence with a long time interval in the actual scenarios. The GRU model, a variant of RNN, can solve the above difficulties, and it has a simple structure and high efficiency. The GRU model, therefore, is a better solution to practical problems. Unlike LSTM, the GRU model combines its internal self-circulating cell and hidden layer Hidden. It reconstructs the input and forgetting gates into update units and adds reset gates, thus modifying the LSTM hidden state calculation method. In a simpler model structure, it effectively shortens the model prediction time. Therefore, the differences between LSTM and GRU include: first, LSTM does better control of information flow than GRU; second, LSTM can maintain a longer sequence of information thanks to the transmission of its cell state; third, LSTM is less efficient. The GRU (Gated Recurrent Unit) model is a variation of the RNN. We show the basic principle of the GRU model in Figure 4.

CNN Model

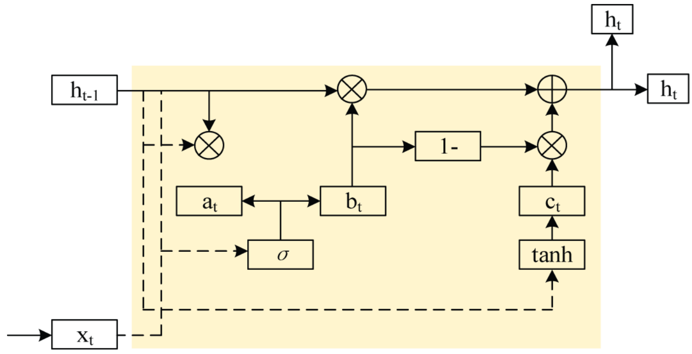
CNN uses the local connection, weight sharing and other methods to solve the problems of general fully connected network, and is effective in reducing the number of parameters and speeding up the

Figure 3. Framework of the ACG model



training speed. Down sampling uses pooling to reduce the number of samples in each layer, thus improving the robustness of the model. For image related tasks, the CNN accelerates the training speed and ensures the processing effect by retaining feature parameters and reducing the number of parameters. However, convolution operation has translation invariance, while pooling operation has

Figure 4. Basic principle of GRU model



local translation invariance, which makes CNN unable to process data in non-Euclidean space and inapplicable to graph structure data.

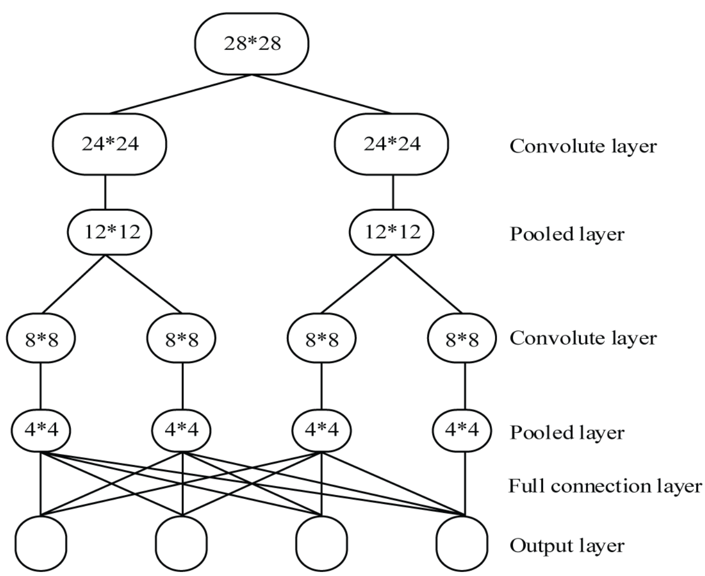
The basic structure of CNN is shown in Figure 5. The general CNN includes convolution, down sampling, and full connection layers. A multi-layer structure is possible with the down sampling layer. The introduction of Residual Learning in ResNet has overcome the degradation in deep networks and enabled researchers to train deeper networks.

ResNet Model

The network degradation happens because the current training method makes finding a good parameter for the deep network difficult.

Some scholars proposed ResNet module to solve the degradation problem. ResNet module can capture S-T dependency and avoid deep network degradation. In addition, ResNet can also mitigate gradient disappearance/explosion by introducing residuals. This process shortens the effective path from loss of input, and directly adds the delta of the endpoint layer of the shortcut to the derivation.

Figure 5. Basic structure of CNN



Compared with VGG and other neural networks, ResNet has a greater advantage of introducing identity mapping because of excessive layers by calculating residuals.

The basic structure of ResNet is shown in Figure 6.

Prediction Model Algorithm Design

After introducing the model's most critical feature modeling unit, the proposed model's process will be explained. The pre-training process uses the source city's training samples to obtain a preliminary urban TFP model. In the training phase, the objective function of model convergence is to minimize the error between the predicted P and the real T. We show the specific training process in Figure 7.

EXPERIMENTAL RESULTS AND ANALYSIS

Experimental Environment Configuration

The environment involved in the experiments includes a personal computer and server in the model training stage. We show the experimental information in Table 1.

Model Super Parameter Settings

We show some super parameter settings in Table 2.

Datasets

Two datasets used in TFP, BikeNYC and TaxiNYC datasets, were selected for experimental verification.

