This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

A Hybrid Deep Learning Model With Attention-Based Conv-LSTM Networks for Short-Term Traffic Flow Prediction

Haifeng Zheng, *Member, IEEE*, Feng Lin, Xinxin Feng, *Member, IEEE*, and Youjia Chen, *Member, IEEE*

***Abstract*— Accurate short-time traffic flow prediction has gained gradually increasing importance for traffic plan and management with the deployment of intelligent transportation systems (ITSs). However, the existing approaches for short-term traffic flow prediction are unable to efficiently capture the complex nonlinearity of traffic flow, which provide unsatisfactory prediction accuracy. In this paper, we propose a deep learning based model which uses hybrid and multiple-layer architectures to automatically extract inherent features of traffic flow data. Firstly, built on the convolutional neural network (CNN) and the long short-term memory (LSTM) network, we develop an attention-based Conv-LSTM module to extract the spatial and short-term temporal features. The attention mechanism is properly designed to distinguish the importance of flow sequences at different times by automatically assigning different weights. Secondly, to further explore long-term temporal features, we propose a bidirectional LSTM (Bi-LSTM) module to extract daily and weekly periodic features so as to capture variance tendency of the traffic flow from both previous and posterior directions. Finally, extensive experimental results are presented to show that the proposed model combining the attention Conv-LSTM and Bi-LSTM achieves better prediction performance compared with other existing approaches.**

***Index Terms*— Traffic flow prediction, deep learning, Conv-LSTM module, attention mechanism, Bi-LSTM.**

I. INTRODUCTION

ACCURATE and real-time short-term traffic flow prediction is of great importance for the daily life of citizens and traffic management. It has the potential to not only help travelers make better route guidance to save money and time but also help government agencies make better route plan to reduce traffic congestion and accidents [1]. Therefore, traffic flow prediction has become one of important function components in intelligent transportation systems (ITSs). However, short-term traffic flow prediction is very challenging due to the stochastic and dynamic traffic condition. In recent years, how to efficiently and accurately predict traffic flow has attracted much attention with the deployment of ITSs.

Manuscript received July 22, 2019; revised February 18, 2020 and March 29, 2020; accepted May 8, 2020. This work was supported in part by NSF China under Grant 61971139, Grant 61571129, and Grant 61601126. The Associate Editor for this article was Y. Lv. *(Corresponding author: Xinxin Feng.)*

The authors are with the Fujian Key Laboratory for Intelligent Processing and Wireless Transmission of Media Information, College of Physics and Information Engineering, Fuzhou University, Fuzhou 350002, China (e-mail: zhenghf@fzu.edu.cn; n171127027@fzu.edu.cn; fxx1116@fzu.edu.cn; youjia.chen@fzu.edu.cn).

Digital Object Identifier 10.1109/TITS.2020.2997352

Previous short-term traffic flow prediction approaches can be roughly classified into three categories: parametric approach, non-parametric approach and hybrid approach [2]. The parametric approach includes time-series methods and Kalman filtering [3], [4]. The widely used models based on time-series methods is the autoregressive integrated moving average (ARIMA) model and its many variants, such as Kohonen ARIMA (KARIMA) [5], subset ARIMA [6], seasonal ARIMA (SARIMA) [7]. However, due to the stochastic and nonlinear nature of traffic flow, these techniques only consider the temporal variation of traffic flow and thus provide unsatisfying prediction performance. The non-parametric approach includes k-nearest neighbor (k-NN) methods, support vector regression (SVR) [9], and artificial neural networks (ANNs). However, it has been shown that the k-NN methods for traffic flow prediction do not outperform the time-series methods. Furthermore, the traditional machine learning based methods utilize human-crafted features to capture the characteristics of traffic flow, which are inadequate to obtain accurate prediction performance. Moreover, the early works based neural networks usually use shallow networks or with only one hidden layer, which are also unable to capture the uncertain and complex nonlinearity of traffic flow.

Recent years have witnessed a great success of deep learning applied in many fields such as computer vision and speech recognition. Compared to the traditional ANN models, deep learning models use multiple-layer architectures to automatically extract inherent features from a large amount of raw data. Recently, deep learning has inspired a surge of interest in transportation research. A variety of deep learning methods have been proposed for traffic flow prediction [11]–[16]. However, the existing works based on deep learning models for traffic flow prediction suffer from the following drawbacks: 1) Some works employ a simple neural network model such as stacked autoencoder (SAE), LSTM or CNN, which cannot fully capture the complex features of traffic flow, thus providing limited prediction performance improvement. For example, CNN is usually utilized to extract spatial feature while LSTM is applied to extract temporal feature. 2) Although some works propose hybrid deep learning methods by combining several models to capture multiple features including spatial, temporal and periodic features for traffic flow prediction, they are usually processed independently. Furthermore, the intricate structures existed in traffic flow data are not exploited fully by these works. For example, the past traffic flow at some locations or times may be more important than any others for forecasting future traffic flow. To solve the problems above, we propose a novel hybrid deep learning model with an attention mechanism for short-term traffic flow prediction. The main contributions of this paper are as follows.

- We develop a hybrid deep learning model with Conv-LSTM networks to exploit the spatial-temporal feature of traffic flow. Different from the existing hybrid model for traffic flow forecasting, Conv-LSTM is able to capture spatial-temporal feature more efficiently since it processes spatial and temporal features as a whole, which improves prediction performance.
- We propose a Bi-LSTM module to extract periodic features for traffic flow prediction by taking both daily periodicity and weekly periodicity into account. The proposed module has the ability of capturing variation tendency of the traffic flow from both previous and posterior directions.
- We design an attention mechanism for the Conv-LSTM module to adaptively allocate different levels of attention to a traffic flow sequence at different times. The proposed mechanism is able to automatically distinguish the importance that each flow sequence contributes to the final prediction performance without auxiliary information.
- We conduct experiments to evaluate the effectiveness of the proposed model by using real-world dataset. The experimental results demonstrate that the proposed model achieves better performance than other existing methods for short-term traffic flow prediction.

The remainder of this paper is organized as follows. In Section II, we introduce the related work. In Section III, we present problem formulations. In Section IV, we propose a novel deep learning approach for traffic flow prediction. In Section V, we conduct experiments on the real-world dataset and compare prediction performance with some existing methods. Finally, we conclude the paper in Section VI.

II. Related Work

In this section, we discuss the most related work regarding to short-term traffic flow prediction. As mentioned above, there are mainly three categories: parametric approach, non-parametric approach and hybrid approach.

In the first category, the widely used model belonging to parametric approach is the autoregressive integrated moving average (ARIMA) model, which is a time-series method. The ARIMA model is applied for predicting traffic flow in expressway and urban arterial roads [22], [23]. Subsequently, many variants of ARIMA model were proposed to enhance prediction performance. For example, a KARIMA model which combines the Kohonen network with ARIMA was proposed for traffic prediction to solve the problem that the ARIMA model cannot handle non-linear traffic data [5]. An ARIMAX model was proposed by combining ARIMA with explanatory variables to improve forecasting performance [7]. A Bayesian seasonal ARIMA was proposed for short-term traffic flow forecasting by using the Bayesian method instead of classic methods such as maximum likelihood estimate and least-squares estimate to improve prediction accuracy [24]. A unified spatio-temporal model based on STARIMA (Space-Time Autoregressive Integrated Moving Average) was also proposed to capture the intricate spatio-temporal correlation structure between road traffic and achieves more accurate prediction performance [8]. The above variants solve the inherent problems of the classic ARIMA model, such as the inability to process non-linear data and poor prediction performance.

In the second category, some approaches belonging to the non-parametric model include k-nearest neighbor (k-NN) methods, support vector regression (SVR) [9], [10], artificial neural networks (ANNs). For example, the k-NN method is applied for short-term freeway traffic prediction but achieves worse performance than the linear time-series approach [25]. A dynamic model based on the k-nearest neighbour non-parametric regression (KNN-NPR) was proposed to predict multi-interval traffic flow by using a wealth of historical data to improve prediction accuracy [26]. An online learning weighted support vector regression (SVR) was proposed for short-term traffic flow forecasting [27]. A hybrid SVR model which applies the hybrid genetic algorithm-simulated annealing algorithm (GA-SA) was presented for traffic flow prediction [9]. Furthermore, a variety of ANN-based models were proposed for traffic flow prediction [28]–[33].

