Article

Traffic Flow Prediction with Attention Mechanism Based on TS-NAS

Cai Zhao [1,2], Ruijing Liu [1], Bei Su [1], Lei Zhao [1], Zhiyong Han [1] and Wen Zheng [1,3,*]

1. Institute of Public-Safety and Big Data, College of Data Science, Taiyuan University of Technology, Taiyuan 030024, China
2. Center of Information Management and Development, Taiyuan University of Technology, Taiyuan 030024, China
3. Center for Big Data Research in Health, Changzhi Medical College, Changzhi 046000, China
* Correspondence: zhengwen@tyut.edu.cn

Abstract: The prediction of traffic flow is of great significance in the traffic field. However, because of the high uncertainty and complexity of traffic data, it is challenging that doing traffic flow prediction. Most of the existing methods have achieved good results in traffic flow prediction, but are not accurate enough to capture the dynamic temporal and spatial relationship of data by using the structural information of traffic flow. In this study, we propose a traffic flow prediction method with temporal attention mechanism and spatial attention mechanism based on neural architecture search (TS-NAS). Firstly, based on temporal and spatial attention mechanisms, we design a new attention mechanism. Secondly, we define a novel model to learn temporal flow and space flow in traffic network. Finally, the proposed method uses different modules about time, space and convolution and neural architecture search to be used for optimizing the model. We use two datasets to test the method. Experimental results show that the performance of the method is better than that of the existing method.

Keywords: prediction of traffic flow; temporal attention mechanism; TS-NAS; spatial attention mechanism

1. Introduction

In recent years, traffic forecasting has become a hot topic. It is necessary to improve the intelligence of road traffic management and public travel services, so that the hiding function of existing traffic flow data can play a role in data mining [1,2]. However, in terms of prediction, the utilization of spatial structure information is insufficient [3].

For traffic flow, it is one of the main parameters reflecting traffic. Traffic flow data has accumulated mass of traffic time series with geographical location information. How to use these data effectively is a problem actively explored in the field of transportation. Under the above background, if people can effectively use the massive traffic data collected and give full play to its value, the prediction of traffic flow will be more accurate and the operational capacity and efficiency of the whole road network will be improved. Many methods have been applied to traffic flow prediction. HA algorithm uses the mean value of the last time slice to predict the future value. ARIMA [4] is a regression method, which is applied to the prediction of traffic flow. The VAR [5] model captures the relationship between traffic flow sequence nodes. In addition, there are other sequence models, such as GRU [6–9]. In particular, both temporal and spatial attention mechanisms are considered in [10]. However, these models are insufficient in considering the spatial and temporal properties and model structure.

Graph neural network can pay attention to the interaction of node information in the process of network optimization, to learn better feature representation, and is applied to man fields [9,11]. LSTM network [6] has a good effect on the learning of sequence information. Attention mechanism [12,13] has shown good performance in many fields, such

as image classification [14], image caption [15] and so on. Neural architecture search [16] can optimize the network structure, so we can optimize our traffic flow network structure through neural architecture search in our custom space to achieve the best effect.

We propose a new structure for traffic flow prediction, which is called TS-NAS. Firstly, we represent the features of spatial nodes, and then put the features and graphs into our search convolution space. Secondly, to better learn the spatial characteristics of traffic flow data, we define the search space for spatial attention. After learning the spatial information of traffic flow data through the network, we pool the spatial features. Similarly, to learn the temporal characteristics of traffic flow data, we define the search space for temporal attention. To better study the space and time of traffic flow network, the whole network structure defines the neural architecture search space. The main contributions are summarized as follows:

1   A novel spatio-temporal attention mechanism is designed based on the spatio-temporal attention mechanism;
2   The residual network is added between the spatio-temporal blocks to avoid the shortcomings of excessive smoothing and feature weakening caused by the increase of the number of spatio-temporal blocks ;
3   Combining GCN and LSTM, a new module is proposed to learn the information of spatio-temporal flows in transportation network;
4   In this paper, a neural network search mechanism is added to select the appropriate spatial convolution module, and the selection standard is to make the prediction model optimal.

The tips of this article are summarized as follows. The Related work section introduces the related work of graph convolution and traffic flow. The Proposed method section introduces the prediction framework of the traffic flow model and detailed sub-task descriptions, including the separate graph convolution module, attention space mechanism, and neural network search mechanism. The Experimental Analysis section introduces the main experimental process and results of this research. The Conclusion section concludes the paper.

2. Related Work

2.1. Convolutional Modules

The methods of graph convolution neural network include spatial methods and spectral methods. Based on spatial methods, a linear method [17] has been proposed and achieved good results. A method [18] has been proposed in action recognition and achieved good results. Based on septal methods, the general graph convolution neural network method is proposed [11], and the gating method is proposed [9], and this method is applied to traffic prediction. This paper focuses on the prediction of traffic flow between urban roads. Since the traffic flow data does not meet the conditions of Convolutional Neural Networor (CNN), so the Graph Convolutional Network (GCN) is used in this paper. The difficulty of GCN architecture is the definition of convolution kernel as shown in the Figure 1.

