

Dilated-Dense Network for traffic congestion prediction Model

Ming Shi, Shaowei Cai, Qingming Yi

School of Information Science and Technology, Jinan University, Guangzhou 510632, China.

ABSTRACT. *It is very important to use the convolutional neural network model for urban short-term traffic congestion forecasting. However, pooling operation process is prone to cause data structure information loss when using CNN model to predict short-term traffic flow. Complete data feature information cannot be transmitted, which reduces the prediction ability of the model. To solve these problems, an innovative model based on dilated convolutional-dense networks is proposed in this paper. Firstly, the model can use the dilated convolution to obtain the characteristics of the larger receptive field with fewer network parameters, and fully extract the complex and variable data features. Then, through up-sampling and densely connected, the problem of parameter degradation in the process of increasing layers of neural network can be solved. Finally, the actual urban road average speed data blocks are taken to verify the validity of the model. Experimental results show that compared with traditional CNN model, the average absolute error of the network prediction structure is reduced by 3% to 23%.*

KEYWORDS: *convolutional-dense network; spatio-temporal characteristics; convolutional neural networks; short-term traffic congestion prediction.*

In the data with uncertainty, non-linearity or time-varying, extracting the spatio-temporal characteristics of the data is an important way to obtain data information. Studying the extraction method of the spatio-temporal characteristics of the data is of great significance to improve the prediction ability of the model [1- 4]. In some complex forecasting scenarios, such as in the prediction of short-term traffic congestion, the vehicle speed data is susceptible to interference factors such as traffic congestion on adjacent roads, holidays, etc., which makes the vehicle speed data fluctuate widely and have a lot of abnormal values. It is just as important as the data features at the time of the anomaly, so an effective method is needed to extract the spatiotemporal feature information of these data. With the widespread application of deep learning models in recent years, more and more deep learning prediction models acquire and use the spatiotemporal characteristics of data through different improvements. Among them, Luo Wenhui [5] and others used CNN (Convolutional Neural Network) to extract the characteristics of the spatiotemporal characteristics of the data, and combined CNN and SVR for short-term traffic flow prediction; Deng S [6] and others proposed a random subspace based on CNN The

learning method converts data into the form of images and extracts the spatio-temporal feature information, thereby improving the prediction ability of the model; Lin F[7] and others use sparse autoencoders to extract the spatio-temporal feature information of the data and use it as LSTM (Long Short- Term Memory), and verified the validity of the model through actual traffic data; Kang D [8] and other data were used as input methods of LSTM at the same time, such as traffic volume, vehicle speed, and detector occupancy, which improved the characteristics of the input data. Diversity and achieve the purpose of different data characteristics to jointly affect the prediction results; An J [9] et al. Used a residual network to increase the number of convolution layers and combined different time interval inputs to establish a fuzzy convolutional neural network depth model, So that the time characteristic information of the data can be fully utilized. The above deep learning models mainly use two models, LSTM and CNN. Because the structure of LSTM itself is more suitable for time series, such as scenes with less volatility, fewer outliers, and more obvious changes, it can be used for prediction. In the scene, the extraction of spatio-temporal feature information is not sufficient. The structure of the CNN model can extract structural information between non-adjacent data due to changes in the receptive field. Therefore, this feature can be used to obtain the spatio-temporal feature information of the data and use it in prediction scenarios with large data volatility and many outliers. Medium [10-11].

Considering the method of extracting the spatio-temporal feature information of data by CNN, there are certain flaws. That is, after the model completes the convolution process, it needs to compress the features through the pooling process and fill the two-dimensional data. This process will make the data internal Loss of structural and spatial hierarchical information. Aiming at this problem, this paper proposes a dilated-dense network structure model. This structure utilizes the characteristics of dilated convolutions [12] (Dilated-Convolutions) to flexibly control the convolution sampling interval, which can reduce the role of the pooling layer while reducing the complexity of the model. In addition, this article opens a second upsampling channel for extracting the salient features of the data, and then densely connects the dilated convolution channel and the upsampling channel in the output part [13-15], so that more spatiotemporal feature information is captured. Transfer, so as to ensure the depth and prediction accuracy of the model structure, so that the model achieves better prediction results in terms of road congestion prediction.