In the third category, some hybrid methods by combining several techniques were proposed for traffic flow prediction [34]–[36]. For example, an aggregation approach utilizing the moving average (MA), ARIMA, exponential smoothing (ES), and neural network (NN) models was proposed for traffic flow prediction [34]. An approach combining the ARIMA model with the expectation maximization and cumulative sum algorithms was proposed for traffic flow forecasting [35]. An adaptive hybrid fuzzy rule-based approach was presented for urban traffic flow prediction [36].

Recently, a variety of deep learning based methods have been proposed for traffic flow prediction, which belong to a type of machine learning methods. The deep learning based approach can automatically extract the inherent spatial and temporal features from raw data without any data preprocessing. A deep belief network (DBN) with the multitask learning method is the first model that applies deep learning technique, where the deep belief network is employed for feature learning and the multitask learning method is employed for integrating several tasks together to jointly train the model [11]. A stacked autoencoder (SAE) model was also proposed for traffic flow prediction, which consists of deep layer structure and uses the layerwise greedy algorithm to learn the spatial and temporal features of traffic flow data [12]. A deep learning approach using LSTM was presented to extract the temporal feature of traffic flow [13]. A deep learning architecture using a convolutional neural network (CNN) was proposed to extract spatial-temporal traffic feature for speed prediction in a large-scale transportation network [14]. A deep irregular convolutional residual LSTM network model called DST-ICRL was proposed for urban traffic passenger flow prediction [15]. To extract both temporal and spatial features of traffic flow, a hybrid deep learning framework using CNN and LSTM was proposed for traffic flow prediction, where CNN and LSTM are used to extract the spatial and temporal features of traffic flow, respectively [16]. Furthermore, an attention mechanism was also proposed to learn the importance of the near-term inputs of traffic flow [16]. However, it requires auxiliary information such as road speed to learn the attention weight matrix, which is not preferred when the auxiliary information is unavailable. In addition, a novel method, Periodic-CRN (PCRN), was proposed with taking recurring periodic patterns of spatio-temporal data in account by using convolutional recurrent network (CRN) for crowd density prediction [17]. Recently, CNN has been extended to graph convolutional neural networks (GCN) to learn feature of graph-structured data based on the spectral graph theory [18]. A deep learning framework, Spatio-Temporal Graph Convolutional Networks (STGCN) was proposed to capture comprehensive spatio-temporal correlations of traffic data for traffic flow forecasting [19]. To capture the complex spatial and non-linear temporal dynamics of traffic data, a deep learning framework, Diffusion Convolutional Recurrent Neural Network (DCRNN), was proposed for traffic flow forecasting, where the traffic flow is modeled as a diffusion process on a directed graph [20].

Different from the existing deep learning models for traffic flow prediction, we propose a novel hybrid model integrating CNN and Bi-LSTM networks which is able to capture both spatial-temporal feature and periodic feature of traffic flow more efficiently. Furthermore, different from our previous work [21], an attention mechanism is introduced into our model to automatically distinguish the importance of traffic flow sequence at different times, which does not utilize the auxiliary information such as speed data as done in [16]. These additional modules significantly improve the performance for short-term traffic flow prediction.

III. Problem Formulation

The purpose of traffic flow prediction is to provide accurate and timely traffic flow information of the near future to improve transportation efficiency. The problem of traffic flow prediction can be formulated as follows. Let X_τ^i denote the traffic flow of the ith observation location during the τth time interval. The observation location can be a road, station or spatial identity. At a current time t, the task is to predict the traffic flow of a point of interest (POI) at time interval $(t + h\Delta)$ for some prediction horizon Δ given the historical traffic flow sequence of observation locations $\{X_\tau^i\}$ ($\tau = t - n\Delta, \dots, t - \Delta, t$ and $i \in O$, where O is the set of observation locations in the transportation network). In this work, we consider $\Delta = 5$ minutes, $n = 15$ and $h = 1, 3, 6, 12$, which means that 75-minute historical data is used to predict the traffic flow of the next 5, 15, 30 and 60 minutes. For the simplicity of description, we denote $t - n$ as $t - n\Delta$ by omitting the symbol Δ in the paper.

Since transportation traffic condition is usually random and nonlinear, a good traffic model should be able to capture such complicate characteristics. Many statistical or machine learning based traffic models have been developed to improve prediction performance. Feature learning is an important procedure for building an efficient traffic model, which extracts and selects the most representative features from the historical traffic flow data. The features of traffic flow usually exhibit spatial-temporal correlation and periodic characteristic. More concretely, the traffic flow of a POI is not only impacted by the traffic conditions of its neighboring observation locations but also influenced by previous time. Furthermore, the traffic flow also exhibits periodic pattern in a daily or weekly manner. For instance, the variation tendency of the traffic flow in the same day of two consecutive weeks is very similar. All these characteristics of traffic flow together determine the future status of the observation location to be predicted. In this paper, we propose a deep learning based model to exploit both spatial-temporal correlation and periodic characteristic for short-term traffic flow prediction.

Before introducing our traffic model, we describe how to construct historical traffic flow to facilitate extracting spatial-temporal and periodic features. Let f_t^p denote the traffic flow of an observation location p at time t. The historical traffic flow of the observation location p from time $t - n$ to t can be represented as $X_t^p = [f_{t-n}^p, f_{t-(n-1)}^p, \dots, f_t^p]^T$. Then we combine the historical traffic flow of its neighboring locations (total m locations including location p) to form a spatial-temporal traffic flow matrix as follows:

$$\boldsymbol{X}_t^s = \begin{bmatrix} X_{t-n}^s \\ X_{t-(n-1)}^s \\ \vdots \\ X_t^s \end{bmatrix}^T = \begin{bmatrix} f_{t-n}^1 & f_{t-(n-1)}^1 & \cdots & f_t^1 \\ f_{t-n}^2 & f_{t-(n-1)}^2 & \cdots & f_t^2 \\ \vdots & \vdots & \ddots & \vdots \\ f_{t-n}^m & f_{t-(n-1)}^m & \cdots & f_t^m \end{bmatrix}, \quad (1)$$

where $X_t^s = [f_t^1, f_t^2, \dots, f_t^m]$ denotes the traffic flow of the prediction region at time t.

In addition, we consider the periodic characteristic of the traffic flow. To construct historical traffic flow data with periodicity, we take both daily and weekly periodicity into account. The traffic data with daily periodicity can be obtained by considering previous and subsequent n time intervals of the same moment as time t in the last day, which can be represented as

$$\boldsymbol{X}_t^d = \begin{bmatrix} f_{t^d-n}^1 & f_{t^d-(n-1)}^1 & \cdots & f_{t^d}^1 & \cdots & f_{t^d+n}^1 \\ f_{t^d-n}^2 & f_{t^d-(n-1)}^2 & \cdots & f_{t^d}^2 & \cdots & f_{t^d+n}^2 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ f_{t^d-n}^m & f_{t^d-(n-1)}^m & \cdots & f_{t^d}^m & \cdots & f_{t^d+n}^m \end{bmatrix}, \quad (2)$$

where t^d denotes the same moment as time t in the last day. Similarly, we construct historical traffic flow data with weekly periodicity by considering previous and subsequent n time intervals of the same moment as time t in the last week as follows

$$\boldsymbol{X}_t^w = \begin{bmatrix} f_{t^w-n}^1 & f_{t^w-(n-1)}^1 & \cdots & f_{t^w}^1 & \cdots & f_{t^w+n}^1 \\ f_{t^w-n}^2 & f_{t^w-(n-1)}^2 & \cdots & f_{t^w}^2 & \cdots & f_{t^w+n}^2 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ f_{t^w-n}^m & f_{t^w-(n-1)}^m & \cdots & f_{t^w}^m & \cdots & f_{t^w+n}^m \end{bmatrix}, \quad (3)$$

where t^w denotes the same moment as time t in the last week.