The convolution operation is extended to the graph structure data by using the spectrogram method, and the data are regarded as the signal on the graph, and then the graph signal is processed directly on the graph. In this model, the role of the GCN layer is to extract spatial characteristics of traffic flow through convolutional networks.

The traditional long short-term memory (LSTM) [6] relies on a feed-forward neural network to calculate the internal correlation. However, for spatial and temporal data, this will bring redundancy, because spatial data has strong local characteristics. In addition, the LSTM cannot obtain the previous information from the back during the process.
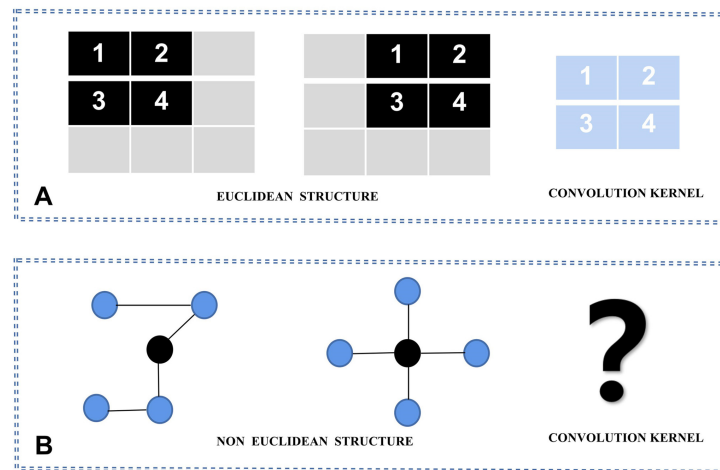
Figure 1. Network Structure and Graph Structure ((**A**) CNN architecture; (**B**) GCN architecture).

2.2. Structural Optimization and Traffic Flow Prediction

Attention mechanism enables the model to learn important features. Therefore, attention mechanism is used in various scenes and develops rapidly. Attention was first introduced [12,13] to help the model learn important information. In [19], the attention mechanism is applied to the image caption task. Recently, various interesting attention mechanism models have been proposed [14,15]. The attention mechanism has also been applied to the target detection task. Neural architecture search is one of the research hot topics in recent years. It is mainly used to solve the parameter optimization problem of neural network learning model. Neural architecture search (NAS) was first proposed in [16]. NAS optimizes the network by optimizing the neural structure. Reinforcement learning was first applied to NAS search method, which produced different optimization learning algorithms for different search strategies. In addition, there are evolutionary algorithm and Bayesian optimization algorithm to optimize the neural architecture search. In recent years, the research on traffic prediction has achieved important results. For example, traditional methods are often insufficient in capturing spatial information [20–23], so they can not obtain satisfactory results in performance. With the extensive application of deep learning in various fields, such as speech recognition, image processing [24], and action recognition [25], data exploration of spatial structure has also achieved remarkable results. With the emergence of deep learning, more and more researchers have designed relevant models to explore structured data, such as [26–29].

3. Proposed Method

In this section, this study proposes a temporal attention mechanism and a spatial attention mechanism traffic flow prediction method based on neural architecture search (TS-NAS). Firstly, spatial features are extracted by GCN, and temporal features are extracted by BiLSTM, so that the network has a more special "two-way memory" function; then, a spatial attention mechanism is established to increase the latitude of time slice feature aggregation and define and propose a temporal attention mechanism; A residual network is added between the spatio-temporal blocks to avoid the shortcomings of over-smoothing and feature weakening caused by the increase of the space-time block features as the number of blocks increases. Finally, Nas is used to search for the optimal solution, and the TS-NAS model is established. See the following subsections for details.

3.1. Notations

We spilt the dataset into days. At a certain point in time , X represents the feature of all nodes, $X \in R^{N \times F}$, N is the number of nodes and F is the dimension of nodes. X_t represents the features of all nodes at time t. X_i is used to represent the feature of the i th

node. Because the traffic flow prediction is the problem of sequence prediction, and the sequence information of the future is mainly predicted by the historical information so that the predicted information is represented as Y, $Y \in \mathrm{R}^{N \times P}$, P is the dimension of prediction node. The traffic network can be represented as G (V,E). V represents space of the nodes and E represents the relationship between the nodes. At time t, the x_i represents feature of i th node. By calculating the Euclidean distance, we construct a new graph matrix A, which is described as:

$$A_i = \frac{e^{||x_i - x_j||_2}}{\sum_j e^{||x_i - x_j||_2}} \tag{1}$$

where $A_{i,j}$ is the value from graph matrix A. It can be used to represent the relationship between nodes in graph space so that can be applied to graph convolution network and graph attention convolution network in our convolution spaces.