1. Hollow-dense network structure construction

1.1 Convolution block and dense block construction

In this paper, we use the method of feature extraction and classifier of CNN, and pass the input through a layer of convolution to the dense layer with multiple dense blocks as the core to fully extract the features. The convolution kernel size is 1x1 between the dense blocks. The convolutional layer is connected, and the output is performed through two fully connected layers. The algorithm structure flow of this

paper is shown in Figure 1. The number of dense blocks depends on the size of the input matrix. The fully connected layer is an ordinary feedforward network. The activation function is For Sigmoid.

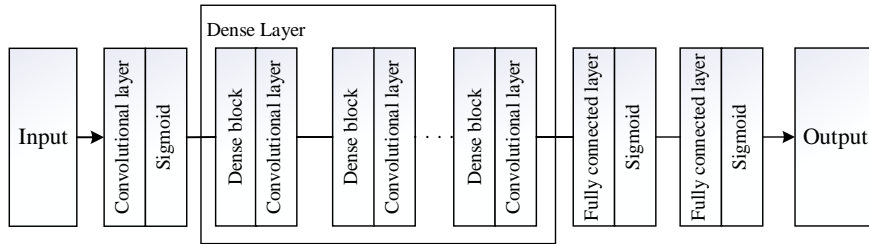


Fig. 1 Flowchart of the proposed algorithm

1.2 Input matrix construction

Considering that the same type of data has a certain continuity in adjacent time intervals and the characteristics of different collected data areas are related, this paper enters the collected data of different areas in a measurement cycle into a two-dimensional matrix I . The expression of the matrix I is shown in formula (1), which $x_{p,t}$ represents the collected data of the P region at the time t , where $1 \leq p \leq i$, $0 \leq t \leq j$, in the matrix, the row data represents the data from the collection time 0 to the collection time j , and the column data represents the data of different collection points at the same time Thus, an input matrix based on spatiotemporal features is constructed and used as the input of the convolution layer.

$$X = \begin{bmatrix} x_{1,0}, x_{1,1}, \dots, x_{1,j} \\ x_{2,0}, x_{2,1}, \dots, x_{2,j} \\ \vdots \\ x_{i,0}, x_{i,1}, \dots, x_{i,j} \end{bmatrix} \quad (1)$$

1.3 Dense block structure

The dense block is mainly composed of two channels. The first dilated convolution channel is composed of three convolution blocks. Each convolution block consists of a layer of dilated convolution layers, batch-normalization layer, and Sigmoid. The activation function layer is composed of the input x and output of the convolution block $f(x)$. The size of the convolution kernel can be changed by changing the value of the dilated convolution k , and the size of the dilated of the convolution kernel is set r . The second upsampling channel is composed of two

maximum pooling layers and a 1×1 convolutional layer; the input of the dense block is X , and the convolutional channel and the upsampling channel are densely connected at the output section of the channel and output Y .

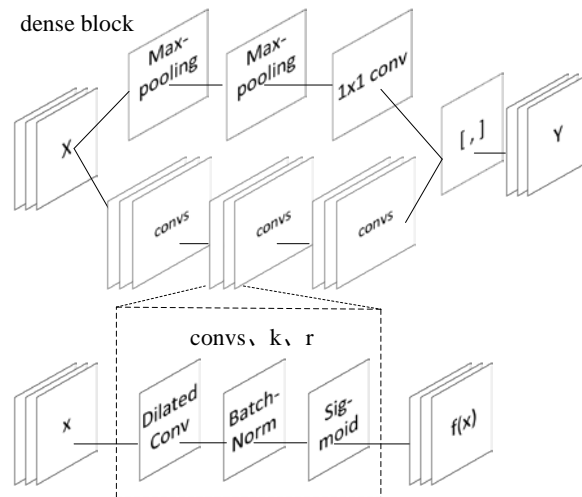


Fig. 2 dense block structure

1.4 Hollow convolution channel structure

Dilated convolution is to add dilateds inside the ordinary convolution kernel, which is mainly achieved by controlling the size of the sampling interval rate of the convolution kernel, as shown in Figure 3, assuming that the sampling interval is 1, 2 and 3, The receptive field under the kernel 3×3 increases with the sampling interval. In addition, with the increase of rate, the number of convolution kernel parameters for dilated convolution is the same as that of traditional convolution kernels, and the output of the convolution process can contain feature information in a larger range of data.