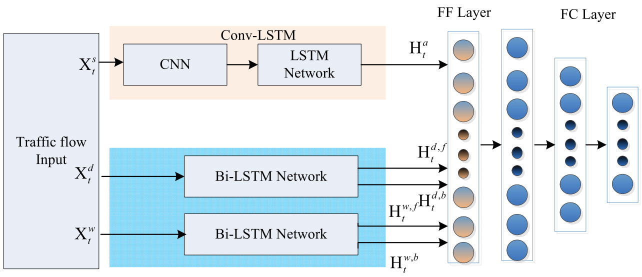


Fig. 1. The proposed deep learning based model for short-term traffic flow prediction.

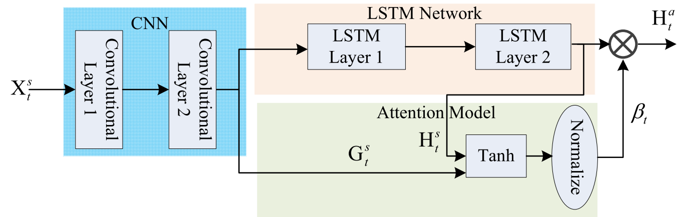


Fig. 2. The Conv-LSTM module with an attention mechanism.

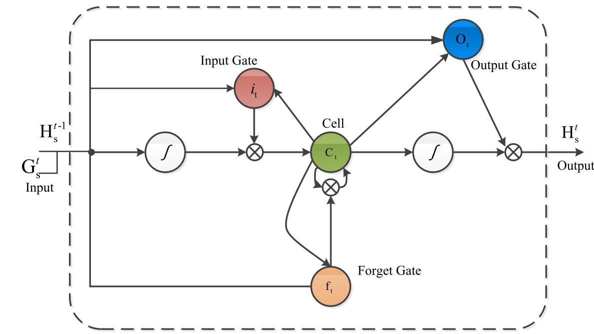


Fig. 3. The LSTM neuron structure.

IV. The Proposed Hybrid Deep Learning Model for Traffic Flow Prediction

A. Overview of the Proposed Model

In this paper, we propose a novel hybrid deep architecture for traffic flow prediction. The proposed model consists of a Conv-LSTM module and two Bi-LSTM modules. Fig. 1 illustrates the overall architecture of the proposed model. The Conv-LSTM module is constructed from the convolution neural network and the LSTM network, where the convolution neural network is utilized to extract the spatial feature of the traffic flow and then is connected to the LSTM network to obtain the short-term temporal feature of the traffic flow. Recently, some similar approaches that combine CNN and LSTM have been proposed for a variety of applications such as traffic information detection [38], network fault prediction [39], gesture recognition [46] and speech emotion recognition [47]. Meanwhile, the Bi-LSTM module is used to extract the daily and weekly periodic features of traffic flow. Afterward, the spatial-temporal feature and the periodicity feature are fused to a feature vector by a feature fusion layer (FF layer). Finally, the feature fusion layer is followed by the two fully-connected layers (FC layer) which are regression layers to perform forecasting. In addition, we design an attention mechanism for the Conv-LSTM module to automatically explore different levels of the importance of flow sequences at different times. We will describe each module in detail in the following subsections.

B. Conv-LSTM

The Conv-LSTM module is the main component of the proposed model that aims to extract the spatial-temporal feature of traffic flow. The Conv-LSTM module incorporates the convolutional neural network and the LSTM network as shown in Fig. 2. The convolutional neural network consists of two convolutional layers and the LSTM network contains two LSTM layers, respectively.

The input of Conv-LSTM is a spatial-temporal traffic flow matrix X_t^s as indicated in Eq. (1), which represents the historical traffic flow of the POI to be predicted and its neighbours. To extract the spatial feature, one dimensional convolution operation is performed over the flow data X_t^s at each time step t. A one-dimensional convolution kernel filter is used to acquire the local perceptual domain by a sliding filter. The process of convolution kernel filter can be expressed as follows:

$$Y_t^s = \sigma(W_s * X_t^s + b_s), \tag{4}$$

where W_s is the weights of the filter, b_s is bias, X_t^s is the input traffic flow at time t, symbol $*$ represents the convolution operation, σ is the activation function and Y_t^s is the output of the convolutional layer. Such a process is conducive to extracting the spatial feature from the neighbouring observation locations.

The pooling layer is not applied after the convolutional layer in our model since the dimension of the spatial feature is not large. Denote G_t^s as the output of the convolutional layer 2. After the spatial information is processed by the two convolutional layers, then the output is connected to an LSTM network.

As is known, traffic flow also exhibits temporal correlations in adjacent times. Recurrent neural network (RNN) is usually introduced for analyzing the hidden temporal feature in sequential data [41]. However, RNN performs poor in case of long term sequences due to the vanishing gradient problem since it forgets the earlier status of the sequence. LSTM provides a practical method to learn long time dependencies for long term sequential data. A structure of LSTM contains a cell to store block cell state and three gates: input gate, out gate and forget gate. The current state is influenced by the forget gate and input gate. LSTM uses the forget gate to determine how much information at the previous cell state is retained in the current cell state, and uses the input gate to decide how much information of the input needs to be saved in the current cell state. The structure of each LSTM neuron is shown in Fig. 3. In this paper, we propose to utilize LSTM to extract temporal feature of traffic flow.

To improve the performance of the deep neural network, the traditional method is to increase the number of layers in the model. In this paper, we stack multiple LSTM layers to the model to capture higher level of the features of the traffic flow. By stacking LSTM layers, each LSTM layer receives the hidden state of the previous layer as its input. For example, as shown in Fig. 2, LSTM layer 1 processes the sequence output from the CNN module $\mathbf{G}_t^s = [G_{t-n}^s, \ldots, G_{t-1}^s, G_t^s]$ from begin to end and calculates the hidden state for each time step $\mathbf{H}_{1,t}^s = [H_{1,t-n}^s, \ldots, H_{1,t-1}^s, H_{1,t}^s]$. Then the hidden state sequence $\mathbf{H}_{1,t}^s$ is input into LSTM layer 2 to calculate the hidden state $H_{2,t}^s$ for time step t as the output of the entire LSTM network H_t^s. The computation of each LSTM layer can be explained through Eq. (5) to Eq. (9) as follows:

$$i_t = \sigma(W_{gi} I_t^s + W_{hi} H_{t-1}^s + W_{ci} \circ C_{t-1} + b_i), \tag{5}$$

$$f_t = \sigma(W_{gf} I_t^s + W_{hf} H_{t-1}^s + W_{cf} \circ C_{t-1} + b_f), \tag{6}$$

$$C_t = f_t \circ C_{t-1} + i_t \circ \tanh(W_{gc} I_t^s + W_{hc} H_{t-1}^s + b_c), \tag{7}$$

$$o_t = sigma(W_{go} I_t^s + W_{ho} H_{t-1}^s + W_{co} \circ C_t + b_o), \tag{8}$$

$$H_t^s = o_t \circ \tanh(C_t), \tag{9}$$

where I_t^s is the input of the LSTM layer at time step t, σ is the activation function, i_t, f_t, o_t are respectively the input gate, the forget Gate and output gate at time t, C_t is the cell state, $\circ$ represents the matrix element-wise product, W's and b's are weights and bias, respectively. The final output of the LSTM layer H_t^s is determined by the output gate and the update cell as described in Eq. (9). The inputs I_t^s of layer 1 and layer 2 are G_t^s and $H_{1,t}^s$, respectively. The outputs H_t^s of layer 1 and layer 2 correspond to $H_{1,t}^s$ and $H_{2,t}^s$, respectively. Finally, we obtain the spatial-temporal feature H_t^s for time step t.