3.2. Search Space about Attention

The attention can be expressed about x_i as a dot product attention model. Spatial and temporal attention models are shown in Figures 2 and 3. We define that the attention in the search space can be expressed as:

$$P_1(x_i, x_j) = x_i x_j^T \tag{2}$$

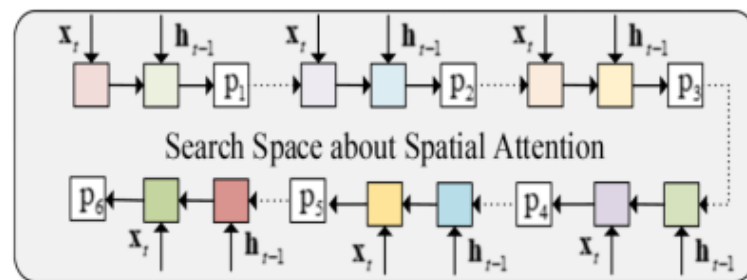p_1 (x_i,x_j) is a dot product attention model.

Figure 2. The search space of network model about spatial attention.

$$P_2(x_i, x_j) = \frac{x_i x_j^T}{\sqrt{d}} \tag{3}$$

p_2 (x_i,x_j) is a scaled dot product attention model. The d takes a fixed value. In our search model, we use it as a learning parameter.

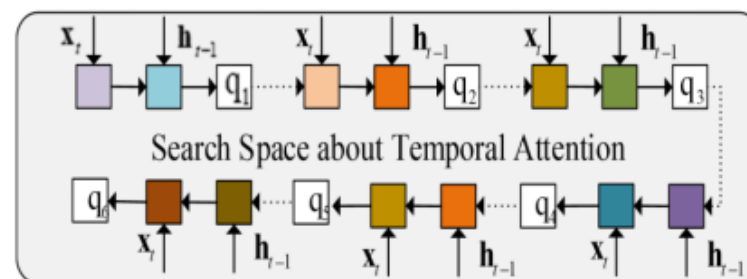$$P_3(x_i, x_j) = x_i W x_j^T \tag{4}$$

Figure 3. The search space of network model about temporal attention.

p_3 (x_i,x_j) is a bilinear attention model, W is a learnable weight matrix, and $W \in \mathrm{R}^{F \times F}$.

$$P_4(x_i, x_j) = sigmoid(x_j W + x_j U v^T) \tag{5}$$

$p_4(x_i, x_j)$ is an additive model. Both W and U are learnable weight matrixs, $W \in \mathrm{R}^{F \times F}$, $U \in \mathrm{R}^{F \times F}$.

$$P_5(x_i, x_j) = x_i W(x_j U)^T \tag{6}$$

$p_5(x_i, x_j)$ is a key value pair attention mechanism model.

$$P_6(x_i, x_j) = \alpha tanh(x_i W(x_i U)^T + (1 - \alpha)tanh(x_i W(x_j U)v^T \tag{7}$$

$p_6(x_i, x_j)$ is designed by ourselves. At the same time, the temporal attention mechanism is expressed as:

$$q_6(f_i, f_j) = \alpha tanh(f_i W(f_i U)^T + (1 - \alpha)tanh(f_i W(f_j U)v^T \tag{8}$$

3.3. Model about TS-NAS

The model architecture as shown in Figure 4 and the data of the model is represented as x_i. By calculating the relationship between them, the representation of graph structure can be obtained as A. Then, we input the features and graphs into the convolution space, which is similar to the attention search space. We use the neural architecture search mechanism to finally select the appropriate convolution module to make the model optimal. Then, the output result of convolution space is input into the search space about pooling, and the pooled module is composed of max ($\cdot$), min ($\cdot$), mean ($\cdot$) or flatten ($\cdot$). Therefore, we can obtain f_i. For the prediction of traffic flow, we need to consider not only the spatial relationship, but also the temporal relationship. Therefore, we define the search space for temporal attention. Then we use the full connection and softmax operation, and finally obtain our prediction results.
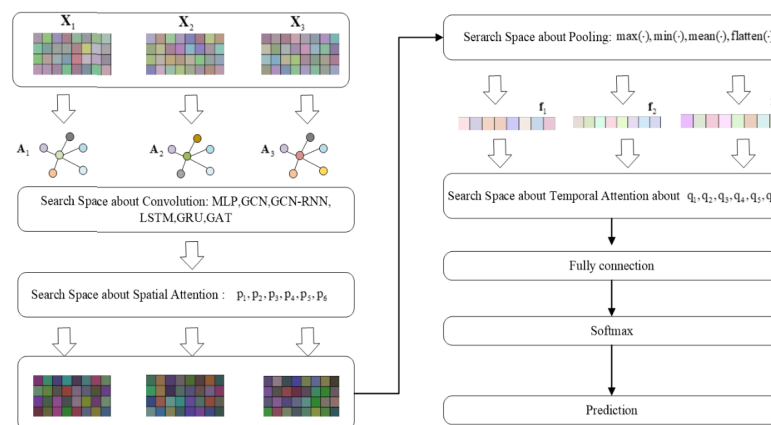

Figure 4. Network model of TS-NAS.

3.4. Prediction

The TS-NAS model was evaluated by MAE and RMSE as follows:

$$MAE = \frac{1}{n} \sum_{i=1}^{n} |\hat{y}_i - y_i| \tag{9}$$

$$RMAE = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (\hat{y}_i - y_i)^2} \tag{10}$$

where $\hat{y}_i$ is the traffic flow data predicted by the proposed method, y_i is the ground truth. n represents the numbers of samples. In addition, we use Adam optimization algorithm to update the gradient.

