Short-Term Traffic Flow Prediction with Conv-LSTM

Yipeng Liu, Haifeng Zheng*, Xinxin Feng, and Zhonghui Chen
College of Physics and Information Engineering, Fuzhou University, Fuzhou, Fujian, China
* Corresponding Author (zhenghf@fzu.edu.cn)

***Abstract*—The accurate short-term traffic flow prediction can provide timely and accurate traffic condition information which can help one to make travel decision and mitigate the traffic jam. Deep learning (DL) provides a new paradigm for the analysis of big data generated by the urban daily traffic. In this paper, we propose a novel end-to-end deep learning architecture which consists of two modules. We combine convolution and LSTM to form a Conv-LSTM module which can extract the spatial-temporal information of the traffic flow information. Furthermore, a Bi-directional LSTM module is also adopted to analyze historical traffic flow data of the prediction point to get the traffic flow periodicity feature. The experimental results on the real dataset show that the proposed approach can achieve a better prediction accuracy compared with the existing approaches.**

***Index Terms*—Traffic Flow Prediction, Conv-LSTM Module, Spatial-Temporal Correlation**

I. Introduction

With the urban traffic system has been highly developed, traffic congestion has become a serious problem in people's daily life. It has been widely recognized that accurate and timely traffic flow information can help people to plan the travel time and travel route, which mitigates the traffic congestion and reduces unnecessary waste. An accurate short-term traffic flow prediction is the key to provide such an accurate and timely traffic condition[1].

In the past few years, a varying number of methods have been proposed to predict traffic flow. Firstly, several works based on Autoregressive Integrated Moving Average Model (ARIMA) have been proposed for traffic flow prediction[2], [3]. However, a large number of studies have found that the traffic flow data are random and nonlinear. The ARIMA algorithm cannot analyze the nonlinear traffic flow data because it is based on the linear relationship. Furthermore, some machine learning methods have also been proposed, such as time series[4], [5], SVM[6], [7], [8], K-NN[9], [10], [11], and so on. Smith et al. applied K-NN to traffic flow prediction, which just compares the short-term traffic flow data with a large amount of data collected before. The algorithm does not analyze the spatial feature or temporal feature of the data. Therefore, the accuracy of prediction is easily affected by the historical data. Castro et al. proposed a Support Vector Regression (SVR) algorithm to predict traffic flow, which analyzed the nonlinear feature of traffic flow data. Yang et al. also applied Support Vector Machine(SVM) for traffic flow prediction which maps short-term traffic flow data to high dimensions, fits the data curve according to Maximum-Margin standard for traffic flow prediction.

In recent years, deep learning has received much attention. A lot of deep learning methods have been proposed for traffic flow prediction, such as SAE[12], [13], DNN[14], [15], DBN[16], [17], LSTM[18], [19], [20], CNN-LSTM[21]. Huang et al. proposed deep belief network (DBN) algorithm to analyze traffic flow information and predict short-term traffic flow[16]. Then SAE (Stack Denoise Autoencoder) was applied to short-term traffic flow prediction. The data matrix which is input to the SAE combines the spatial and temporal information of traffic flow data. The columns of data matrix represent the spatial data of traffic flow and the rows represent temporal traffic flow information. Long Short-Term Memory (LSTM) has achieved a good performance in the field of natural language process[22]. Zhao et al. proposed an LSTM based method for traffic flow data prediction, which uses LSTM to extract the temporal feature of traffic flow data and get the time trend of traffic flow data. This method only uses the LSTM to process the data, which can only extract the temporal feature of traffic flow data. The analysis of the data is not comprehensive, which leads to an inaccurate traffic flow prediction. Wu et al. proposed a hybrid deep learning framework, which contains CNN module and LSTM module. They use CNN module and LSTM module to extract different features of the data respectively. The CNN module extracts the spatial feature of the traffic flow data in the adjacent area. The LSTM module extracts the temporal feature of traffic flow data. Finally, the features extracted by CNN module and LSTM module are combined to get the traffic flow data prediction. The disadvantage of this method is that the features are not fully fused.