C. Attention Mechanism

Attention model has been proposed to explore the inherent features of data and improve the efficiency of information processing. For example, an attention mechanism was introduced in neural machine translation by assigning different weights to text fragments and make information more efficiently encoded [40]. It lays the groundwork for variants of subsequent attention mechanisms. As is known, the information provided by the traffic flow at different times may be not equally important for prediction performance. In other words, the traffic condition of some observation locations at different times may have a different influence on the traffic flow of the POI to be predicted. However, the standard LSTM cannot detect which is the important part for a traffic flow sequence. To solve this problem, we design an attention mechanism for the Conv-LSTM module to automatically exploit different levels of importance of a traffic flow sequence at different times.

Fig.4 shows an illustration on how the attention mechanism is incorporated in the Conv-LSTM module. As illustrated in Fig. 4, the output of Conv-LSTM at each time step t is computed as a weighted summation of the output of the LSTM network H_t^s as follows:

$$H_t^a = \sum_{k=1}^{n+1} \beta_k H_{t-(k-1)}^s, \tag{10}$$

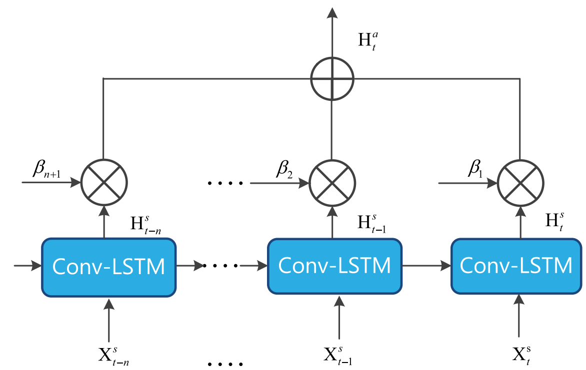


Fig. 4. The attention mechanism with Conv-LSTM networks.

where $n + 1$ is the length of flow sequence and β_k is the temporal attention value at time step $t - (k - 1)$. The attention value β_k can be computed as

$$\beta_k = \frac{\exp(s_k)}{\sum_{k=1}^{n+1} \exp(s_k)}. \tag{11}$$

The scores $s = (s_1, s_2, \cdots, s_{n+1})^T$ indicate the importance of each part in the traffic flow sequence, which can be obtained as

$$s_t = V_s^T \tanh(W_{hs} G_t^s + W_{ls} H_t^s), \tag{12}$$

where V_s, W_{xs} and W_{hs} are the learnable parameters and H_t^s is the hidden output from the Conv-LSTM network.

From Eqs. (11) and (12), we can see that the attention value β at time step t depends on the input G_t^s and the hidden variables H_t^s at the current time step t and its previous n time steps. The attention value β can be also viewed as the activation of the flow selection gate. The set of gates control the amount of information from each flow to enter the LSTM network. The larger the activation value, the more important the flow contributes to the final prediction result. Note that the proposed attention mechanism does not require other auxiliary information such as road speed [16] to learn the attention weights.

D. Bi-Directional LSTM

As discussed in the previous section, the behaviors of citizens usually show a certain regularity in their daily life, which results in the similar or repeated pattern of the traffic flow in a daily or weekly manner. In addition, the traffic flow at a given time not only depends on the flow of its previous time, but also in turn affects the flow of its upcoming time. In order to explore these characteristics, we propose a module based on bi-directional LSTM networks to extract periodic features and capture such a temporal dependency from the historical traffic flow.

The structure of bi-directional LSTM contains two uni-directional LSTMs stacked up and down, where one is for forward pass and another is for backward pass. We also use multiple LSTM layers as discussed above to deal with the historical periodic traffic flow, where each bi-directional LSTM network consists of two LSTM layers for both forward

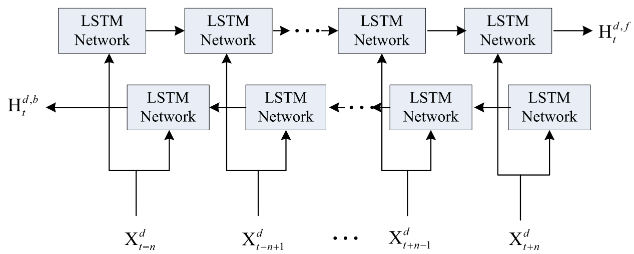


Fig. 5. The structure of Bi-LSTM networks.

pass and backward pass. The input data at a predicted time are the flow of its previous time and upcoming time of the last day and the last week, which are respectively denoted by Eqs. (2) and (3). Here only the historical periodic traffic flow of the prediction location is considered. Then they are fed to the forward and backward passes of the bi-directional LSTM, respectively. The hidden states of forward pass and backward pass are finally combined together as the output. In this way, more features from both directions can be captured, which improves the prediction performance. Fig. 5 illustrates the overall structure of the proposed bi-directional LSTM module used in our model, where $\boldsymbol{X}_t^d$ and $\boldsymbol{X}_t^w$ denote the input of LSTM, $\boldsymbol{H}_t^{d,f}$ and $\boldsymbol{H}_t^{w,f}$ denote the output of forward LSTM, and $\boldsymbol{H}_t^{d,b}$ and $\boldsymbol{H}_t^{w,b}$ denote the output of backward LSTM when the input of LSTM are $\boldsymbol{X}_t^d$ and $\boldsymbol{X}_t^w$, respectively.

E. Feature Fusion

As shown in Fig.1, after the processing by the attention Conv-LSTM and Bi-LSTM modules, we obtain the spatial-temporal features H_t^a, the daily periodicity features $\boldsymbol{H}_t^{d,f}$, $\boldsymbol{H}_t^{d,b}$ and the weekly periodicity features $\boldsymbol{H}_t^{w,f}$ and $\boldsymbol{H}_t^{w,b}$. Then all these features are concatenated into a feature vector and then is input by two regression layers to perform forecasting. The objective function of regression is a loss function to calculate the mean squared error of the predicted traffic flow and the true traffic flow, which will be discussed later.

V. Performance Evaluation

A. Datasets

In this section, we evaluate the performance of the proposed model by using real-world dataset for short-term traffic flow prediction. The dataset is obtained from the Performance Measurement System (PeMS) sponsored by the California Department of Transportation (Caltrans), which is widely used for traffic prediction [42]. The dataset used for the experiments is collected from September 11th, 2017 to March 4th, 2018 for about 6 months, which includes the data of both weekdays and weekends. The traffic data are aggregated every 5 minutes. The proposed model is trained and evaluated by the dataset collected from sensors that are respectively located in the freeway and urban area, representing two different scenarios. One area is located at Freeway SR99-S District 10 and the other one is located at Street I980 District 4 in Oakland city.

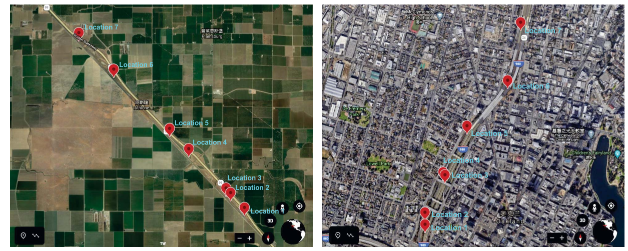


Fig. 6. Sensor distribution on Freeway SR99-S (left) and Street I980 (right).

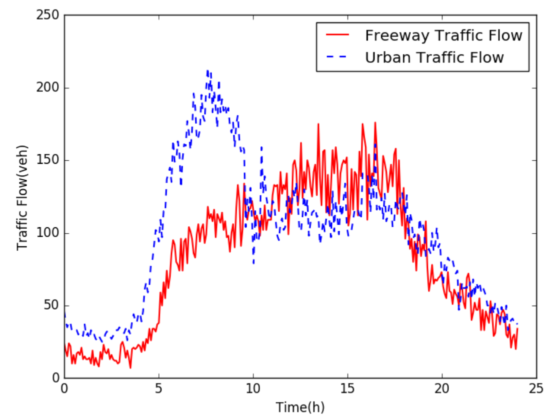


Fig. 7. Comparison of the freeway and urban traffic flow.