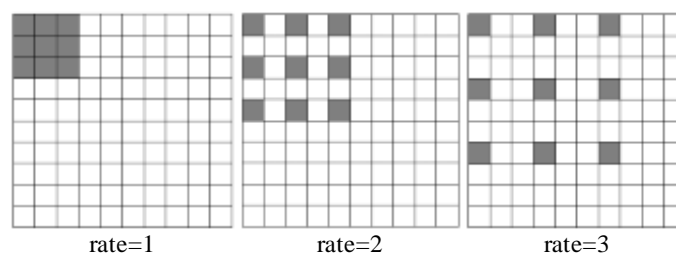


Fig. 3 Convolution kernels with different dilation rate as 1, 2 and 3.

Assume that the convolution kernel size of (k_1, k_2) the dilated convolution is and the input two-dimensional data size is (n, m) ; then, as the rate changes, the convolution kernel size becomes (k'_1, k'_2) , the feature map size becomes (n_1, m_2) , and the influence of the convolution kernel size with the change of the rate meets In formulas (2) and (3), the feature size extracted by the convolution kernel satisfies formulas (4) and (5). Where is the size of r rate, P the step size of the convolution kernel movement, and the number of pixels s filled in the feature map.

$$k'_1 = k_1 + r * (k_1 - 1) \quad (2)$$

$$k'_2 = k_2 + r * (k_2 - 1) \quad (3)$$

$$n_1 = \frac{n + 2p - k - r(k - 1)}{s + 1} \quad (4)$$

$$m_1 = \frac{m + 2p - k - r(k - 1)}{s + 1} \quad (5)$$

Assume that the size of the dilated convolution input I' is $M \times N$, the size of the convolution kernel W is $m \times n$, and $M \geq m$; $N \geq n$, then the convolution operation is shown in formula (6), where $1 \leq i \leq M - m + 1$, $1 \leq j \leq N - n + 1$, the elements obtained by the convolution are C_{ij} .

$$C_{ij} = \sum_{s=1}^m \sum_{t=1}^n I'_{i+m+s} \times W_{st} \quad (6)$$

In order to avoid the parameter change in the dilated convolution block causing the output of the dilated convolution layer to change drastically, this paper adds a batch normalization layer after the dilated convolution layer, and adjusts the output by introducing the mean and standard deviation of the small batch data, So that the overall model output is more stable. Denote the output of the dilated convolutional layer by $f_1(x)$, the output of the batch normalization process is $f_2(x)$, and the output of the batch normalization layer is $f_3(x)$. The data standardization process is shown in formulas (7) and (8).

$$f_2(x) = \frac{f_1(x) - \mu_\theta}{\sqrt{\sigma_\theta^2 + \xi}} \quad (7)$$

$$f_3(x) = \gamma * f_2(x) + \beta \quad (8)$$

$$f(x) = \text{Sigmoid}(w * f_3(x) + b) \quad (9)$$

Where μ_θ is the mean of the mini-batch data, σ_θ^2 is the variance of the mini-batch data, and ξ is a relatively small constant. γ is the lift coefficient and β is the offset parameter. The importance of the normalization layer can be adjusted by γ and parameters β . The output of the normalization layer is $f(x)$ obtained by the activation function of the convolution block. The activation function is formula (9), where w is the weight coefficient and b is the offset coefficient.

1.5 Upsampling Channel Structure

Because the maximum pooling layer has the significant characteristics of retaining data and can be scaled down [2], this paper selects the maximum pooling layer to up-sample the input dense block data, and passes the up-sampled features through the volume The convolution layer with the kernel size is 1×1 used as the output of the channel, so that significant features can be transferred to the output part, and the feature block can be adapted to the output feature block size of the hollow convolution layer.