4. Experimental Analysis

4.1. Experimental Data

In this project, PEMSD4 and PEMSD8 were utilized as the benchmark data set to verify performance of the proposed method in this paper. To be specific, the data were recorded from the sensor information and the collected time series data, and the collection frequency was 30s/time. There are various data types recorded in PEMS. In this paper, the traffic flow is selected as the feature of the data set.

The PEMSD4 data set in this paper is derived from the traffic data of the San Francisco Bay Area Expressway. The traffic network in this area contains 29 highways and a total of 3848 data collection devices. The collection time range is 59 days from January to February 2018. The experiment selected data from January to February 2018 for implementation, with the first 50 days as the training set and the last 9 days as the test set. The experiment was analyzed in detail so that 47 detection points on PEMSD4 were selected as research objects through the spatial geographical distribution of detection points.

The PEMSD8 data set is derived from traffic data from San Bernardino, with 1979 detectors installed on eight roads. Data from July to August 2016 were selected for the experiment. In this paper, the data on the first 50 days are used as the training set and the data on the last 12 days is the test set.

4.2. Settings

The input feature dimension is $47 \times 1 \times 12$, The output prediction dimension is 47×12. The graph convolution operation used 64 convolution cores of the same size. The standard convolution operation of time dimension used 64 convolution kernels of the same size. These convolution kernels were 1 along the spatial axis and 12 along the time axis. In the training process, the Adam optimizer was used to optimize the model, and the learning rate was set as 0. 0001. The number of epoch is 100. In addition, 10% of the training set as the validation set was selected to adjust other parameters. For fair comparison, all experiments were conducted on server with 128G memory, 2. 4 GHz GPU and a GeForce GTX 1080Ti.

4.3. Baseline Methods

Compare our algorithm by the following baseline methods:

- HA: It is a historical average. As a comparison algorithm, the average of the last 12 time slices is used to predict the next value.
- ARIMA: This is an autoregressive synthetic moving average method. Time series analysis used to predict future value.
- VAR: In order to capture the pairwise relationship between all traffic flow series, a more advanced time series model is proposed, which is called vector autoregressive algorithm.
- LSTM: This is a special RNN model, as a long-term and short-term memory network.
- GRU: This is a gated recurrent unit network, a special RNN model.
- STGCN: In order to extract spatial information, a spatiotemporal convolution model based on spatial method is proposed.
- GLU-STGCN: In order to forecast traffic, a graph convolution network with gating mechanism is proposed, and good performance is achieved.
- ASTGCN: A novel attention based spatiotemporal graph convolute.

4.4. Experimental Evaluation

To evaluate the performance of the proposed model, we used PEMD4 and PEMD8 datasets. In addition, we use eight baseline methods to compare with our model through MAE and RMSE metrics. The results are shown in Figures 5–8.

Compared with other benchmark methods, the proposed method in this paper achieves the best performance in two evaluation metrics. The RMSE and MAE of TS-NAS improved by 1.67% and 4.31%, respectively, compared to the optimal prediction on PEMSD4 in Figures 5 and 6 from the benchmark method. We compare the same models with the 8 baseline methods on PEMSD8 in Figures 7 and 8. With the reference method, RMSE and MAE increased by 9.39% and 11.83%, respectively. It can also be observed that the traditional time series prediction methods, such as HA, ARIMA and VAR, which are not ideal because of their limited modeling ability. The method based on TS-NAS achieves better prediction results than traditional methods. Due to the consideration of spatio-temporal dependence of data, the prediction results of the models including spatio-temporal graph convolution mechanisms ST-GCN, GU-STGCN, MSTGCN and ASTGCN are superior to the general deep learning models. The performance of the proposed method is much better than other models.
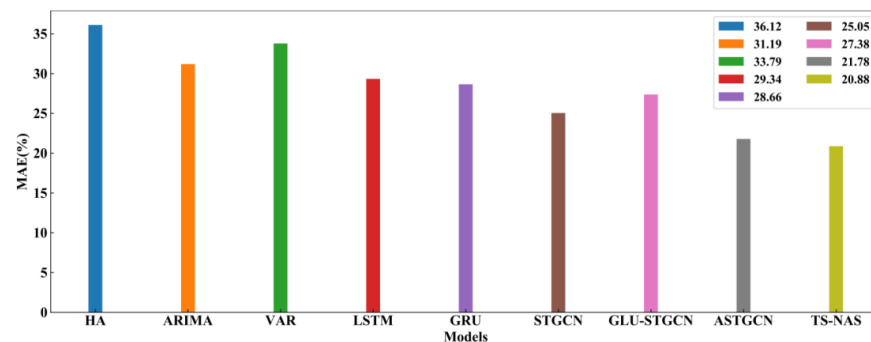
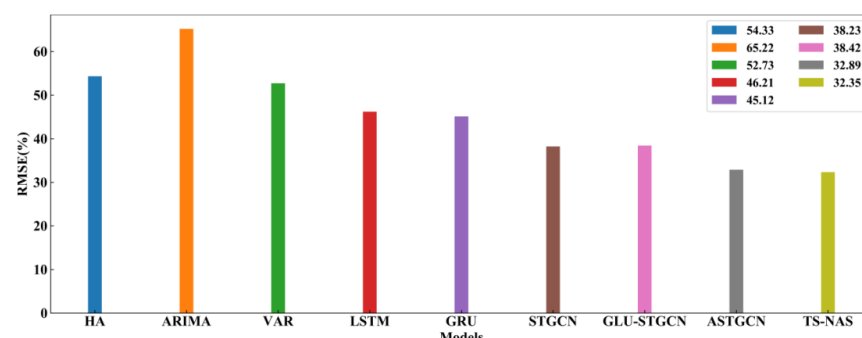**Figure 5.** Comparison MAE of Baseline Method for PEMSD4.