BikeNYC has 9 million bicycle tracks. Among them, there are more than 600 bicycle stations and 10000 bicyclesEach bicycle track contains information such as start-stop, time stamp, longitude and latitude. The data from the first 11 months are used for training and validation; the rest is for testing. TaxiNYC has more than 160 million taxi tracks, with an average of more than 1.3 million taxi tracks per month. Each data contains information such as boarding and alighting time, boarding and alighting locations, and track distance. Again, the data from the first 11 months are used for training and validation, and the rest is used for testing.

Figure 6. Basic structure of ResNet module

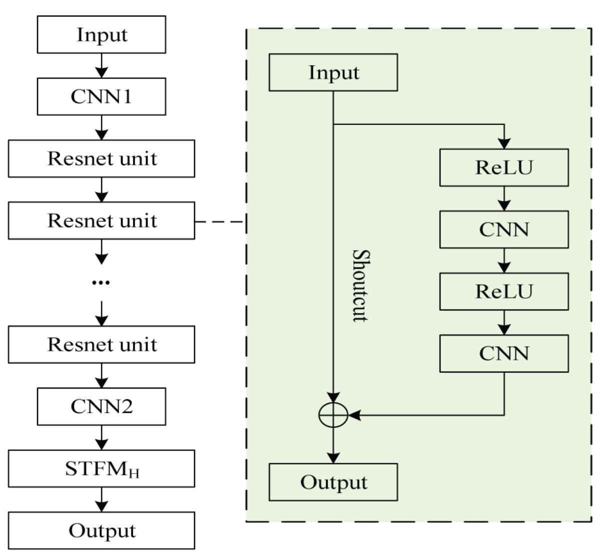
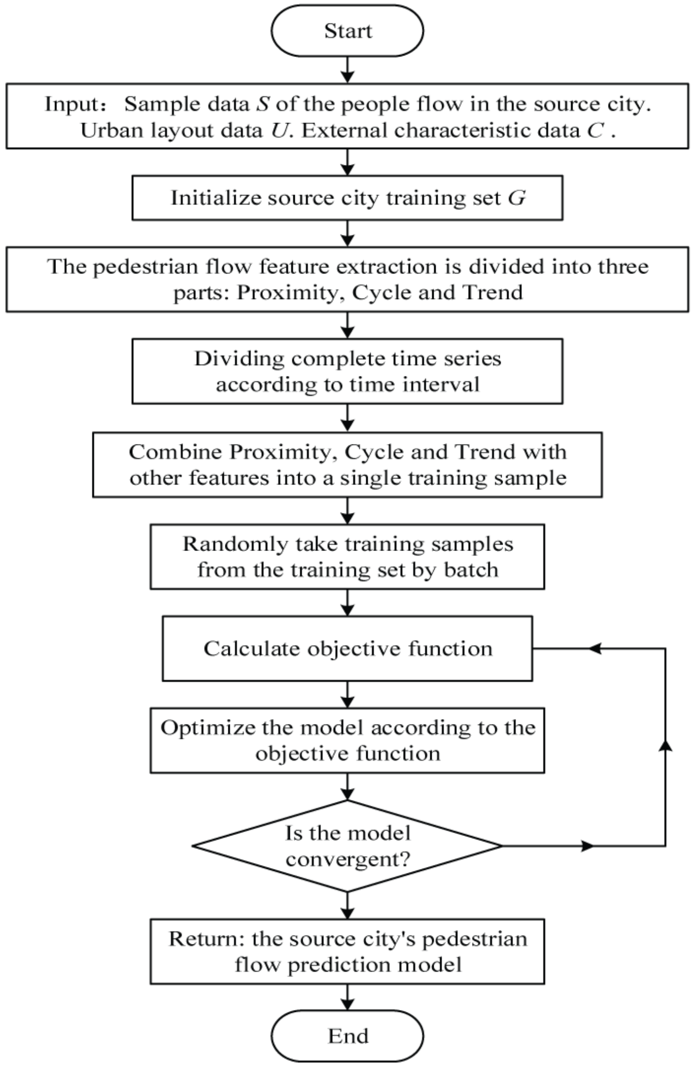


Figure 7. Algorithm of the urban TFP model



The experiment also uses external information data, such as weather and holidays. The weather data includes precipitation, snow, temperature, and wind speed, and considers the effects of working days, weekends, and holidays on the crowd flow model. Table 3 describes the details of the above two datasets and external information.

Evaluation Indicators

Generally, the performance evaluation of urban TFP methods includes three criteria, namely, Root Mean Square Error (RMSE), Mean Absolute Error (MAE), and Mean Absolute Percentage Error (MAPE). Their calculation methods are shown in formula (2), (3) and (4), respectively:

$$RMSE = \sqrt{\frac{1}{m}\sum_{k=1}^{m}\left(y_t - \hat{y}_t\right)} \qquad (2)$$

Table 1. Experimental environment configuration

Name	Configuration
Operating system	Windows 10
Computer's CPU	@ 2.90GHz 2.90 GHz
Computer's processor	Intel(R) Core(TM) i7-10700
Computer's RAM	16GB
Server's processor	Intel(R) Xeon(R)
Server's CPU	E5-2620 v4 @ 2.10GHz
Server's RAM	64GB
GPU	NVIDIA TESLA P40
Programing language	Python 3.7.6
Environment	Tensorflow
Framework of model	Keras 2.1.6