In this paper we propose a deep learning based approach to predict short-term traffic flow. The traffic flow data contains three main characteristics: temporal characteristic, spatial characteristic and periodicity characteristic. We combine Convolution and LSTM to generate a Conv-LSTM module for extracting the spatial-temporal feature of traffic flow and then use a Bi-LSTM to extract the period feature of traffic flow. The main contributions of this paper are list as follows:

- We propose a novel deep learning architecture for short-term traffic flow prediction which combines Convolution and LSTM to generate a Conv-LSTM module to extract the spatial-temporal feature of traffic data, which can fully fuse the temporal characteristics and the spatial

978-1-5386-2062-5/17/$31.00 ©2017 IEEE

characteristics of traffic flow in the adjacent region of the prediction point. We then add Bi-LSTM module to the deep learning model as an auxiliary module to extract the periodic features of traffic data.

- We use the spatial-temporal feature and periodic features of traffic flow extracted by Conv-LSTM and Bi-LSTM module to predict short-term traffic flow, respectively. We use Conv-LSTM module to process the short-term traffic flow data in adjacent region for extracting the spatial-temporal feature and use the bidirectional LSTM to process the historical traffic data of the spot to extract the periodic feature of traffic flow data which outperforms the LSTM based algorithm.
- We propose an end-to-end deep learning architecture for short-term traffic flow prediction without data pre-processing and data features extraction by manual. We present simulation experiments on a real-world data. The results demonstrate that our method can achieve a better performance than the other existing methods.

The remainder of this paper is organized as follows. In Section II, we introduce the related work. In Section III, we introduce the details of our method. In Section IV, we conduct simulations on a real dataset and compare the result with some existing methods. Finally, we draw a conclusion and discuss our future work in Section V.

II. System Model

A. Data Formulation

Since the spots of traffic flow data are not evenly distributed in the region, the location of the collected traffic data presents an asymmetric state and the interval is large. We map the traffic flow data to a one-dimensional vector. We put traffic flow data of the prediction spot into the center of the vector center, and fill the vector by traffic flow data of other spots according to the distance from the prediction spot. The deep neural network is used to analyze the vector to find the spatial correlation between the prediction spot and other spot. We represent a one-dimensional spatial information vector as follows:

$$F_t = (F_1, F_2, F_3, \dots, F_t). \tag{1}$$

The one-dimensional spatial information vector at different moments can be combined together to form a matrix as follows:

$$\mathbf{F} = \begin{bmatrix} F_1^1 & F_1^2 & \dots & F_1^s \\ F_2^1 & F_2^2 & \dots & F_2^s \\ \vdots & \vdots & \ddots & \vdots \\ F_t^1 & F_t^2 & \dots & F_t^s \end{bmatrix} \tag{3.1}$$

where s represents the spot within the region and t represents the time.

Periodic data can be expressed as the following matrices

$$\mathbf{D} = \begin{bmatrix} F_1^1 & F_1^2 & \dots & F_1^s \\ F_2^1 & F_2^2 & \dots & F_2^s \\ \vdots & \vdots & \ddots & \vdots \\ F_d^1 & F_d^2 & \dots & F_d^s \end{bmatrix} \tag{3.2}$$

and

$$\mathbf{W} = \begin{bmatrix} F_1^1 & F_1^2 & \dots & F_1^s \\ F_2^1 & F_2^2 & \dots & F_2^s \\ \vdots & \vdots & \ddots & \vdots \\ F_w^1 & F_w^2 & \dots & F_w^s \end{bmatrix} \tag{3.3}$$

where d represents the same moment in yesterday and w represents the same moment in the last week.

B. Overview

We input the above data matrix into the proposed deep neural network in this paper. As shown in Fig.1, our method mainly incorporates two parts, where the first part is Conv-lstm module and the second part is Bi-LSTM module. We use the Conv-LSTM to extract the spatial-temporal feature of traffic flow data and use Bi-LSTM to extract periodicity feature. Finally, spatial-temporal feature and periodicity feature are concentrated to predict the traffic flow. The overall structure of the proposed deep neural network is shown in Fig. 1.