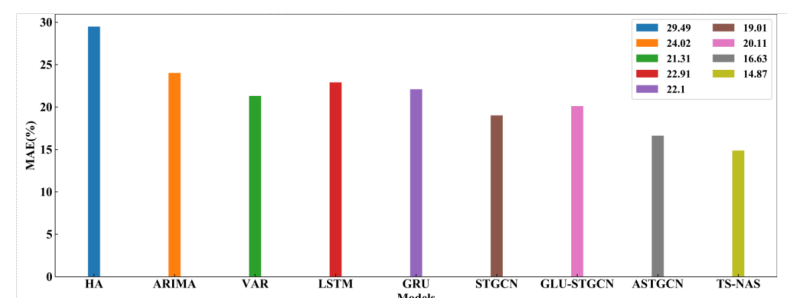**Figure 6.** Comparison RMSE of Baseline Method for PEMSD4.

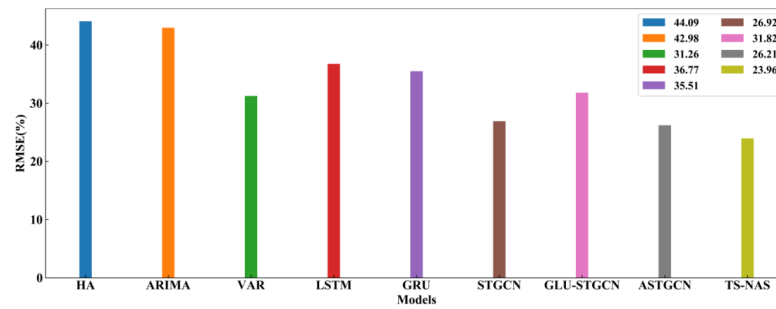**Figure 7.** Comparison of Accuracy of Baseline Method for PEMSD8.

Figure 8. Comparison of Accuracy of Baseline Method for PEMSD8.

4.5. Our Analysis

In this paper, 47 representatives and evenly distributed detection points will be selected within the research scope. The checkpoints were spread over five streets, located in the downtown area of Oakland in western California. The longitude ranges from 122.20 W to 122.30 W, while the latitude ranges from 37.85 N to 37.75 N. The distribution is marked with red, yellow, blue, green, and black checkpoints on Streets 1, 2, 3, 4, and 5, to record the change in demand for its traffic flow over time as shown in Figure 9.
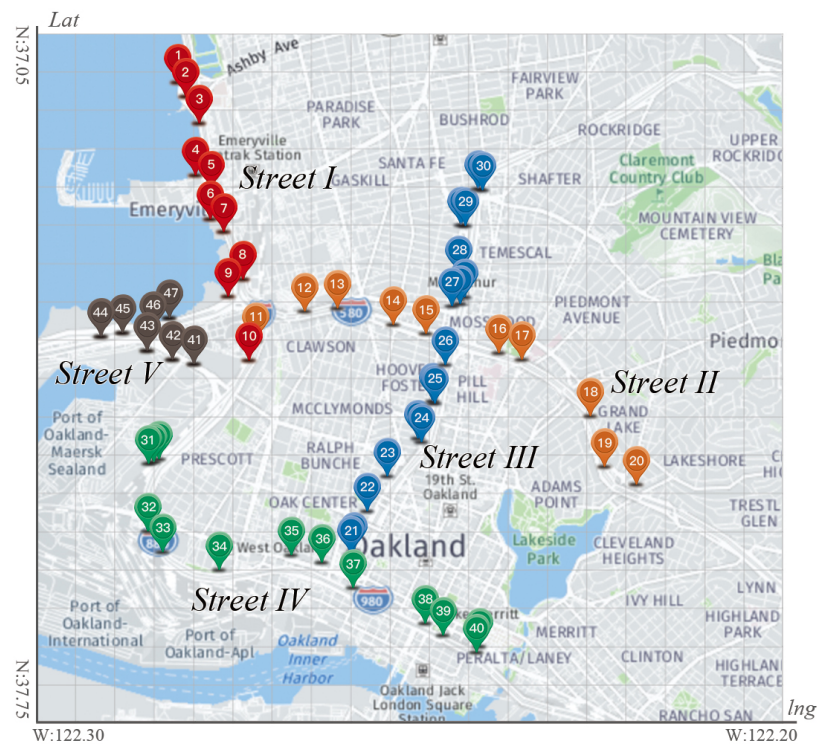
Figure 9. Spatial Geographic Map of the Research Area.