Table 2. Super parameter settings of the prediction model

Parameter	Field	Value
Loss	Objective function	keras.losses.huber_loss
Epochs	Maximum iteration	5000
Batch_size	Number of batch samples	64
Pic_shape	Picture shape	(256, 256, 3)
Learning_rate	Learning rate	0.001
Lr_r_patience	Learning rate attenuation	10
Lr_r_factor	Learning rate decay rate	0.5
ES_patience	Early stop round	20

Table 3. Datasets and external information

Data set	BikeNYC	TaxiNYC
Longitude range	-74.02~-73.95	-74.02~-73.95
Dimension range	40.67~40.77	40.67~40.77
Time range	1/1/2015~31/12/2015	1/1/2015~31/12/2015
Time interval	1 hour	1 hour
Grid size	(16,16)	(16,16)
Number of tracks	9 million	160 million
Number of time intervals	8754	8754
Weather information	precipitation, snowfall, temperature	
Date	Working days, weekends, holidays	

$$MAE = \frac{1}{m}\sum_{k=1}^{m}\left|y_t - \hat{y}_t\right| \tag{3}$$

$$MAPE = \frac{1}{m}\sum_{k=1}^{m}\left|\frac{y_t - \hat{y}_t}{y_t}\right| \times 100\% \tag{4}$$

Model Training

First, the convergence of the proposed urban TFP on BikeNYC and TaxiNYC are simulated, and we show the results in Figure 8.

It can be seen from Figure 8 and Figure 9 that in the two datasets, BikeNYC dataset converges after the 60th epoch, while TaxiNYC dataset converges in about 75 epochs. However, the loss curve

Figure 8. Change curve of loss under BikeNYC dataset

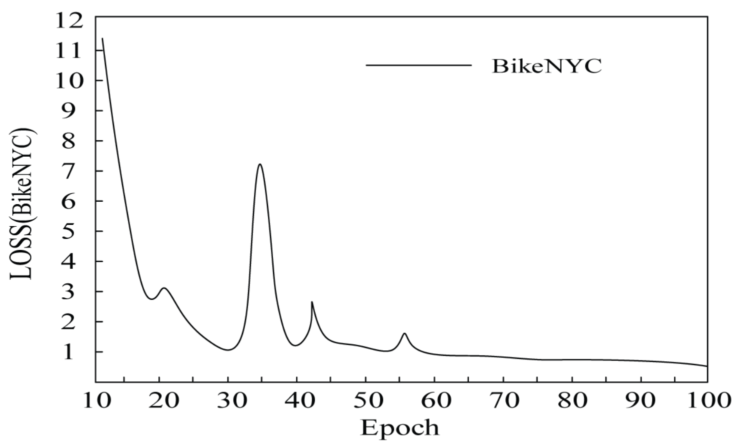
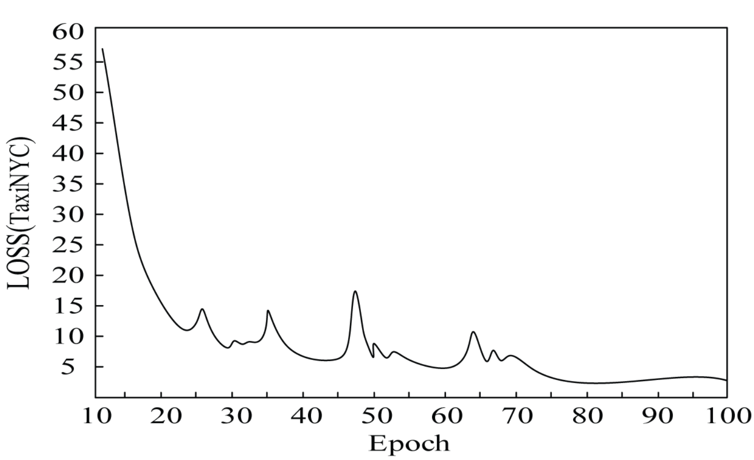


Figure 9. Change curve of loss with the TaxiNYC dataset



obtained from the two datasets does not decline smoothly because it has many losses during training, especially the confrontation loss is difficult to train. This paper used 100 epochs to train the proposed model in the following experiments.

To get the optimal batch size, work was done with two datasets with a batch sizes of 50, 100, 200, 300, 400, 500, 600, 700, 800, 900, and 1000, respectively. The results are shown in Figures 10 and 11, respectively.

In Figures 10 and 11, when two different datasets are used, the three error values of the model are low when the batch size is about 50 and 500. RMSE is about 17.2, MAE is about 10.3, and MAPE is about 5.1. However, since the duration of each iteration is shorter when the batch size is 50 than when the batch size is 500, we set the batch size to 50. After batch size increases, likely, different batches do not change in the gradient direction, which readily results in a local minimum. When the same

Figure 10. Indicator values of models with different batch sizes using the BikeNYC dataset