Considering the current feature block connection methods mainly include residual and dense connections. Compared with residual networks, dense connections combine the advantages of residual connections to make the network better converge, while only requiring fewer parameters. Can reduce the number of overall model parameters. Therefore, in the dense block of this paper, the two channels are densely connected in the output part. The dense connection is shown in formula (10), where Y is the output and the output of the hollow convolution channel is f and the output of the upsampling channel y are densely connected.

$$Y = [f, y] \quad (10)$$

1.6 Network training

The process of training a dilated-dense network structure consists of two parts. The input matrix is forward propagated through the network model to obtain prediction data for the next detection period. Mean square error analysis is performed on these prediction data. The loss function is shown in formula (11) And will get the error back-propagated.

$$L = \frac{1}{2N} \sum_{s=1}^m \|O_i - Z_i\|_2^2 \quad (11)$$

In formula (2.11), N is the batch size of the input sample, O_i is the output value of the dilated-dense network structure, and Z_i is the true value of the data. Considering that the measured data selected in this paper has the characteristics of complex and changeable and noisy, this paper chooses Adam (Adaptive Moment

Estimation) [16] gradient descent algorithm to update the model parameters. Assuming the moment t , the first derivative of the loss function L for the parameter is g_t , The first and second moment estimates of the gradient are shown in equations (12) to (15).

$$m_t = \beta_1 m_{t-1} + (1 - \beta_1) g_t \quad (12)$$

$$v_t = \beta_2 v_{t-1} + (1 - \beta_2) g_t^2 \quad (13)$$

$$m_t' = \frac{m_t}{1 - \beta_1^t} \quad (14)$$

$$v_t' = \frac{v_t}{1 - \beta_2^t} \quad (15)$$

Where m_t is the exponential moving average of the gradient, v_t is the squared gradient, β_1 is the exponential decay rate of the first-order moment estimation, β_2 is the exponential decay rate of the second-order moment estimation, m_t' and v_t' is the value after the modified moment estimation, so that the weight of the parameter θ_t is updated to θ_{t+1} , As shown in formula (16). In the formula, Adam algorithm can adapt different learning rates to different model parameters through the first-order moment estimation and second-order moment estimation of the gradient, so that the model parameter changes are relatively stable, which is beneficial to the extraction of data features.

$$\theta_{t+1} = \theta_t - \frac{\eta m_t'}{\sqrt{v_t'} + \epsilon} \quad (16)$$

2. Experiment and analysis

2.1 data preprocessing

In order to verify the effectiveness of the dilated-dense network structure, this paper chooses to predict the situation of urban traffic congestion. The data in this article are from the 214 anonymous sections of Guangzhou provided by the OpenITS Alliance from August 1, 2016 to September 30, 2016. They are mainly measured speed data of trunk roads and expressways. The data monitoring period is 10 minutes. The vehicle speed data during the period of 00:00-24:00 is selected for training and prediction. A total of 221,760 data in the first 55 days of the data are used as training data, and a total of 24,192 data in the next 6 days are used as prediction data. Because the data value of the vehicle will fluctuate greatly when

congestion occurs, the model will be greatly affected by the local data, resulting in a poor degree of fit between the prediction result and the real data, and low prediction accuracy. Therefore, this paper uses Z-score normalization method to preprocess the input data. The preprocessed data conforms to the standard normal distribution, which reduces the impact of numerical differences. The preprocessing is shown in formula (17), where The data is x , the mean of the input matrix samples is μ and the standard deviation of the input matrix samples is σ . After normalization, the output data value μ' is between [0, 1].

$$\mu' = \frac{x - \mu}{\sigma} \quad (17)$$

2.2 Network Structure Parameters

Considering the influence of the input matrix size and the sampling interval of the dilated convolution on the prediction accuracy of the model, this paper sets the model's convolution kernel to a uniform size, the model's learning rate is 1e-2, the batch size is 32, and the number of dense blocks is 3. Experiment in Python's IDE PyCharm. By testing the number of different roads under three different sampling intervals, the number of roads is increased by 4 each time, and the number of training iterations of the model is 8000. The loss function values are shown in Table 1.

The loss function value obtained in Table 1 is, and the smaller the loss function value, the smaller the error between the true value and the predicted value. Because the number of roads in the input matrix determines the number of rows in the input matrix, when the number of roads is 28 and the sampling interval is 1, 2, 4, or 1, 2, 5, the requirements of the structure of this paper cannot be met, so the value of the loss function is empty, which is suitable for sampling The interval is 1, 2, 3, and the same is true in other cases. As can be seen from the table, as the input matrix increases, the sampling interval needs to be increased to reduce the impact on prediction accuracy.