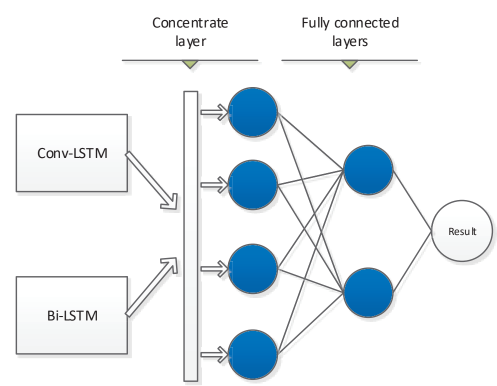

Fig. 1. The proposed deep architecture model for short-term traffic flow prediction

In the training phase, we use MSE (Mean Squared Error) as the loss function for our methods, which can accurately describe the difference between the true value and prediction value. MSE can be expressed as follows:

$$MSE = \frac{1}{n} \sum_{t=1}^{n} (F_p - F_t)^2 \tag{2}$$

where F_p is the prediction value of short-term traffic flow, and F_t is the true value of traffic flow.

Then we use the RMSprop algorithm to optimize the deep neural network. RMSprop optimization algorithm can adjust the learning rate adaptively. Firstly the RMSprop algorithm uses the variable MeanSquare (w, t) to preserve the mean square of the gradient squared for each period of the previous time. According to this variable, the learning rate is self-adaptive. The RMSprop function is shown as follows:

$$MeanSquare(w,t) = 0.9MeanSquare(w,t-1)+0.1(\frac{\partial E}{\partial w^t})^2 \quad (3)$$

where w is the weights of deep neural work, t is the training steps, $\frac{\partial E}{\partial w^t}$ is the gradient of w in the training step t.

C. Conv-LSTM

The input matrix of Conv-LSTM is shown in (3.1). The matrix can be expressed as a time series vector $F_t^s = (F_1^s, F_2^s, F_3^s, \ldots, F_t^s)$. Each element in the vector represents the traffic flow data of all spots in the adjacent region of the spot to be predicted at the same time. Each element in the vector can be represented as $F^s = (F_1, F_2, F_3, \ldots, F_t)$.

We use 1D-Conv to process each element in F_t^s. A one-dimensional convolution kernel filter is used to acquire convolution information of the local perceptual domain by sliding filter. Such a process is beneficial to extract the spatial characteristics between near spots. Then, the local features are aggregated to form the global feature. A unit node generated by each step slides over the vector. The unit node can be represented as:

$$G(i) = F(Aw + B) \quad (4)$$

where w is the filter weights of the node, B is bias, A is the value of the input node and F is the activation function.

The pooling filter is applied over data vector. The difference is that pooling filter does not perform complex convolution operations. Instead, the values are simply averaged. Pooling reduces the size of vectors effectively. Some unnecessary information is filtered out during the pooling process to obtain more abstract data. The generated feature sequence C is reduced to the half size of its original dimension through the pooling. Such two feature extractions make the deep neural network have higher distortion tolerance in dealing with traffic flow data.

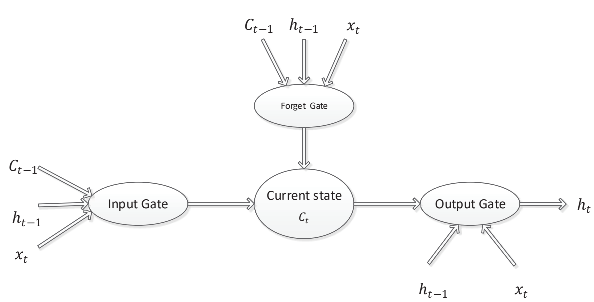

Fig. 2. The LSTM Structure

After convolution and pooling processes each element of time series vector respectively, the output results become time series vector as $C_t = (C_1, C_2, C_3, \ldots, C_t)$. Each element in the vector is the spatial correlation of traffic flow between spots in the region. The time series C is used as the input to