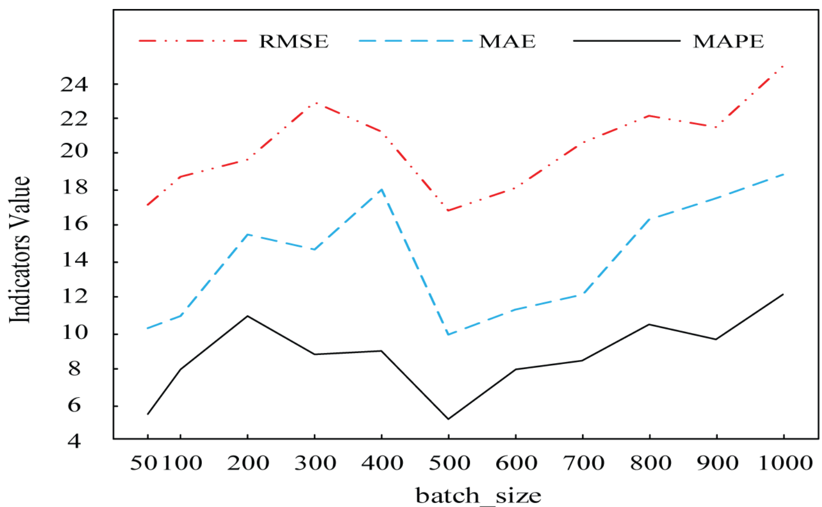
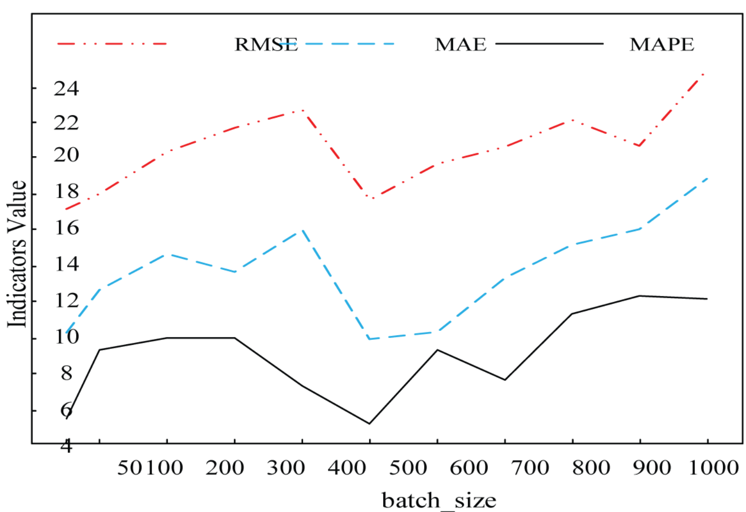


Figure 11. Indicator values of models with different batch sizes under TaxiNYC dataset



accuracy is achieved, if the batch size increases, the corresponding iteration times must be increased, and the required training time will also be extended. We conducted this experiment by changing the batch size based on the same number of iterations. The accuracy, therefore, will decrease when the batch size increases.

Verification of Model Effectiveness

To verify the proposed TFP method, the following experiments are conducted with the two datasets compared with the methods in Zhang et al. (2017), Du et al. (2020) and Bai et al. (2020). With RMSE, MAE, and MAPE as evaluation criteria, the prediction performance using periods of 0.25 hour, 0.5 hour and 1 hour were analyzed. We show the final calculation results of evaluation indicators of different algorithms in Tables 4 and 5, respectively.

In Tables 4 and 5, when BikeNYC and TaxiNYC datasets are used, the proposed big data-driven urban TFP method based on D-L outperforms the other three comparison methods in three evaluation indicators. When using the BikeNYC dataset, the lowest RMSE, MAE, and MAPE of the proposed method reach 18.32, 10.66, and 5.34, respectively. In addition, when predicting 0.25 hours in advance, the error drops considerably compared to the other three comparison methods.

Table 4. Results of different methods using BikeNYC dataset

Method		Proposed Method	ST-ResNet	DST-ICRL	AGCRN
RMSE	0.25h	18.32	35.64	29.88	37.61
	0.5h	18.98	36.51	31.64	38.24
	1.0h	19.56	37.22	32.18	38.96
MAE	0.25h	10.66	25.31	18.67	22.64
	0.5h	11.34	26.87	19.66	23.74
	1.0h	11.87	27.43	20.31	24.55
MAPE	0.25h	5.34	10.67	8.52	9.33
	0.5h	5.96	10.76	8.76	9.64
	1.0h	6.25	10.95	8.97	9.85

Table 5. Results of different methods using TaxiNYC dataset

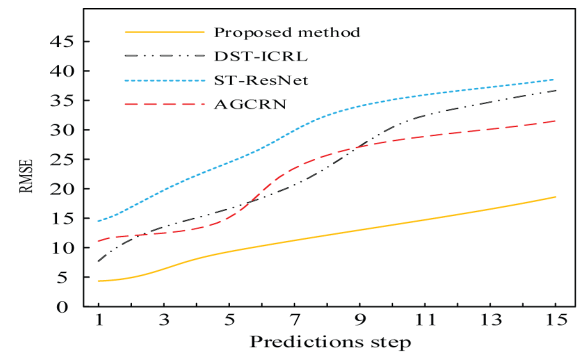
Method		Proposed Method	ST-ResNet	DST-ICRL	AGCRN
RMSE	0.25h	17.72	34.64	28.63	36.07
	0.5h	18.35	35.49	30.31	36.67
	1.0h	18.91	36.18	30.83	37.36
MAE	0.25h	10.31	24.60	17.89	21.71
	0.5h	10.97	26.12	18.83	22.77
	1.0h	11.48	26.66	19.46	23.54
MAPE	0.25h	5.16	10.37	8.16	8.95
	0.5h	5.76	10.46	8.39	9.24
	1.0h	6.04	10.64	8.59	9.45

Figure 12 shows the relationship between the measurement results of each evaluation indicator of different algorithms and the prediction step when using the BikeNYC dataset.