Fig. 6 shows the distribution of the sensors used for prediction at Freeway SR99-S District 10 and Street I980 District 4 in Oakland city, respectively. Fig. 7 shows the corresponding traffic flow of two separated sensors for 24 hours in one day, which are deployed in the freeway (Location 2) and urban (Location 1) area, respectively. The data is obtained by averaging the traffic volume of ten days at each point. From Fig. 7, we observe that the traffic flow of freeway and urban area exhibits similar tendency. For instance, the traffic flow in the early morning and evening is the lowest, and then increases slowly with time and reaches its peak around the afternoon. Fig. 8 plots the traffic flow of Location 1 at Street I980 District 4 from September 18th, 2017 to September 22nd, 2017. From Fig. 8, we can see that traffic flow at the same location also shows strong periodic characteristic.

B. Index of Performance

In order to evaluate the performance of the proposed model, we use three performance indexes that are commonly used to evaluate the traffic forecasting performance: Mean Absolute Error (MAE), Mean Absolute Percentage Error (MAPE), Root Mean Square Error (RMSE), which can be defined as follows:

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

$$MAPE = \frac{1}{n}\sum_{t=1}^{n}|\frac{F_t - F_p}{F_t}| \times 100\%, \tag{14}$$

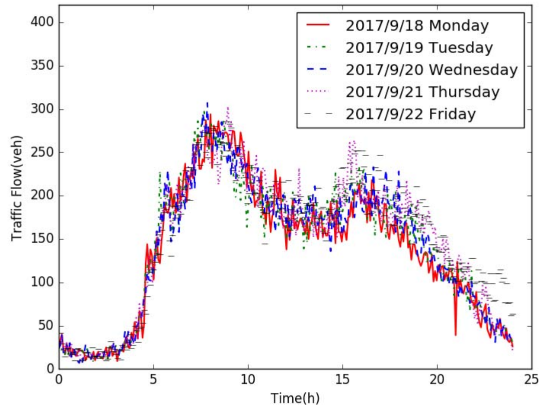


Fig. 8. The traffic flow of Location 1 on Street I980 District 4 in 2017/9/18-2017/9/22.

$$RMSE = \sqrt{\frac{1}{n}\sum_{t=1}^{n}(F_p - F_t)^2}, \tag{15}$$

where F_p represents the predicted traffic flow and F_t represents the true traffic flow.

C. Training Configuration

In the following section, we introduce some details on the training of the proposed model. During the training phase, we set a loss function to update the parameters in our model, which contains a Mean Squared Error (MSE), ℓ_1 weight regularization and ℓ_2 weight regularization. The loss function is defined as follows:

$$Loss = MSE + \frac{\lambda_1}{n}\|\boldsymbol{w}\|_1 + \frac{\lambda_2}{n}\|\boldsymbol{w}\|_2, \tag{16}$$

where λ_1, λ_2 are regularization parameters and $\boldsymbol{w}$ is the weight. In the loss function, MSE is defined as the mean squared error of the predicted traffic flow and the true traffic flow:

$$MSE = \frac{1}{n}\sum_{t=1}^{n}(F_p - F_t)^2, \tag{17}$$

where F_p is the predicted traffic flow, F_t is the true traffic flow, and n is the size of dataset.

The goal of ℓ_1 regularization included in the loss function is to obtain a sparse model, which can prevent overfitting by the deep model. Furthermore, ℓ_2 regularization in the loss function can prevent the occurrence of numerical oversize parameters in the model and avoid a particular feature to dominate the prediction performance of the model. ℓ_1 regularization and ℓ_2 regularization can be respectively defined as follows:

$$\|w\|_1 = \sum_{i=1}^{n}|W_i| \quad \text{and} \quad \|w\|_2 = \sum_{i=1}^{n}\sqrt{W_i^2}. \tag{18}$$

Then, the loss function can be rewritten as follows

$$Loss = \frac{1}{n}\left(\sum_{t=1}^{n}(F_p - F_t)^2 + \lambda_1\sum_{i=1}^{n}|W_i| + \lambda_2\sum_{i=1}^{n}\sqrt{W_i^2}\right) \tag{19}$$

TABLE I

PREDICTION PERFORMANCE WITH VARIOUS PROPOSED MODULES FOR URBAN TRAFFIC FLOW PREDICTION

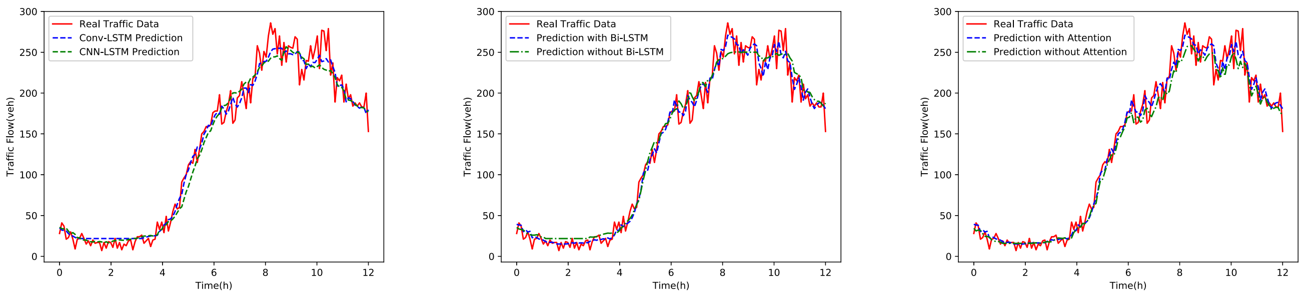
Algorithm	Indexes	5 min	15 min	30 min	60 min
CNN-LSTM	MAE	12.25	13.32	14.98	17.55
	MAPE	12.8%	14.1%	14.8%	18.3%
	RMSE	16.87	18.31	20.37	23.03
Conv-LSTM	MAE	12.02	13.19	14.35	15.79
	MAPE	12.5%	13.5%	14.5%	17.8%
	RMSE	16.41	18.03	19.63	21.39
With Bi-LSTM	MAE	11.57	12.526	14.02	15.03
	MAPE	12.0%	13.0%	14.0%	16.0%
	RMSE	15.78	17.61	19.08	20.21
With Attention	MAE	11.21	12.13	13.57	14.27
	MAPE	11.8%	12.6%	13.6%	14.7%
	RMSE	15.56	17.21	18.56	19.56

In the proposed model, we use the Adam optimization algorithm to optimize the model parameters, which can adaptively adjust the learning rate.

D. Prediction Performance of the Proposed Model

In this section, we employ the proposed model with various modules including Conv-LSTM, Bi-LSTM, attention mechanism to predict the traffic flow of one POI (Location 3) at Street I980 District 4 in Oakland city. The proposed model is implemented by using Tensorflow framework [45]. The experiments are conducted on a workstation with an Intel Core i7-8700 CPU and one Nvidia GeForce RTX 2080Ti Graphics Card. In the experiment, the convolutional layer has 10 filters with the size of each filter being 3. The stride of the sliding window for the input flow data is set to 1. The batch size for training data is set to 128. The Rectified linear activation unit (ReLU) is adopted as the activation function. In all the experiments, the size of the time window is set to 15, which means that 75-minute historical data is used as a training unit.

We first evaluate the ability of extracting spatio-temporal feature with the proposed Conv-LSTM module for traffic flow prediction. As discussed above, the spatial traffic data and the temporal traffic data are input together into the Conv-LSTM module to extract spatial-temporal feature. To verify the effectiveness of the proposed approach, we compare the prediction performance of Conv-LSTM module with a CNN-LSTM module (without including other modules). With the CNN-LSTM module, the spatial traffic data is individually input into a convolutional neural network to extract spatial feature while the temporal traffic data is individually input into an LSTM network to extract temporal feature. As a result, the spatial feature and the temporal feature are separated and fused in a fully connected layer. Fig. 9(a) visualizes an illustration of performance comparison in terms of predicted traffic volume from 0:00 to 12:00 AM for 5-minute prediction horizon. It can be seen that the traffic volume predicted by the Conv-LSTM module is more approximate to the ground truth than the CNN-LSTM module especially in those periods when there is a large fluctuation in traffic volume. Tab. I also shows the prediction performance in terms of MAE, MAPE and RMSE. All the results are obtained by averaging on each day for a week. We observe that the proposed module can achieve smaller prediction error than the CNN-LSTM module with



(a) Conv-LSTM vs. CNN-LSTM (b) w/o Bi-LSTM (c) w/o Attention

Fig. 9. Performance comparison with various proposed modules for urban traffic flow prediction.

all three metrics for all prediction horizons. This is because the spatial and temporal features of traffic flow are usually interweaved with each other, which can be captured more efficiently by the Conv-LSTM module. However, the spatial feature and the temporal feature are separately processed by two independent neural networks in the CNN-LSTM module.