Through the prediction of our model, we draw the flow size diagram of traffic flow, as shown in Figure 10. In addition, we predict the different locations of street 3 and street 4. Street 3 predicts the 21st to 30th detection points as shown in Figure 11, Street 4 predicts the 31st to 40th detection points as shown in Figure 12. Obviously, TS-NAS model shows excellent performance.
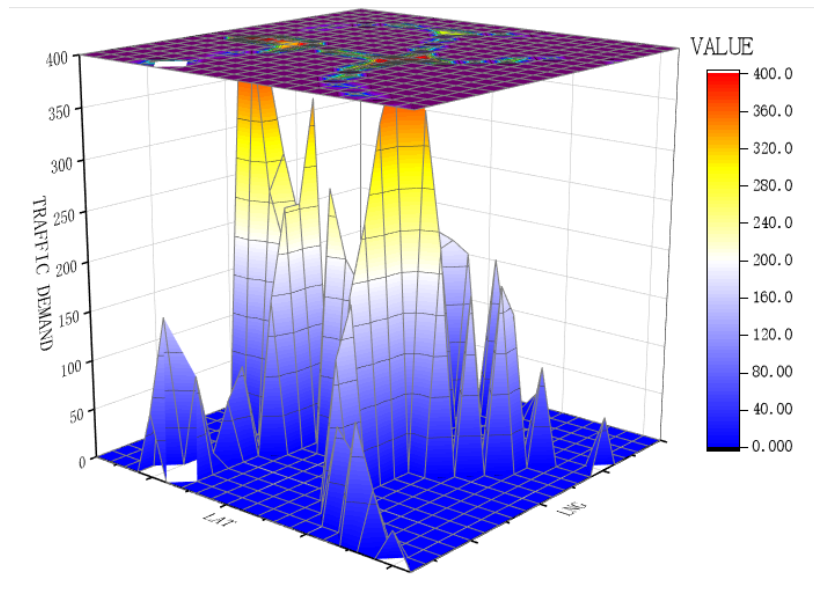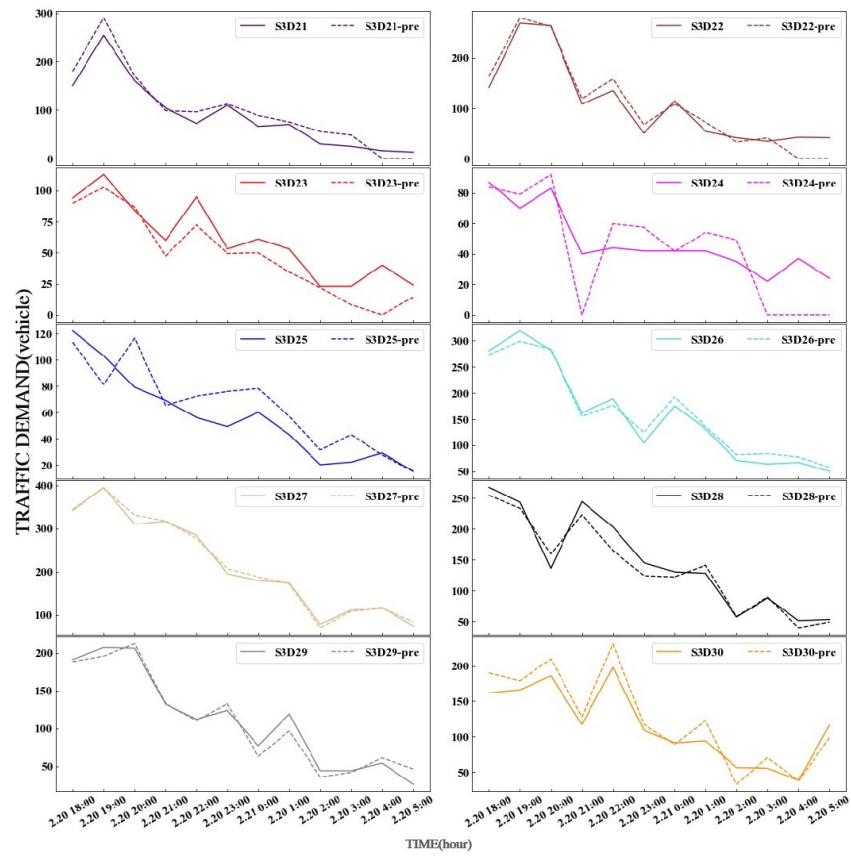
Figure 10. Flow size of traffic flow.