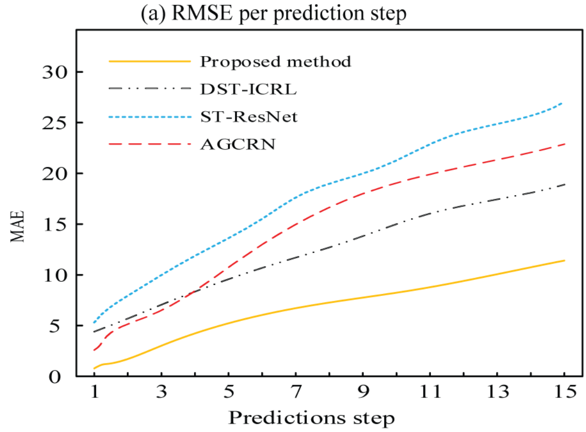
Figure 13 shows the relationship between the measurement results of each evaluation indicator of different algorithms and the prediction step when using the TaxiNYC dataset.

As can be seen from Figure 12 and Figure 13, when two different datasets are used, respectively, with the continuous growth of the prediction step, the proposed urban TFP method performs better than the other three comparison methods. The growth rate and value of the three evaluation indicators

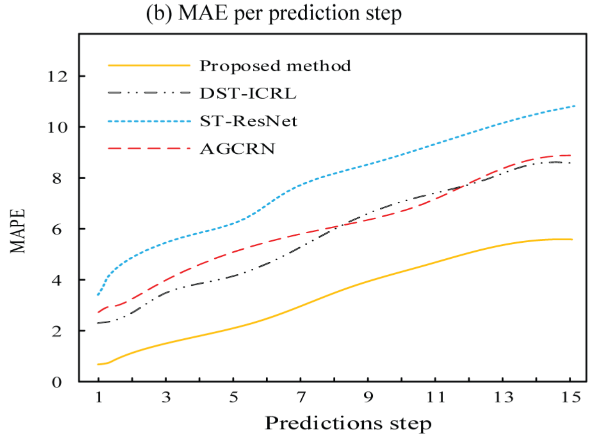
Figure 12. Measurement results of each evaluation indicator of different algorithms under BikeNYC dataset



(a) RMSE per prediction step

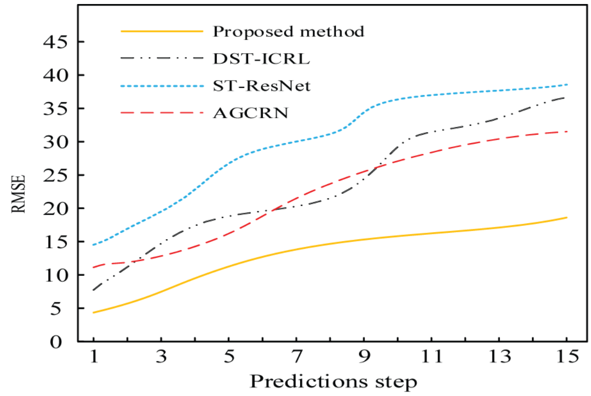


(b) MAE per prediction step

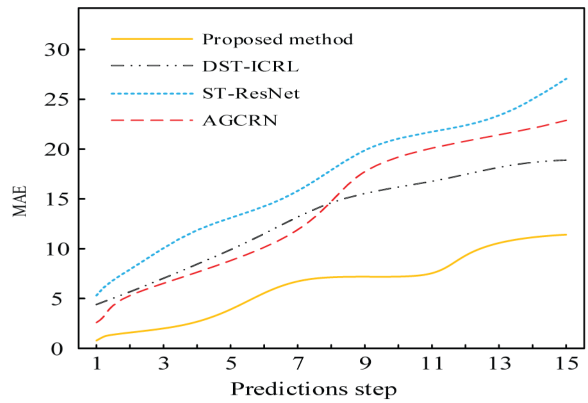


(c) MAPE per prediction step

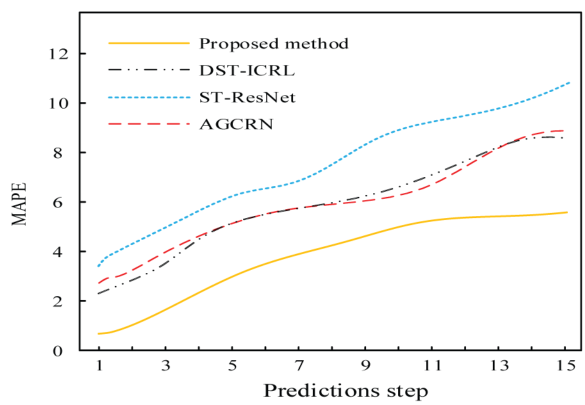
Figure 13. Measurement results of each evaluation indicator of different algorithms under TaxiNYC dataset



(a) RMSE per prediction step



(b) MAE per prediction step



(c) MAPE per prediction step

are the smallest, and the trend of change is the most stable. This stability is because the traditional statistical methods only use the time series data of each node without considering the spatial dependence. Furthermore, the model does not consider the attention mechanism since it can increase the model complexity and the training time and cannot achieve the best effect.

With different datasets, the results of different methods for four crowded periods, including weekday, vacation, peak, and off-peak periods, are shown in Table 6 and Figure 14.

As seen in Table 6 and Figure 14, the proposed urban TFP method shows better predictability than the other three comparison methods in three different evaluation indicators for the four crowded periods of weekday, vacation, peak and peak and off-peak, with the lowest of 18.32, 11.02 and 5.67 respectively.