Next, we evaluate prediction performance of the proposed Bi-LSTM module. As aforementioned, the traffic flow always shows strong periodic tendency. The advantage of the proposed Bi-LSTM module is that it is able to capture not only the temporal dependency but also the daily and weekly periodicities of the historical traffic flow. Furthermore, we incorporate the traffic flow of the subsequent times of the prediction time (to mimic future flow data) to capture the future variance tendency. As a result, more periodic features can be exploited by forward and backward passes of the Bi-LSTM module to improve prediction performance. Fig. 9(b) shows the prediction performance with and without the Bi-LSTM module. It can be obviously observed that the Bi-LSTM module provides a significant performance improvement, which obtains more accurate prediction flow at most of predicted times than that without Bi-LSTM. This also illustrates that periodic features are very conducive to improving prediction performance for traffic flow forecasting.

Furthermore, we evaluate prediction performance of the attention mechanism for the proposed model. As discussed in the previous section, the Conv-LSTM module with attention mechanism is able to automatically identify which are the important parts for traffic flow forecasting. Fig. 9(c) illustrates performance comparison of the proposed model with and without attention mechanism. It is evident that when the traffic flow has relatively large fluctuation the attention mechanism can provide better prediction performance than that without attention mechanism. Taking Fig. 9(c) as an example, the predicted traffic volume is obviously approximate to the ground truth from 8:00 to 12:00 AM. Tab. I also shows the prediction performance in terms of MAE, MAPE and RMSE. It shows that prediction error can be further reduced when the attention mechanism is incorporated in our model.

E. Performance Comparison With Different Prediction Algorithms in Urban Scenario

In this section, we carry out experiments to investigate the prediction performance by the proposed model compared to the existing methods for short-term traffic flow prediction including Support Vector Regression (SVR) [37], [43], Stack Auto Encoders (SAE) [12], LSTM networks [13], [44], DNN based traffic flow prediction model (DNN-BTF) [16], and Diffusion Convolutional Recurrent Neural Network (DCRNN) [20]. For SVR, we adopt the radial basis function (RBF) as the kernel function, and the penalty parameter is set to empirical value 1. For SAE, we use greedy layerwise unsupervised learning algorithm to train the deep SAE network. For the LSTM network, it contains two LSTM layers. For the DNN-BTF model, a convolution neural network is built to capture spatial feature of the near term, last day and last week of traffic flow, and two LSTM networks to process the same data to capture temporal and periodic features. The speed data of the corresponding traffic flow is utilized to learn the weights required by the attention mechanism. For the DCRNN model, the default settings of hyperparameters as in [20] are adopted for training. We use the traffic flow data of 7 sensors on Street I980 District 4 in Oakland city and then calculate the pairwise distances between sensors on the target road to build the adjacency matrix and construct the graph network. All the results are averaged on each location and each day for a week.

The prediction error in terms of performance indexes of different algorithms is shown in Tab. II. The experimental results demonstrate that SVR has largest prediction error among the above prediction algorithms with all three metrics for all the prediction horizons. This is because SVR utilizes the kernel function to map a large amount of uncertain traffic flow data into a high dimensional space, which cannot fully exploit spatial and periodic features for traffic forecasting, thus leading to poor prediction performance. Compared to SVR, LSTM achieves a better prediction performance, where MAPE, MAE and RMSE have a roughly average 20.4%, 5.3%, 1.6% reduction, respectively. LSTM improves its prediction performance due to its unique neuron unit and neural network structure for extracting temporal feature of traffic flow. However, LSTM is still unable to capture the other inherent features such as spatial and periodic features. Compared to the above two algorithms, SAE further reduces the prediction error in terms of MAPE, MAE and RMSE. This performance improvement may be due to the reason that SAE is able to extract more nonlinear features from the flow data. DNN-BTF has the ability to extract not only spatial-temporal feature but also periodic feature

TABLE II

PERFORMANCE COMPARISON OF DIFFERENT ALGORITHMS FOR URBAN TRAFFIC FLOW PREDICTION

T	Indexes	SVR	LSTM	SAE	DNN-BTF	DCRNN	**AT-Conv-LSTM**
5 min	MAE	16.39	14.77	14.32	14.05	13.79	**13.49**
	MAPE	16.7%	11.1%	11.1%	10.9%	10.7%	**10.1%**
	RMSE	20.82	20.05	19.61	19.32	18.88	**18.56**
15 min	MAE	18.03	16.50	16.25	15.55	14.79	**14.34**
	MAPE	18.0%	12.4%	12.1%	11.5%	11.5%	**10.8%**
	RMSE	23.14	22.57	22.12	21.37	20.43	**20.08**
30 min	MAE	20.03	19.00	17.86	16.97	16.05	**15.48**
	MAPE	19.3%	14.7%	13.4%	12.8%	12.4%	**11.4%**
	RMSE	25.92	25.59	24.54	23.06	22.18	**21.26**
60 min	MAE	24.00	23.60	21.31	19.12	18.43	**16.65**
	MAPE	21.9%	18.3%	16.2%	14.8%	14.2%	**12.3%**
	RMSE	30.90	30.75	28.57	25.88	25.74	**23.26**

TABLE III

PERFORMANCE COMPARISON OF DIFFERENT ALGORITHMS FOR FREEWAY TRAFFIC FLOW PREDICTION

T	Indexes	SVR	LSTM	SAE	DNN-BTF	DCRNN	**AT-Conv-LSTM**
5 min	MAE	11.66	10.91	10.67	8.47	7.34	**7.14**
	MAPE	22.1%	16.7%	16.1%	12.4%	11.0%	**10.8%**
	RMSE	15.03	14.67	14.44	11.26	9.88	**9.69**
15 min	MAE	12.51	11.69	11.42	11.25	10.90	**10.66**
	MAPE	23.9%	17.8%	17.3%	16.7%	16.0%	**15.7%**
	RMSE	16.39	16.04	15.73	15.44	14.99	**14.75**
30 min	MAE	13.62	12.96	12.46	11.98	11.91	**11.23**
	MAPE	25.3%	20.6%	18.9%	18.2%	17.3%	**16.3%**
	RMSE	18.09	17.98	17.42	16.60	16.51	**15.87**
60 min	MAE	15.84	15.31	14.25	13.56	13.46	**12.03**
	MAPE	28.7%	24.9%	22.2%	21.0%	19.9%	**17.5%**
	RMSE	21.01	20.76	19.65	18.56	18.93	**16.89**

since it is built on a hybrid deep neural network including a convolutional neural network and two LSTM networks. The model uses the road speed to learn the weights that represent the correlation between the past spatial-temporal flow and the future one. Consequently, DNN-BTF outperforms the above three algorithms. As discussed before, DCRNN models the dynamics of the traffic flow as a diffusion process and uses the diffusion graph convolution to capture spatial dependency and recurrent neural network to capture the temporal dynamics. Consequently, DCRNN achieves the best performance among the above four algorithms.

Different from all the above algorithms, the proposed model, called as attention Conv-LSTM (AT-Conv-LSTM), can extract spatial-temporal feature and periodic feature more efficiently than DNN-BTF and DCRNN. More specially, spatial and temporal flow data are processed together by the Conv-LSTM model to extract spatial-temporal feature and Bi-LSTM can capture periodic features in both previous and posterior directions. Furthermore, the proposed attention mechanism can also automatically assign different degrees of importance to each flow sequence at different times to improve prediction performance. From Tab. II, it can be seen that AT-Conv-LSTM outperforms all the above algorithms with all three metrics for all prediction horizons. In particular, performance improvement becomes more evident with the increase of prediction horizon.