Table 1 Road number and sampling interval configuration

roads	Loss value		
	rate		
	1, 2, 3	1, 2, 4	1, 2, 5
28	0.0023	-	-
32	0.0026	0.0026	-
36	0.0029	0.0031	0.0027
38	0.0034	0.0032	0.0030
40	0.0037	0.0033	0.0033
42	0.0039	0.0036	0.0035

2.3 Comparative analysis of intuitive effects

In this paper, the data of two days adjacent to September 26 and September 27 are selected as the test data sets of holidays and normal days and expressed in the same block diagram, where September 26 is a holiday and September 27 is a normal day. According to the data in Table 1, the number of roads is 28, and the sampling intervals are 1, 2, and 3 for observation. As shown in Figure 2, the abscissa represents time, the ordinate represents speed, the solid line is the true value, and the dotted line is the predicted value. The value corresponding to 0 to 24 hours in the ordinate is the change in vehicle speed during the monitoring period on September 26, and the value of the ordinate corresponding to 24 to 48 hours indicates the corresponding in the monitoring period on September 27. Changes in vehicle speed. This article randomly lists the measured and predicted values of the actual speed of four roads. As shown in Figure 2, the four graphs are the actual and predicted changes in the vehicle speed of the four roads. It can be seen from the figure that the real value represented by the solid line and the predicted value represented by the dotted line have a similar trend of change, and the degree of fitting is high. When the road is congested at about 8 o'clock on the first day and about 12 o'clock on the second day, the error between the predicted value of the vehicle speed and the true value is small, and in other time periods where the congestion is relatively gentle, the degree of coincidence between the predicted value curve and the true value is higher.

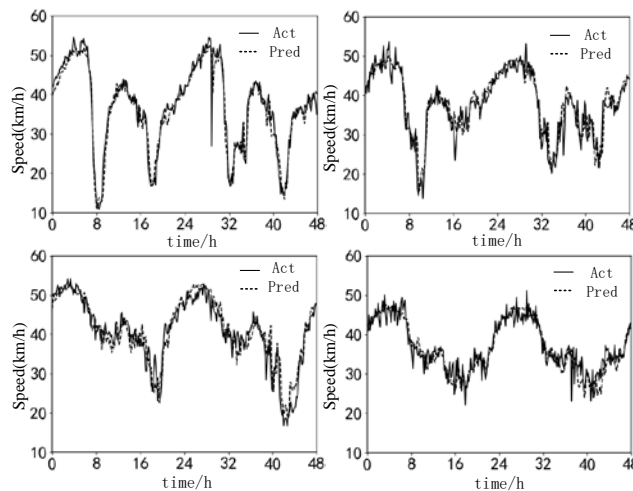


Fig. 4 Compare the true and predicted values of multiple roads

The prediction curve of the hollow-dense structure can more closely reflect the change of the real traffic flow on normal working days and holidays, indicating that the hollow-dense structure can effectively extract the characteristics of large fluctuation points or outliers in the input matrix, and can correctly predict Changes

in the constant part. In addition, the prediction data of the model is composed of prediction data of 28 roads, so the model in this paper can predict 28 roads without increasing the amount of calculation, which greatly reduces the complexity of the overall vehicle congestion prediction system.

2.4 Comparative analysis of model residuals

In order to verify the robustness of the model, this paper introduces the LeNet model, dilated convolution model, dilated-residual structure model, and dilated-dense structure model. The true and predicted values of these models on normal working days and holidays The residual values are analyzed. According to the data in Table 1, the number of roads is 28, and the sampling intervals are 1, 2, and 3 for analysis. As shown in Figure 3, the abscissa represents time and the ordinate represents vehicle speed.