Table 7 and Figure 15 compare the prediction errors for crowded periods when different methods use different datasets. Table 7 and Figure 15 show that when the BikeNYC and TaxiNYC datasets are used, the proposed urban TFP method performs better than the other three comparison methods, with the smallest error reaching 8.64 and 9.02, respectively. The proposed Attention-CNN-GRU-ResNet model includes convolution to extract S-T features rather than just considering temporal features. The attention-based ConvGRU module captures the time correlation of TF by comprehensively considering spatial attention, GRU, and channel attention, greatly improving the accuracy of urban TFP and reducing the prediction error.

CONCLUSION

With the wide application of GPS, Beidou, and other global positioning systems, an enormous amount of urban S-T traffic data have been generated. These data are crucial for practical applications like traffic planning and crime prediction. However, the existing methods based on D-L require massive data to support training, and we cannot apply these models in some cities or fields since they do not account for sparse data. At the same time, the existing multi-task learning method of traffic prediction only splices the features of related tasks, which cannot well model the correlation between related urban traffic prediction tasks. To overcome the large errors and inaccurate prediction for long-term flow that comes with the traditional urban TFP method, we propose a big data-driven TFP method for sustainable urban development based on deep learning (D-L).

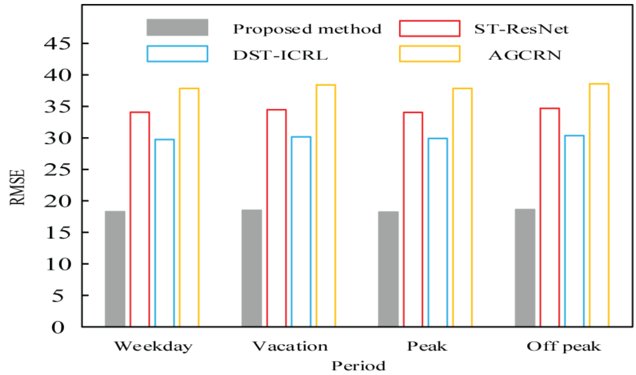
The experiment results show that:

1. Using CNN to classify the input information according to its hierarchical structure can effectively reduce the model calculation.

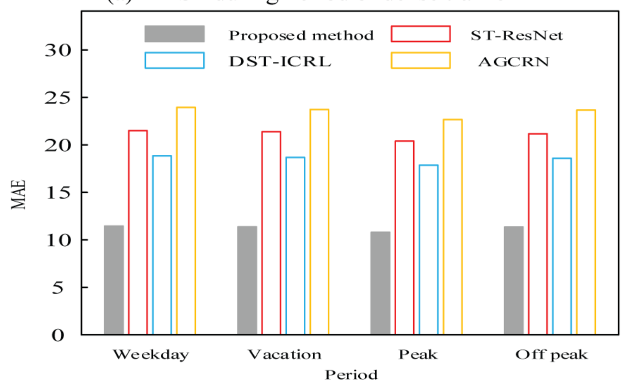
Table 6. Prediction results of different methods for periods of dense traffic

Period		Proposed Method	ST-ResNet	DST-ICRL	AGCRN
Weekday	RMSE	18.32	34.08	29.86	37.92
	MAE	11.64	21.65	18.97	24.09
	MAPE	5.83	10.84	9.50	12.07
Vacation	RMSE	18.56	34.52	30.25	38.42
	MAE	11.54	21.46	18.81	23.89
	MAPE	5.94	11.05	9.68	12.30
Peak	RMSE	18.33	34.09	29.88	37.94
	MAE	11.02	20.50	17.96	22.81
	MAPE	5.67	10.55	9.24	11.74
Off peak	RMSE	18.68	34.74	30.45	38.67
	MAE	11.49	21.37	18.73	23.78
	MAPE	5.82	10.83	9.49	12.05

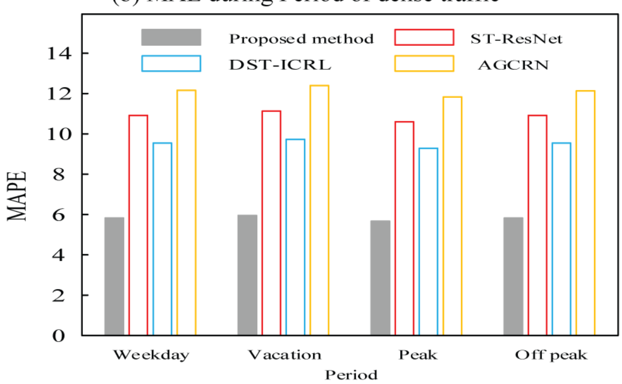
Figure 14. Prediction results of different methods for periods of dense traffic