F. Performance Comparison Under Different Scenarios

To illustrate the generalization of the proposed model, we investigate the prediction performance under different scenarios. We select the traffic flow data at Freeway SR99-S District 10 for comparison. The sensor distribution of the freeway area to be predicted is illustrated in Fig. 6 (left). Tab. III shows the average prediction error on each location with different algorithms. We notice that the average prediction error of the freeway scenario is much smaller than that of the urban scenario. This may be due to the reason that the traffic conditions in the urban scenario are more complicated than that in the freeway scenario and the proposed model performs better in the freeway scenario. It can also be observed that the proposed model still outperforms the other algorithms with all three metrics for all prediction horizons especially for longer prediction horizons (e.g., 60 minutes). Fig. 10 visualizes an example of performance comparison for 5-minute ahead prediction with the three hybrid models. We observe that although these three models all perform well at most of time in this scenario, AT-Conv-LSTM approximates much more closely to the ground truth at most of moments of acute

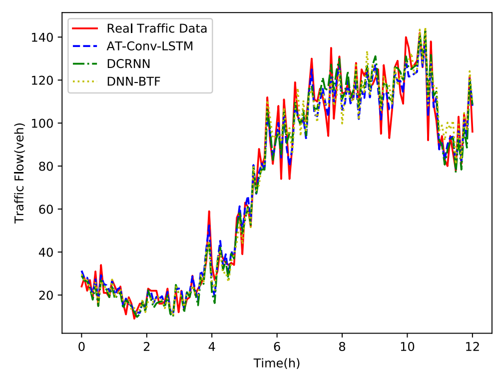


Fig. 10. Prediction comparison of different algorithms for freeway traffic flow.

traffic fluctuation and DNN-BTF performs worst especially at around 11:00 AM. In general, the proposed model provides good universality for traffic flow prediction under different scenarios.

VI. Conclusion and Future Work

In this paper, we studied the short-term traffic flow forecasting for ITSs. To deal with the complex nonlinear characteristic of traffic flow, we proposed a hybrid deep learning model constructed from the convolutional neural network and the LSTM network. We find that spatial-temporal features can be extracted more efficiently when they are processed together in the proposed Conv-LSTM module. Furthermore, it has been shown that the proposed attention mechanism is also benefit for the Conv-LSTM module which is able to enhance prediction performance. Meanwhile, the proposed Bi-LSTM module can also efficiently capture the daily and weekly periodic features to improve prediction accuracy. The extensive experimental results demonstrate that the proposed model achieves superior forecasting performance compared with the existing approaches.

In this work, the road network we consider is relatively simple and small, where the sensors are usually deployed along a straight line. However, the road networks in practice are more complex and large-scale than the one given in this work. As such, the conventional CNN and (Bi-)LSTM networks may not be able to fully exploit the complex and dynamic features of traffic flow. A class of attention-based graph neural networks such as GMAN [48] provides a potential alternative to the conventional CNN-LSTM hybrid networks for traffic prediction in complex and large-scale road networks. Thus, it will be left as our future work.

References

[1] J. Zhang, F.-Y. Wang, K. Wang, W.-H. Lin, X. Xu, and C. Chen, "Data-driven intelligent transportation systems: A survey," *IEEE Trans. Intell. Transp. Syst.*, vol. 12, no. 4, pp. 1624–1639, Dec. 2011.

[2] A. Koesdwiady, R. Soua, and F. Karray, "Improving traffic flow prediction with weather information in connected cars: A deep learning approach," *IEEE Trans. Veh. Technol.*, vol. 65, no. 12, pp. 9508–9517, Dec. 2016.

[3] J. Guo and B. Williams, "Real-time short-term traffic speed level forecasting and uncertainty quantification using layered Kalman filters," *Transp. Res. Rec., J. Transp. Res. Board*, vol. 2175, pp. 28–37, 2010.

[4] N. Barimani, A. Kian, and B. Moshiri, "Real time adaptive non-linear estimator/predictor design for traffic systems with inadequate detectors," *IET Intell. Transp. Syst.*, vol. 8, no. 3, pp. 308–321, 2014.

[5] M. Van Der Voort, M. Dougherty, and S. Watson, "Combining Kohonen maps with ARIMA time series models to forecast traffic flow," *Transp. Res. C, Emerg. Technol.*, vol. 4, no. 5, pp. 307–318, 1996.

[6] S. Lee and D. Fambro, "Application of subset autoregressive integrated moving average model for short-term freeway traffic vol. forecasting," *Transp. Res. Rec.*, vol. 1678, pp. 179–188, 1999.

[7] B. M. Williams and L. A. Hoel, "Modeling and forecasting vehicular traffic flow as a seasonal ARIMA process: Theoretical basis and empirical results," *J. Transp. Eng.*, vol. 129, no. 6, pp. 664–672, Nov. 2003.

[8] P. Duan, G. Mao, W. Liang, and D. Zhang, "A unified spatio-temporal model for short-term traffic flow prediction," *IEEE Trans. Intel. Transp. Syst.*, vol. 20, no. 9, pp. 3212–3223, Sep. 2019.

[9] W.-C. Hong, Y. Dong, F. Zheng, and S. Y. Wei, "Hybrid evolutionary algorithms in a SVR traffic flow forecasting model," *Appl. Math. Comput.*, vol. 217, no. 15, pp. 6733–6747, 2011.

[10] X. Feng, X. Ling, H. Zheng, Z. Chen, and Y. Xu, "Adaptive multi-kernel SVM with spatial-temporal correlation for short-term traffic flow prediction," *IEEE Trans. Intel. Transp. Syst.*, vol. 20, no. 6, pp. 2001–2013, Jun. 2019.

[11] W. Huang, G. Song, H. Hong, and K. Xie, "Deep architecture for traffic flow prediction: Deep belief networks with multitask learning," *IEEE Trans. Intell. Transp. Syst.*, vol. 15, no. 5, pp. 2191–2201, Oct. 2014.

[12] Y. Lv, Y. Duan, W. Kang, Z. Li, and F. Wang, "Traffic flow prediction with big data: A deep learning approach," *IEEE Trans. Intel. Transp. Syst.*, vol. 16, no. 2, pp. 865–873, Apr. 2015.

[13] Z. Zhao *et al.*, "LSTM network: A deep learning approach for short-term traffic forecast," *IET Intell. Transp. Syst.*, vol. 11, no. 2, pp. 68–75, Mar. 2017.

[14] X. Ma *et al.*, "Learning traffic as images: A deep convolutional neural network for large-scale transportation network speed prediction," *Sensors*, vol. 17, no. 4, 2017.

[15] B. Du *et al.*, "Deep irregular convolutional residual LSTM for urban traffic passenger flows prediction," *IEEE Trans. Intel. Transp. Syst.*, vol. 21, no. 3, pp. 972–985, Mar. 2020.

[16] Y. Wu, H. Tan, L. Qin, B. Ran, and Z. Jiang, "A hybrid deep learning based traffic flow prediction method and its understanding," *Transp. Res. C, Emerg. Technol.*, vol. 90, pp. 166–180, May 2018.

[17] A. Zonoozi, J. Kim, X. Li, and G. Cong, "Periodic-CRN: A convolutional recurrent model for crowd density prediction with recurring periodic patterns," in *Proc. Int. Joint Conf. on Artif. Intell. (IJCAI)*, 2018, pp. 3732–3738.

[18] J. Bruna, W. Zaremba, A. Szlam, and Y. LeCun, "Spectral networks and locally connected networks on graphs," in *Proc. Int. Conf. Learn. Represent. (ICLR)*, 2014, pp. 1–8.

[19] B. Yu, H. Yin, and Z. Zhu, "Spatio-temporal graph convolutional networks: A deep learning framework for traffic forecasting," in *Proc. Int. Joint Conf. Artif. Intell. (IJCAI)*, 2018, pp. 1–7.