Figure 11. Comparison Curve of the Actual and Predicted Traffic Flow at the Aetection Point 3 of Street.
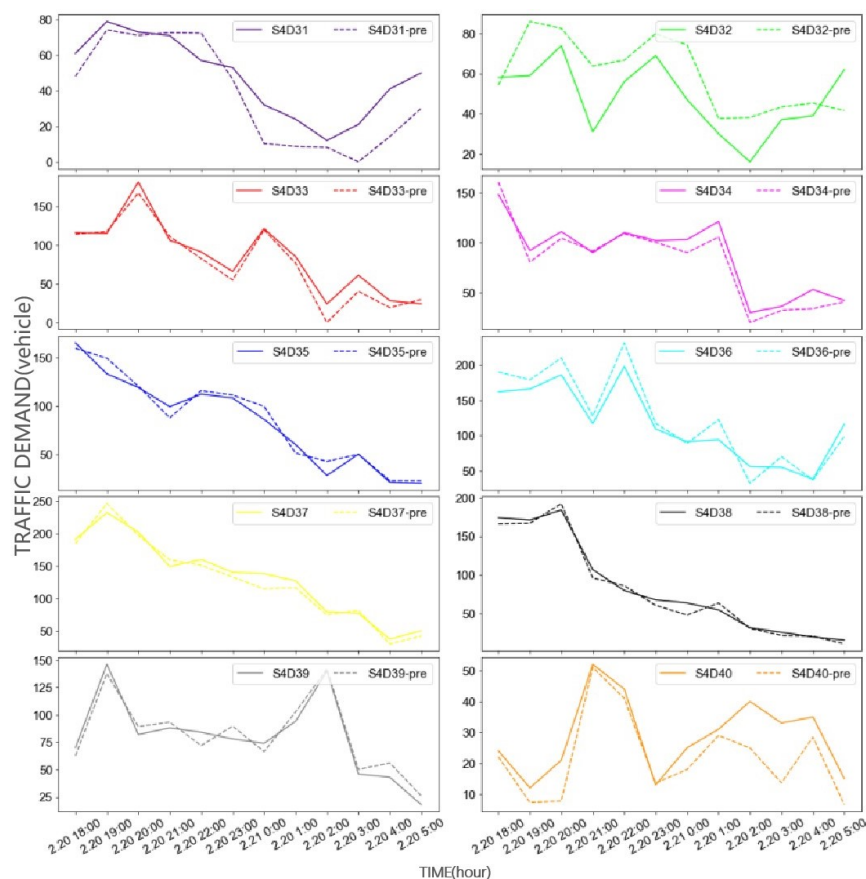
Figure 12. Comparison Curve of the Actual and Predicted Traffic Flow at the Aetection Point 4 of Street.

5. Conclusions

In this paper, we propose a novel method, which is called TS-NAS. We use different network convolution modules and neural architecture to search space by redefining temporal attention mechanism and spatial attention mechanism. Firstly, based on the existing attention, we design a new temporal attention mechanism and spatial attention mechanism according to the temporal and spatial characteristics. Secondly, we design a new model to learn the temporal and spatial characteristics of traffic flow. Finally, the proposed model uses different temporal attention, spatial attention, convolution modules and utilizes neural structure search to optimize the model. TS-NAS can capture the temporal and spatial characteristics of traffic flow and improve the prediction ability of traffic flow at different detection points. The experimental results are on two standard data sets. Compared with the baseline method, it shows that our model has better prediction ability.

Author Contributions: C.Z., R.L., B.S., L.Z., Z.H. and W.Z. designed the project; C.Z. and L.Z. performed the experiment and analyzed the data; C.Z., R.L., B.S. and Z.H. wrote the paper. All authors have read and agreed to the published version of the manuscript.

Informed Consent Statement: Not applicable.