(a) RMSE during Period of dense traffic



(b) MAE during Period of dense traffic

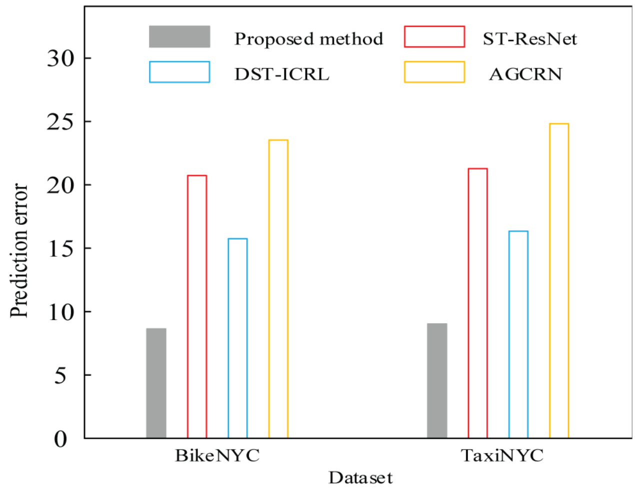


(c) MAPE during Period of dense traffic

Table 7. Prediction errors of different methods when using different datasets

Data Sets	Method			
	Proposed Method	ST-ResNet	DST-ICRL	AGCRN
BikeNYC	8.64	20.64	15.79	23.47
TaxiNYC	9.02	21.22	16.35	24.72

Figure 15. Prediction error of different methods when using different datasets



2. The introduction of attention-based GRU model into the detection model can capture the time correlation of TF more carefully.
3. By introducing residuals, the ResNet module can alleviate gradient disappearance/explosion, avoid deep network degradation, and dramatically improve the model's prediction performance.

Future work will focus on using domain knowledge or other means to ease data sparsity and consider the spatial and temporal complexity of S-T data based on graph structure to further improve the accuracy of the model prediction.

AUTHOR NOTE

The data used to support the findings of this study are included within the article.

The author declares that there is no conflict of interest regarding the publication of this paper.

REFERENCES

Ahmed, M. S., & Cook, A. R. (1979). Analysis of freeway traffic time-series data by using Box-Jenkins techniques. *Transportation Research Record: Journal of the Transportation Research Board*, *722*(3), 1–9.

Bai, L., Yao, L., Li, C., Wang, X., & Wang, C. (2020). Adaptive graph convolutional recurrent network for traffic forecasting. Advances in Neural Information Processing Systems, 33(5), 322-330.

Cheng, L., Zang, H., Ding, T., Sun, R., Wang, M., Wei, Z., & Sun, G. (2019). Ensemble recurrent neural network based probabilistic wind speed forecasting approach. *Energies*, *11*(8), 195–204. doi:10.3390/en11081958

Cheng, Z., Lu, J., Zhou, H., Zhang, Y., & Zhang, L. (2021). Short-term traffic flow prediction: An integrated method of econometrics and hybrid deep learning. *IEEE Transactions on Intelligent Transportation Systems*, *33*(6), 5231–5244. doi:10.1109/TITS.2021.3052796

Cui, Z., Ke, R., Pu, Z., & Wang, Y. (2020). Stacked bidirectional and unidirectional LSTM recurrent neural network for forecasting network-wide traffic state with missing values. *Transportation Research Part C, Emerging Technologies*, *118*, 102674. doi:10.1016/j.trc.2020.102674

Du, B., Peng, H., Wang, S., Bhuiyan, M. Z. A., Wang, L., Gong, Q., Liu, L., & Li, J. (2020). Deep irregular convolutional residual LSTM for urban traffic passenger flows prediction. *IEEE Transactions on Intelligent Transportation Systems*, *21*(3), 972–985. doi:10.1109/TITS.2019.2900481

Du, S., Li, T., Gong, X., & Horng, S.-J. (2020). A hybrid method for traffic flow forecasting using multimodal deep learning. *International Journal of Computational Intelligence Systems*, *13*(1), 85–97. doi:10.2991/ijcis.d.200120.001

Han, D., Chen, J., & Sun, J. (2019). A parallel spatiotemporal deep learning network for highway traffic flow forecasting. *International Journal of Distributed Sensor Networks*, *15*(2), 58–66. doi:10.1177/1550147719832792

Han, L., & Huang, Y.-S. (2020). Short-term traffic flow prediction of road network based on deep learning. *IET Intelligent Transport Systems*, *14*(6), 495–503. doi:10.1049/iet-its.2019.0133

He, Z., Chow, C.-Y., & Zhang, J.-D. (2020). STNN: A spatio-temporal neural network for traffic predictions. *IEEE Transactions on Intelligent Transportation Systems*, *22*(7), 1–10. doi:10.1109/TITS.2020.3006227

Huang, C., Wu, X., Zhang, C., Yin, D., Zhao, J., & Chawia, N. V. (2019). Mist: A multiview and multimodal spatial-temporal learning framework for citywide abnormal event forecasting. *Proceedings of The World Wide Web Conference, 13*(7), 717–728. doi:10.1145/3308558.3313730

Kamarianakis, Y., & Prastacos, P. (2018). Forecasting traffic flow conditions in an urban network: Comparison of multivariate and univariate approaches. *Transportation Research Record: Journal of the Transportation Research Board*, *1857*(1), 74–84. doi:10.3141/1857-09

Li, G., Cao, N., Zhu, P., Zhang, Y., Zhang, Y., Li, L., Li, Q., & Zhang, Y. (2021). Towards smart transportation system: A case study on the rebalancing problem of bike sharing system based on reinforcement learning. *Journal of Organizational and End User Computing*, *33*(3), 35–49. doi:10.4018/JOEUC.20210501.oa3

Li, J., Guo, F., Sivakumar, A., Dong, Y., & Krishnan, R. (2021). Transferability improvement in short-term traffic prediction using stacked LSTM network. *Transportation Research Part C, Emerging Technologies*, *124*(3), 102977. doi:10.1016/j.trc.2021.102977