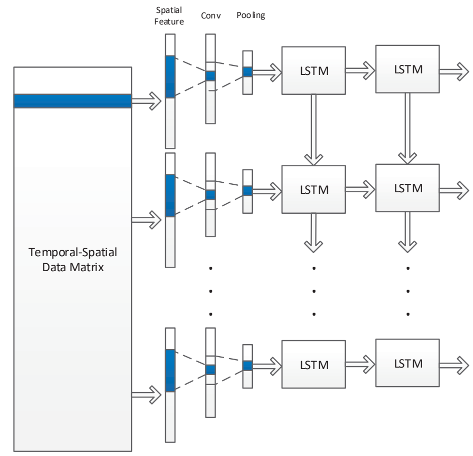

Fig. 3. The Conv-LSTM Structure

the LSTM layer. The structure of the LSTM is shown in Fig. 2. LSTM block contains a cell which stores block cell state, and three gates which control the updates of the input of the cell state and the output of LSTM block, respectively. Firstly, the LSTM cell executes the Forget Gate to determine which part of the information is lost from the previous state of the cell. Secondly, the cell then needs to decide what information should be kept in the input data. The Input Gate determines which values in the LSTM module are updated. Thirdly, a tanh activation function is used to create an intermediate state vector S_t with the acceleration data at this moment and the output of the LSTM module at the last moment. Then the cell value is updated. The value of Forget Gate and the cell value of the $t-1$ moment are multiplied to obtain the values preserved at the last moment. The value of Input Gate and the state vector generated by the tanh function are multiplied to obtain the new value which is added to the cell. The two values are added together to the updated value of the cell. Finally the output of LSTM is obtained by the output Gate. The output is only related to the current cell value and the current cell value depends on the previous Cell value. The whole procedure can be represented as follows:

$$ForgetGate \qquad G_f = \sigma(W_f \cdot [T_{t-1}, I_t] + b_f) \quad (5)$$

$$InputGate \qquad G_i = \sigma(W_i \cdot [T_{t-1}, I_t] + b_i) \quad (6)$$

$$Intermediatestate \qquad S_t = tanh(W_c \cdot [T_{t-1}, I_t] + b_c) \quad (7)$$

$$UpdateCell \qquad C_t = G_f * C_t - 1 + G_i * S_t \quad (8)$$

$$OutputGate \qquad G_o = \sigma(W_o \cdot [T_{t-1}, I_t] + b_o) \quad (9)$$

$$Output \qquad T_t = G_o * tanh(C_t) \quad (10)$$

where σ is the activation function, and G_f, G_i, G_o are the Forget Gate, Input Gate and Output Gate, respectively, W_f, W_i and W_o are their weights, b_f b_i and b_o are their bias, S_t is the Intermediate state of t moment, and S_t is the output of t moment, and W_c and b_c is the weight and bias of Intermediate state. Each layer of the LSTM structure has a Dropout structure which can be used to effectively avoid overfitting. The Conv-LSTM module structure is shown in Fig. 3.

D. Bi-directional LSTM

The traffic flow also has strong periodic features. In this paper, periodic features of traffic flow will be added as supplementary information to predict short term traffic flow. When extracting the periodic feature of traffic flow data, we will deal with the traffic flow information at the same time on the last day and at the same time of the last week.

Unlike the real-time data, we only have data before the prediction time. The historical data contains the full data of that time period. The ordinary LSTM analyzes the information in the time window and get the prediction of the next moment. In this way, the output value of the time series information is only related to the information of the previous time. After dealing with historical information, we obtain full time series data. We use Bi-LSTM to extract periodic feature. The structure of bidirectional LSTM is composed of two unidirectional LSTM stacked up and down. Therefore, the Bi-LSTM input contains the time series before and after the forecast time. At each point the T input is supplied to the two opposite directions of the LSTM, and the output is determined by the two LSTM. Each one-way propagation is the same as the forward LSTM propagation algorithm introduced in the previous section. The Bi-LSTM structure is shown in Fig. 4, where x_i denotes the input of LSTM, O_f denotes the output of forward LSTM and O_b denotes the output of back LSTM.