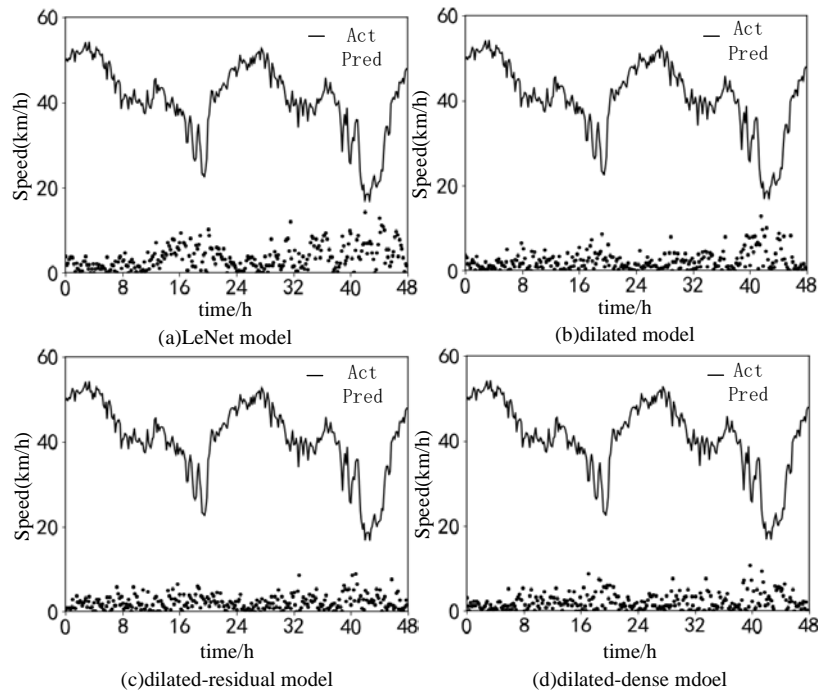


Fig. 5 Comparison of the residuals between four models on normal working days and holiday

The curve in Figure 5 shows the actual vehicle speed change over two days, and the scattered points are the residual values corresponding to the time of the sampling point. It can be seen from the figure that the residual value of the LeNet model at

different times is larger than that of other models using dilated convolution, and the residual value changes of the four models are more obvious in the time period where the speed change is obvious. The dilated-residual structure model and the dilated-dense structure model have smaller residual values than the dilated-convolution model, and the dilated-dense convolution structure model has smaller overall residual values than the dilated-residual model. From the test results, compared with the other three models, the residual value of the cavity-dense structure model can keep a small value compared with the other three models.

2.5 Comparative analysis of objective parameters

In this paper, LeNet model, dilated convolution model, and dilated-residual structure model are analyzed by Mean Absolute Error (MAE) and Root Mean Square Error (RMSE) to verify the validity of the model structure and ensure the consistency of network model parameters. The four models were trained on the training set to predict vehicle speeds on September 26 and 27, respectively. Select the first road data in Figure 2 for average absolute error (MAE) and root mean square error (RMSE) analysis. MAE reflects the actual situation of the predicted value error, while RMSE reflects the predicted error distribution, MAE and RMSE values. The smaller it is, the better the prediction model fits and the higher the prediction accuracy. MAE and RMSE are shown in formulas (18) and (19), where are the predicted values is F_i , the actual measured values is Z_i , and the samples is N.

$$MAE = \frac{1}{2N} \sum_{s=t} |F_i - Z_i|^2 \quad (18)$$

$$RMSE = \frac{1}{2N} \sqrt{\sum_{s=t} (F_i - Z_i)^2} \quad (19)$$

The MAE and RMSE analysis results of the four models are shown in Table 2. Compared with the LeNet model and the dilated convolution model, the MAE and RMSE of the dilated-residual network structure model and the dilated-dense structure model both have a certain decrease, and the dilated- The dense structure has smaller MAE and RMSE values than the cavity residual structure model. In the test set, the MAE of the cavity-dense model is reduced by about 3% to 23% , and the RMSE is reduced by 2% to 26% .

Table 2 The comparison of MAE and RMSE between dilated convolution-dense network and the three other models

	MAE			RMSE		
	Holiday	Normal	Test set	Holiday	Normal	Test set
LeNet	3.16	3.29	3.17	4.93	5.01	4.67
Dilated model	2.61	2.65	2.72	3.73	3.79	3.86
Dilated-residual	2.44	2.46	2.51	3.41	3.52	3.50
Dilated-dense	2.35	2.43	2.44	3.32	3.49	3.43

3. concluding remarks

This paper proposes a traffic congestion prediction model that combines dilated convolution and dense connections to extract the spatio-temporal characteristics of the data. The data is preprocessed into the model in the form of a two-dimensional matrix; the model uses dilated convolution to better extract complex data Spatio-temporal feature information and dense network structure can transfer the feature information more completely, and obtain better prediction results in the test set. This paper compares different models by predicting actual urban traffic congestion, and uses the speed data of multiple roads in different detection periods in Guangzhou as the input of the model and analyzes. The experimental results show that the hollow-dense network structure can better Predict road vehicle speed, the error between the predicted value and the true value is relatively small. In addition, the model can predict multiple roads at the same time, and it can be used as an effective method in the prediction of congested and massive urban road congestion. Because the different interference factors of urban road sections and high-speed road sections show different requirements for the model structure, combining the relevant factors of high-speed road sections, improving the generalization ability of the model, and studying the applicability of the model in other prediction scenarios are the next steps. Important content of research.