Li, T., Ni, A., Zhang, C., Xiao, G., & Gao, L. (2021). Short-term traffic congestion prediction with Conv–BiLSTM considering spatio-temporal features. *IET Intelligent Transport Systems*, *14*(14), 1978–1986. doi:10.1049/iet-its.2020.0406

Liu, J., Huang, J., Sun, R., Yu, H., & Xiao, R. (2020). Data fusion for multi-source sensors using GA-PSO-BP neural network. *IEEE Transactions on Intelligent Transportation Systems*, *22*(10), 6583–6598. doi:10.1109/TITS.2020.3010296

Nguyen, T., Nguyen, G., & Nguyen, B. M. (2020). EO-CNN: An enhanced CNN model trained by equilibrium optimization for traffic transportation prediction. *Procedia Computer Science*, *176*(1), 800–809. doi:10.1016/j.procs.2020.09.075

Shanshan, P., Chao, M., Haitao, Z., & Kun, L. (2022). An evaluation system based on user big data management and artificial intelligence for automatic vehicles. *Journal of Organizational and End User Computing*, *34*(10), 1–21. doi:10.4018/JOEUC.309135

Shi, X., Chen, Z., Wang, H., Yeung, D.-Y., Wong, W.-K., & Woo, W.-C. (2021). Convolutional LSTM network: A machine learning approach for precipitation nowcasting. Advances in Neural Information Processing Systems, 28, 158-166.

Voort, M., Dougherty, M., & Watson, S. (1996). Combining Kohonen maps with ARIMA time series models to forecast traffic flow. *Transportation Research Part C, Emerging Technologies*, *4*(5), 307–318. doi:10.1016/S0968-090X(97)82903-8

Wang, S., Cao, J., & Yu, P. S. (2020). Deep learning for spatio-temporal data mining: A survey. *IEEE Transactions on Knowledge and Data Engineering*, *34*(8), 3681–3700. doi:10.1109/TKDE.2020.3025580

Wang, S., Chen, H., Cao, J., Zhang, J., & Yu, P. S. (2020). Locally balanced inductive matrix completion for demand-supply inference in stationless bike-sharing systems. *IEEE Transactions on Knowledge and Data Engineering*, *32*(12), 2374–2388. doi:10.1109/TKDE.2019.2922636

Williams, B. M., & Hoel, L. A. (2003). Modeling and forecasting vehicular traffic flow as a seasonal ARIMA Process: Theoretical basis and empirical results. *Journal of Transportation Engineering*, *129*(6), 664–672. doi:10.1061/(ASCE)0733-947X(2003)129:6(664)

Yao, H., Tang, X., Wei, H., Zheng, G., & Li, Z. (2019). Revisiting spatial-temporal similarity: A deep learning framework for traffic prediction. *Proceedings of the AAAI Conference on Artificial Intelligence*, *4*(2), 132–140. doi:10.1609/aaai.v33i01.33015668

Zhang, H., Chen, L., Cao, J., Zhang, X., & Kan, S. (2021). A combined traffic flow forecasting model based on graph convolutional network and attention mechanism. *International Journal of Modern Physics C*, *32*(12), 238–246. doi:10.1142/S0129183121501588

Zhang, J., Zheng, Y., & Qi, D. (2017). Deep spatio-temporal residual networks for citywide crowd flows prediction. *Proceedings of Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence, 31*(1). doi:10.1609/aaai.v31i1.10735

Zhang, J., Zheng, Y., Sun, J., & Qi, D. (2020). Flow prediction in spatio-temporal networks based on multitask deep-learning. *IEEE Transactions on Knowledge and Data Engineering*, *32*(3), 468–478. doi:10.1109/TKDE.2019.2891537

Zhang, K., Wu, L., Zhu, Z., & Deng, J. (2020). A multitask learning model for traffic flow and speed forecasting. *IEEE Access: Practical Innovations, Open Solutions*, *8*(3), 80707–80715. doi:10.1109/ACCESS.2020.2990958

Zhang, W., Yu, Y., Qi, Y., Shu, F., & Wang, Y. (2019). Short-term traffic flow prediction based on spatio-temporal analysis and CNN deep learning. *Transportmetrica A: Transport Science*, *15*(2), 1688–1711. doi:10.1080/23249935.2019.1637966

Zhang, Y., & Huang, G. (2019). Traffic flow prediction model based on deep belief network and genetic algorithm. *IET Intelligent Transport Systems*, *12*(6), 533–541. doi:10.1049/iet-its.2017.0199

Zhang, Y., Wang, S., Chen, B., Cao, J., & Huang, Z. (2021). Trafficgan: Network-scale deep traffic prediction with generative adversarial nets. *IEEE Transactions on Intelligent Transportation Systems*, *22*(1), 219–230. doi:10.1109/TITS.2019.2955794

Zhao, W., Gao, Y., Ji, T., Wan, X., Ye, F., & Bai, G. (2020). Deep temporal convolutional networks for short-term traffic flow forecasting. *IEEE Access : Practical Innovations, Open Solutions*, *7*(13), 114496–114507. doi:10.1109/ACCESS.2019.2935504

Zhou, T., Han, G., Xu, X., Han, C., Huang, Y., & Qin, J. (2019). A learning-based multimodel integrated framework for dynamic traffic flow forecasting. *Neural Processing Letters*, *49*(1), 407–430. doi:10.1007/s11063-018-9804-x