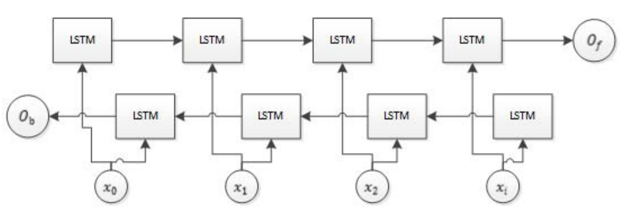

Fig. 4. The Bi-LSTM Structure

III. Simulations

A. Data description

In this paper, we propose an end-to-end deep learning architecture for short-term traffic flow prediction. Deep learning architecture are trained and evaluated by using the Caltrans Performance Measurement System (PeMS) dataset. The detectors used in PeMS collect traffic flow data every 30s. We conduct experiments on the data at two spots where the first one is freeway traffic data and the other is the urban traffic data. The freeway traffic data we use is collected at Freeway SR99-S District 10 and the urban traffic flow data we use is collected at Street I980-E District 4 in Oakland City. The Fig.5 shows a satellite map of the locations which are Freeway SR99-S District 10 and street I980-E District 4 in city of Oakland. The Fig.6 shows the comparison of the traffic flow data of freeway and urban street. Fig. 7 shows a line chart of traffic flow information of I980-E District 4 traffic flow data in 2017/9/4-2017/9/8. From the line chart, we can see that traffic flow data shows strong periodic characteristics and the traffic flow data at the same time tends to show a similar state.

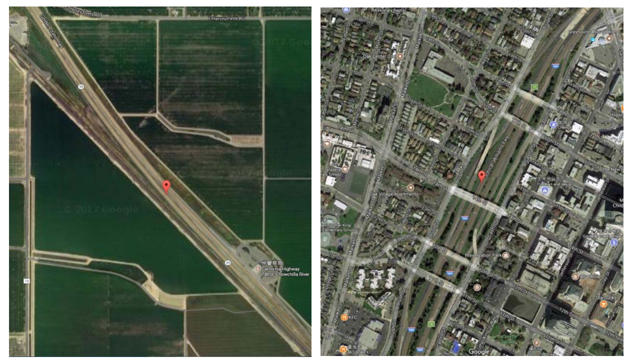

Fig. 5. Freeway SR99-S and Street I980-E Satellite Map

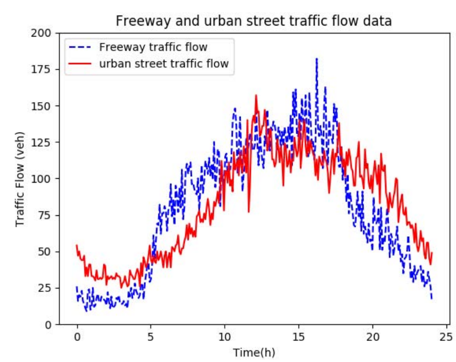

Fig. 6. Freeway and Urban traffic flow comparison

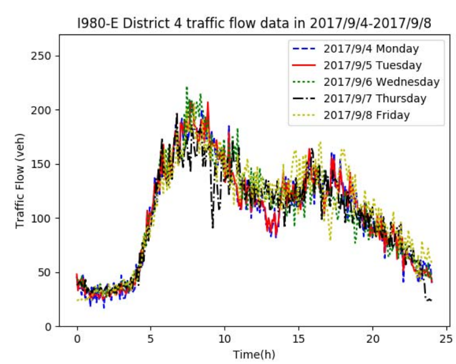

Fig. 7. I980-E District 4 traffic flow data in 2017/9/4-2017/9/8

B. Performance metrics

In order to evaluate the performance of the proposed methods, we use three performance metrics to evaluate the prediction results: Mean Absolute Error(MAE), Mean Absolute Percentage Error(MAPE), Root Mean Square Error(RMSE), which can be expressed as follows:

$$MAE = \frac{1}{n}\sum_{t=1}^{n}|F_p - F_t| \tag{11}$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{t=1}^{n}(F_p - F_t)^2} \tag{12}$$

$$MAPE = \frac{100}{n}\sum_{t=1}^{n}|\frac{F_t - F_p}{F_t}| \tag{13}$$

where F_p represents the prediction value of short-term traffic flow, and F_t represents the true value of traffic flow.