References

1. Yu, K.; Jia, L.; Chen, Y.; Xu, W. Deep Learning: Yesterday, Today and Tomorrow. *Comput. Res. Dev.* **2013**, *50*, 1799.
2. Zha, Y.; Wang, Y. Citation Recommendation Model Based on Bert and GCN. *Comput. Appl. Softw.* **2021**, *38*, 41–45.50.
3. Wang, J.; Li, C.; Xiong, Z.; Shan, Z. Overview of Data-Centered Smart City Research. *Comput. Res. Dev.* **2014**. [CrossRef]
4. Billy M.; Williams, B.M.; Hoel, L.A. Modeling and Forecasting Vehicular Traffic Flow as a Sea- sonal ARIMA Process: Theoretical Basis and Empirical Results. *J. Transp. Eng.* **2003**, *129*, 664–672.
5. Zivot, E.; Wang, J. Vector autoregressive models for multivariate time series. *Model. Financ. Time Ser. S-Plus*, **2006**, 385–429. [CrossRef]
6. Hochreiter, S.; Schmidhuber, J. Long short-term memory. *Neural Comput.* **1997**, *9*, 1735–1780. [CrossRef] [PubMed]
7. Hung, J.; Gulcehre, C.; Cho, K.; Bengio, Y. Empirical Evaluation of Gated Recurrent Neural Networks on Sequence Modeling. *arXiv* **2014**, arXiv:1412. 3555.
8. Liu, B.; Ramsundar, B.; Kawthekar, P.; Shi, J.; Gomes, J.; Luu Nguyen, Q.; Ho, S.; Sloane, J.; Wender, P.; Pande, V. Retrosynthetic reaction prediction using neural sequence-to-sequence models. *ACS Central Ence* **2017**, *3*, 1103–1113. [CrossRef]
9. Yu, B.; Yin, H.; Zhu, Z. Spatio-Temporal Graph Convolutional Networks: A Deep Learning Framework for Traffic Forecasting . In Proceedings of the International Joint Conference on Artificial Intelligence, Stockholm, Sweden, 13–19 July 2018; pp. 3634–3640.
10. Guo, S.; Lin, Y.; Feng, N.; Song, C.; Wan, H. Attention based spatial-temporal graph convolutional networks for traffic flow forecasting. In Proceedings of the AAAI Conference on Artificial Intelligence, Honolulu, HI, USA, 27 January–1 February 2019; Volume 33, pp. 922–929.
11. Bruna, J.; Zaremba, W.; Szlam, A.; Lecun, Y. Spectral networks and locally connected networks on graphs. In Proceesings of the International Conference on Learning Representations, Banff, AB, Canada, 14–16 April 2014.
12. Bahdanau, D.; Cho, K.; Bengio, Y. Neural Machine Translation by Jointly Learning to Align and Translate. In Proceedings of the ICLR 2015, San Diego, CA, USA, 7–9 May 2015; p. 115.
13. Mnih, V.; Heess, N.; Graves, A. Recurrent Models of Visual Attention. *Adv. Neural Inf. Process. Syst.* **2014**, *3*. [CrossRef]
14. Choromanski, K.; Likhosherstov, V.; Dohan, D.; Song, X.; Gane, A.; Sarlos, T.; Hawkins, P.; Davis, J.; Mohiuddin, A.; Kaiser, L.; et al. Rethinking Attention with Performers. *arXiv* **2021**, arXiv:2009.14794v3.
15. Galassi, A.; Lippi, M.; Torroni, P. Attention in Natural Language Processing. *IEEE Trans. Neural Netw. Learn. Syst.* **2020**, *32*, 4291–4308. [CrossRef]
16. Zoph, B.; Le, Q.V. Neural architecture search with reinforcement learning. *arXiv* **2016**, arXiv:1611.01578.
17. Niepert, M.; Ahmed, M.; Kutzkov, K. Learning convolutional neural networks for graphs. In Proceedings of the International Conference on Machine Learning, New York, NY, USA, 20–22 June 2016; pp. 2014–2023.
18. Zhang, Y.; Guan, S.; Xu, C.; Liu, H. Based on spatio-temporal graph convolution networks with residual connection for intelligence behavior recognition. *Int. J. Electr. Eng. Educ.* **2021**, 0020720921996600. [CrossRef]
19. Xu, K.; Ba, J.; Kiros, R.; Cho, K.; Courville, A.; Salakhudinov, R.; Zemel, R.; Bengio, Y. Show, Attend and Tell: Neural Image Caption Generation with Visual Attention. In Proceedings of the International Conference on Machine Learning, Lille, France, 6–11 July 2015.
20. Vlahogianni, E.I.; Karlaftis, M.G.; Golias, J.C. Short-term traffic forecasting: Where we are and where we're going. *Transp. Res. Part C Emerg. Technol.* **2014**, *43*, 3–19. [CrossRef]
21. Castro-Neto, M.; Jeong, Y.S.; Jeong, M.K.; Han, L.D. Online-SVR for short-term traffic flow prediction under typical and atypical traffic conditions. *Expert Syst. Appl.* **2008**, *36*, 6164–6173. [CrossRef]
22. Van Lint, J.W.C.; Van Hinsbergen, C.P.I.J. Short-term traffic and travel time prediction models. *Artif. Intell. Appl. Crit. Transp. Issues* **2012**, *22*, 22–41.
23. Jeong, Y.S.; Byon, Y.J.; Castro-Neto, M.M.; Easa, S.M. Supervised weighting- online learning algorithm for short term traffic flow prediction. *IEEE Trans. Intell. Transp. Syst.* **2013**, *14*, 1700–1707. [CrossRef]
24. Felzenszwalb, P.F.; Huttenlocher, D.P. Efficient graph-based image segmentation. *Int. J. Comput. Vis.* **2004**, *59*, 167–181. [CrossRef]
25. Shi, L.; Zhang, Y.; Cheng, J.; Lu, H. Skeleton based action recognition with directed graph neural networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 15–20 June 2019; pp. 7912–7921.

26. Längkvist, M.; Karlsson, L.; Loutfi, A. A review of unsupervised feature learning and deep learning for time-series modeling. *Pattern Recognit. Lett.* **2014** , *42*, 11–24. [CrossRef]
27. Zhang, J.; Zheng, Y.; Qi, D.; Li, R.; Yi, X.; Li, T. Predicting citywide crowd flows using deep spatio-temporal residual networks. *Artif. Intell.* **2018**, *259*, 147–166. [CrossRef]
28. Yao, H.; Tang, X.; Wei, H.; Zheng, G.; Yu, Y.; Li, Z. Modeling spatial-temporal dynamics for traffic prediction. *arXiv* **2018**, arXiv:1803.01254.
29. Yao, H.; Wu, F.; Ke, J.; Tang, X.; Jia, Y.; Lu, S.; Gong, P.; Ye, J. Deep multiview spatial-temporal network for taxi demand prediction. In Proceedings of the AAAI Conference on Artificial Intelligence, New Orleans, LA, USA, 2–7 February 2018.