Acknowledgments

Fund Project: Youth Program of National Natural Science Foundation of China (61603153), Guangzhou Science and Technology Commission Project (201802010028).

References

- [1] Chen P, Ding C, Lu G, et al. Short-Term Traffic States Forecasting Considering Spatial–Temporal Impact on an Urban Expressway [J]. *Transportation Research Record*, 2016, 2594 (1): 61-72.
- [2] Zhang W, Yu Y, Qi Y. Short-term traffic flow prediction based on spatio-temporal analysis and CNN deep learning [J]. *Transportation Research Record*. 2019, 15 (2):1688-1711.
- [3] Chen C, Wu Q, Gao S. Short-Term Shipping Freight Volume Prediction Based on Temporal-Spatial Features [J]. *Journal of Shanghai Jiaotong University*, 2019, 53 (5):556-562.
- [4] Feng X, Ling X, Zheng H, et al. Adaptive Multi-Kernel SVM With Spatial–Temporal Correlation for Short-Term Traffic Flow Prediction [J]. *IEEE Transactions on Intelligent Transportation Systems*, 2018, 20(6): 2001-2013.
- [5] Luo W, Dong B, Wang Z. Short-term Traffic Flow Prediction Based on CNN-SVR Hybrid Deep Learning Model [J]. *Journal of Transportation Systems Engineering and Information Technology*, 2017, 17(05):68-74.

- [6] Deng S, Jia S, Chen J. Exploring spatial-temporal relations via deep convolutional neural networks for traffic flow prediction with incomplete data [J]. *Applied Soft Computing*, 2019, 78: 712-721.
- [7] Lin F, Xu Y, Yang Y, et al. A Spatial-Temporal Hybrid Model for Short-Term Traffic Prediction [J]. *Mathematical Problems in Engineering*, 2019, 2019: 1-12.
- [8] Kang D, Lv Y, Chen Y. Short-term traffic flow prediction with LSTM recurrent neural network [C]. 2017 IEEE 20th International Conference on Intelligent Transportation Systems (ITSC). IEEE, 2017: 1-6.
- [9] An J, Fu L, Hu M, et al. A Novel Fuzzy-Based Convolutional Neural Network Method to Traffic Flow Prediction with Uncertain Traffic Accident Information [J]. *IEEE Access*, 2019, 7: 20708-20722.
- [10] Wang P, Li L, Jin Y, et al. Detection of unwanted traffic congestion based on existing surveillance system using in freeway via a CNN-architecture trafficNet [C]// 2018 13th IEEE Conference on Industrial Electronics and Applications (ICIEA). IEEE, 2018: 1134-1139.
- [11] Liu Q, Wang B, Zhu Y. Short-Term Traffic Speed Forecasting Based on Attention Convolutional Neural Network for Arterials [J]. *Computer-Aided Civil and Infrastructure Engineering*, 2018, 33 (11): 999-1016.
- [12] Yu F, Koltun V. Multi-Scale Context Aggregation by Dilated Convolutions [C]. *International Conference on Learning Representations (ICLR)*. 2016,1-13.
- [13] Huang G, Liu Z, Van Der Maaten L, et al. Densely connected convolutional networks [C]. *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2017: 4700-4708.
- [14] Dolz J, Gopinath K, Yuan J, et al. HyperDense-Net: A hyper-densely connected CNN for multi-modal image segmentation [J]. *IEEE transactions on medical imaging*, 2018.
- [15] Zhang C, Zhang H, Yuan D, et al. Citywide cellular traffic prediction based on densely connected convolutional neural networks [J]. *IEEE Communications Letters*, 2018, 22(8): 1656-1659.
- [16] Kingma D K, Ba J L, Adam: A Method for Stochastic Optimization [C]. *the 3rd International Conference for Learning Representations*. 2018, 6980—698.