C. Performance with different deep learning architecture

For the Conv-LSTM module, we conduct the experiment on the different combination of convolution and LSTM. The first one is the module proposed in our paper, and the other is the CNN-LSTM model proposed by Wu et al.[12]. The difference between our module and their module is that they combine Convolution and LSTM in two parallel modules to extract the temporal feature and spatial feature individually. Instead, we combine convolution and LSTM as one module, which can fully extract the temporal-spatial feature of short-term traffic flow data. Fig.8 shows the comparison results of the above two models. It can be seen that the prediction accuracy of the Conv-LSTM module is much higher than the CNN-LSTM module.

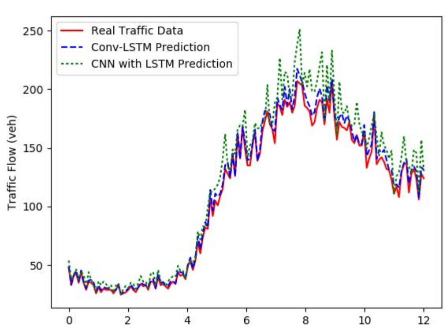

Fig. 8. Prediction performance of the models of Conv-LSTM and CNN-LSTM.

Since the traffic flow data has periodic feature, in our method we use the Bi-LSTM module to extract the periodic feature of the traffic flow data. Fig. 12 shows the prediction performance of the short-term traffic flow with and without the Bi-LSTM module. It is obvious that the Bi-LSTM module can improve the prediction accuracy of the proposed deep learning based architecture.

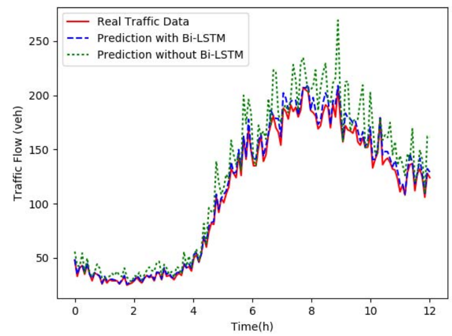

Fig. 9. Prediction performance with Bi-LSTM and without Bi-LSTM

D. Performance comparison with different prediction methods

In this section, we conduct simulations to compare the prediction performance of our method with the other existing methods including ARIMA, SVM, SAE, LSTM, CNN-LSTM by using the traffic flow data of I980-E District 4 in the city of Oakland. The comparison results are shown in Table 1. The results demonstrate that our method can achieve better performance than the other existing methods. We also conduct simulations on the traffic flow data of the Freeway SR99-S District 10. Table 2 shows the prediction performance of our method on the urban traffic data and freeway traffic data. It can be seen that our method can also obtain an accurate prediction performance for freeway traffic data. Figs.10-15 show the traffic flow prediction results of freeway short-term traffic flow.

TABLE I
TRAFFIC FLOW PREDICTION PERFORMANCE OF DIFFERENT METHODS.

Algorithm	MAE/	MAPE/%	RMSE
ARIMA	10.5814	18.7362	12.8993
SAE	8.097386	16.2639	10.38461
LSTM	7.372552	14.5863	9.709646
SVM	6.334922	12.51	9.355609
CNN-LSTM	5.51258	10.2608	8.51564
Our Method			
Without Bi-LSTM	5.253052	9.5281	8.1577
With Bi-LSTM	**4.408279**	**6.98939**	**6.419477**

TABLE II
TRAFFIC FLOW PREDICTION PERFORMANCE OF DIFFERENT SPOTS.

Spots	MAE/	MAPE/%	RMSE
Freeway	4.953052	7.8281	7.4577
Urban	4.408279	6.98939	6.419477

Among the neural network algorithms, the accuracy using SAE neural network is worst. When using the SAE algorithm to predict the traffic flow, we use greedy layerwise unsupervised learning algorithm to train the deep neural network. When we train the first layer, we use the training set as input, which is trained as a single autoencoder. When we train the second layer, it is the same as the training method of the first layer. The hidden layers are connected together to fine-tuning.

Since LSTM's short-term memory characteristics are suitable for processing time-series data, the performance of LSTM is better than SAE. However, we notice that the prediction accuracy of LSTM is poor since LSTM only considers the temporal feature. In contrast, since CNN-LSTM can process the spatial-temporal and periodic features it can achieve a higher accuracy. However, since two independent modules deal with temporal and spatial feature respectively the temporal and spatial characteristics are not fused sufficiently in this approach. Instead, our method can fully extract the spatial-temporal feature of traffic flow data and combine with periodic feature. Therefore, our approach can obtain the best prediction accuracy when using the same dataset.

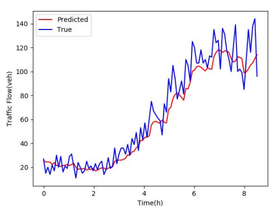

Fig. 10. ARIMA

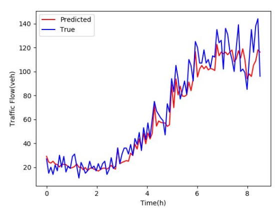

Fig. 11. SAE

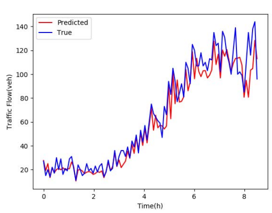

Fig. 12. LSTM

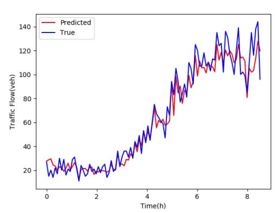

Fig. 13. SVM

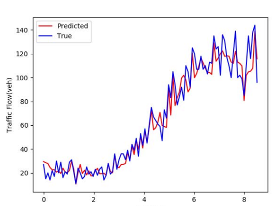

Fig. 14. CNN-LSTM

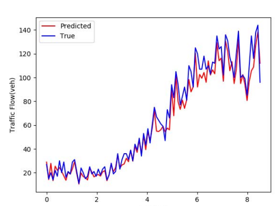

Fig. 15. The proposed method

IV. Conclusion and Future Work

In this paper, we proposed an end to end deep learning method with Conv-LSTM module and Bi-LSTM module for short-term traffic flow prediction. The existing approaches have shortcomings such as incomplete feature extraction, insufficient and incomplete feature fusion and so on. Unlike those methods, we predict traffic flow information by fully integrating the temporal and spatial characteristics of traffic flow information, and supplementing by periodic features. For the future work, we will deal with more types of traffic data to improve the accuracy of short-term traffic flow prediction.

References

[1] J. Zhang, F. Wang, K. Wang,"Data-Driven Intelligent Transportation Systems: A Survey," *IEEE Transactions on Intelligent Transportation Systems*, vol.12, no.4, pp.1624-1639, 2011.
[2] M. Ahmed, A. Cook, "Analysis of Freeway Traffic Time-Series Data by Using Box-Jenkins Techniques,"*Transportation Research Record: Journal of the Transportation Research Board* vol.722,no.1,pp.1-9, 1979.
[3] W. Billy, "Multivariate Vehicular Traffic Flow Prediction: Evaluation of ARIMAX Modeling," *Transportation Research Record: Journal of the Transportation Research Board* vol.1776,no.1, pp.194-200, 2001.
[4] M. Voort, M. Dougherty, S. Watson, "Combining kohonen maps with arima time series models to forecast traffic flow," *Transportation Research Part C: Emerging Technologies*, vol.4, no.5, pp.307-318, 1996.
[5] Z. Zhu, B. Peng, C. Xiong, L. Zhang, "Short-term traffic flow prediction with linear conditional Gaussian Bayesian network,"*Journal of Advanced Transportation* vol.50, no.5, pp.1111-1123, 2016.
[6] M. Castro-Neto, Y. Jeong, M. Jeong, L. Han, "Online-SVR for short-term traffic flow prediction under typical and atypical traffic conditions," *Expert Systems with Applications*, vol.36, no.3, pp.6164-6173, 2009.
[7] Y. Jeong, Y. Byon, M. Castro-Neto, S. Easa, "Supervised Weighting-Online Learning Algorithm for Short-Term Traffic Flow Prediction," *IEEE Transactions on Intelligent Transportation Systems*, vol.14, no.4, pp.1700-1707, 2013.
[8] W. Hong, Y. Dong, and F. Zheng, "Hybrid evolutionary algorithms in a SVR traffic flow forecasting model," *Applied Mathematics and Computation, vol. 217*, no.15, pp.6733-6747, 2011.
[9] H. Chang ,Y. Lee, B. Yoon, S. Baek, "Dynamic near-term traffic flow prediction: systemoriented approach based on past experiences," *IET Intelligent Transport Systems*, vol.6, no.3, pp.292-305, 2012.
[10] B. Smith, B. Williams, R. Oswald, "Comparison of parametric and non-parametric models for traffic flow forecasting," *Transportation Research Part C*, vol.10, no.4, pp.303-321, 2002.
[11] D. Fan, X. Zhang, "Short-term Traffic Flow Prediction Method Based on Balanced Binary Tree and K-Nearest Neighbor Nonparametric Regression," *International Conference on Modelling, Simulation and Applied Mathematics*, 2017.
[12] Y. Lv, Y. Duan, W. Kang, Z. Li, FY. Wang, "Traffic Flow Prediction With Big Data: A Deep Learning Approach," *IEEE Transactions on Intelligent Transportation Systems*, vol.16, no.2, pp.865-873, 2015.
[13] H. Shin, MR. Orton, D. Collins, S. Doran, M. Leach, "Stacked Autoencoders for Unsupervised Feature Learning and Multiple Organ Detection in a Pilot Study Using 4D Patient Data," *IEEE Transactions on Pattern Analysis & Machine Intelligence*, vol.35, no.8, pp.1930-1943, 2013.
[14] J. Zhang, Y. Zheng, D. Li, "DNN-based prediction model for spatio-temporal data," *ACM Sigspatial International Conference on Advances in Geographic Information Systems* pp.92, 2016
[15] M. Jun, L. Xiao, Y. Meng, "Research of Urban Traffic Flow Forecasting Based on Neural Network," *Acta Electronica Sinica*, vol.37, no.5, pp.1092-1094, 2009.
[16] W. Huang, G. Song, H. Hong, K. Xie, "Deep Architecture for Traffic Flow Prediction: Deep Belief Networks With Multitask Learning," *IEEE Transactions on Intelligent Transportation Systems*, vol.15, no.5, pp.2191-2201, 2014.
[17] H. Tan, X. Xuan, K, Wu, Y. Zhong, "A Comparison of Traffic Flow Prediction Methods Based on DBN," *16th COTA International Conference of Transportation*, pp.273-283, 2016.
[18] R. Fu, Z. Zhang, L. Li, "Using LSTM and GRU neural network methods for traffic flow prediction," *Chinese Association of Automation IEEE*, pp.324-328, 2017.
[19] N. Polson, V, Sokolov, "Deep learning for short-term traffic flow prediction," *Transportation Research Part C: Emerging Technologies*, vol.79, pp.1-17, 2017.
[20] Z. Zhao, W. Chen, X. Wu, P. Chen, J. Liu, "LSTM network: a deep learning approach for short-term traffic forecast,' *IET Intelligent Transport Systems*, vol.11, no.2, pp.68-75, 2017.
[21] Y. Wu, H. Tan, "Short-term traffic flow forecasting with spatial-temporal correlation in a hybrid deep learning framework," arXiv:1612.01022, 2016.
[22] R. Zhao, R. Yan, J. Wang, K. Mao, "Learning to Monitor Machine Health with Convolutional Bi-Directional LSTM Networks," *Sensors*, vol.17, no.2, pp.273, 2017.