[20] Y. Li, R. Yu, C. Shahabi, and Y. Liu, "Diffusion convolutional recurrent neural network: Data-driven traffic forecasting," in *Proc. Int. Conf. Learn. Represent. (ICLR)*, 2018, pp. 1–16.

[21] Y. Liu, H. Zheng, X. Feng, and Z. Chen, "Short-term traffic flow prediction with Conv-LSTM," in *Proc. Int. Conf. Wireless Commun. Signal Process.*, Oct. 2017, pp. 1–16.

[22] M. Levin and Y.-D. Tsao, "On forecasting freeway occupancies and volumes," *Transp. Res. Rec.*, vol. 773, pp. 47–49, Oct. 1980.

[23] M. M. Hamed, H. R. Al-Masaeid, and Z. M. B. Said, "Short-term prediction of traffic vol. in, urban Arterials," *J. Transp. Eng.*, vol. 121, no. 3, pp. 249–254, 1995.

[24] B. Ghosh, B. Basu, and M. O'Mahony, "Bayesian time-series model for short-term traffic flow forecasting," *J. Transp. Eng.*, vol. 133, no. 3, pp. 180–189, 2007.

[25] G. A. Davis and N. L. Nihan, "Nonparametric regression and short-term freeway traffic forecasting," *J. Transp. Eng.*, vol. 117, no. 2, pp. 178–188, Mar. 1991.

[26] H. Chang, Y. Lee, B. Yoon, and S. Baek, "Dynamic near-term traffic flow prediction: Systemoriented approach based on past experiences," *IET Intell. Transp. Syst.*, vol. 6, no. 3, pp. 292–305, 2012.

[27] Y.-S. Jeong, Y.-J. Byon, M. Mendonca Castro-Neto, and S. M. Easa, "Supervised weighting-online learning algorithm for short-term traffic flow prediction," *IEEE Trans. Intell. Transp. Syst.*, vol. 14, no. 4, pp. 1700–1707, Dec. 2013.

[28] E. I. Vlahogianni, M. G. Karlaftis, and J. C. Golias, "Optimized and metaoptimized neural networks for short-term traffic flow prediction: A genetic approach," *Transp. Res. C, Emerg. Technol.*, vol. 13, no. 3, pp. 211–234, Jun. 2005.

[29] K. Y. Chan, T. S. Dillon, J. Singh, and E. Chang, "Neural-network-based models for short-term traffic flow forecasting using a hybrid exponential smoothing and Levenberg-Marquardt algorithm," *IEEE Trans. Intel. Transp. Syst.*, vol. 13, no. 2, pp. 644–654, Jun. 2012.

[30] W. Zheng, D.-H. Lee, and Q. Shi, "Short-term freeway traffic flow prediction: Bayesian combined neural network approach," *J. Transp. Eng.*, vol. 132, pp. 114–121, Sep. 2006.

[31] M. Zhong, S. Sharma, and P. Lingras, "Short-term traffic prediction on different types of roads with genetically designed regression and time delay neural network models," *J. Comput. Civil Eng.*, vol. 19, no. 1, pp. 94–103, 2005.

[32] H. Yin, S. C. Wong, J. Xu, and C. K. Wong, "Urban traffic flow prediction using a fuzzy-neural approach," *Transp. Res. C, Emerg. Technol.*, vol. 10, no. 2, pp. 85–98, 2002.

[33] K. Kumar, M. Parida, and V. Katiyar, "Short term traffic flow prediction for a non urban highway using artificial neural network," *Proc. Social Behav. Sci.*, vol. 104, pp. 755–764, Dec. 2013.

[34] M.-C. Tan, S. C. Wong, J.-M. Xu, Z.-R. Guan, and P. Zhang, "An aggregation approach to short-term traffic flow prediction," *IEEE Trans. Intell. Transp. Syst.*, vol. 10, no. 1, pp. 60–69, Mar. 2009.

[35] M. Cetin and G. Comert, "Short-term traffic flow prediction with regime switching models," *Transp. Res. Rec.*, vol. 1965, pp. 23–31, Apr. 2006.

[36] L. Dimitriou, T. Tsekeris, and A. Stathopoulos, "Adaptive hybrid fuzzy rule-Based system approach for modeling and predicting urban traffic flow," *Transp. Res. C, Emerg. Technol.*, vol. 16, pp. 554–573, Oct. 2008.

[37] C. Wu, J. Ho, and D. T. Lee, "Travel time prediction with support vector regression," *IEEE Trans. Intell. Transp. Syst.*, vol. 5, no. 4, pp. 276–281, Dec. 2004.

[38] Y. Chen, Y. Lv, X. Wang, L. Li, and F. Wang, "Detecting traffic information from social media texts with deep learning approaches," *IEEE Trans. Intell. Transp. Syst.*, vol. 20, no. 8, pp. 3049–3058, Aug. 2019.

[39] Z. Tan and P. Pan, "Network fault prediction based on CNN-LSTM hybrid neural network," in *Proc. Int. Conf. Commun., Inf. Syst. Comput. Eng. (CISCE)*, 2019, pp. 486–490.

[40] D. Bahdanau, K. Cho, and Y. Bengio, "Neural machine translation by jointly learning to align and translate," in *Proc. Int. Conf. Learn. Represent. (ICLR)*, 2015, pp. 1–15.

[41] K.-I. Funahashi and Y. Nakamura, "Approximation of dynamical systems by continuous time recurrent neural networks," *Neural Netw.*, vol. 6, no. 6, pp. 801–806, 1993.

[42] *Caltrans PeMS*. Accessed: Oct. 15, 2016. [Online]. Available: http://pems.dot.ca.gov/

[43] D. Basak, S. Pal, and D. C. Patranabis, "Support vector regression," in *Proc. Neural Inf. Process. Lett. Rev.*, 2007, vol. 11, no. 10, pp. 203–224.

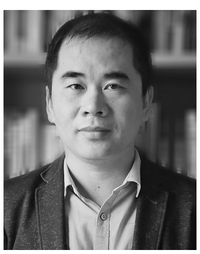
[44] Y. Tian and L. Pan, "Predicting short-term traffic flow by long short-term memory recurrent neural network," in *Proc. IEEE SmartCity*, Dec. 2015, pp. 153–158.

[45] M. Abadi *et al.*, "TensorFlow: Large-scale machine learning on heterogeneous distributed systems," 2016, *arXiv:1603.04467*. [Online]. Available: http://arxiv.org/abs/1603.04467

[46] L. Zhang, G. Zhu, L. Mei, P. Shen, S. Shah, and M. Bennamoun, "Attention in convolutional LSTM for gesture recognition," in *Proc. 32nd Conf. Neural Inf. Process. Syst. (NeurIPS)*, 2018, pp. 1–8.

[47] S. Zhang, X. Zhao, and Q. Tian, "Spontaneous speech emotion recognition using multiscale deep convolutional LSTM," *IEEE Trans. Affect. Comput.*, early access, Oct. 17, 2019, doi: 10.1109/TAFFC.2019.2947464.

[48] C. Zheng, X. Fan, C. Wang, and J. Qi, "GMAN: A graph multi-attention network for traffic prediction," in *Proc. AAAI Conf. Artif. Intell.*, 2020, pp. 1–8.



Haifeng Zheng (Member, IEEE) received the B.Eng. and M.S. degrees in communication engineering from Fuzhou University, Fuzhou, China, and the Ph.D. degree in communication and information systems from Shanghai Jiao Tong University, Shanghai, China. He is currently a Professor with the College of Physics and Information Engineering, Fuzhou University. He was a Visiting Scholar with the State University of New York at Buffalo from October 2015 to September 2016. His research interests include traffic modeling, machine learning, and intelligent transportation systems.

Feng Lin received the B.Eng. degree in electronic and information engineering from Jimei University, Xiamen, China, in 2017. He is currently pursuing the M.S. degree in electronic and communication engineering, Fuzhou University, Fuzhou, China. His research interests include intelligent transportation systems, deep learning, and the Internet of Things.

Xinxin Feng (Member, IEEE) received the B.S. and M.S. degrees in communication engineering from the Nanjing University of Science and Technology, China, in 2006 and 2008, respectively, and the Ph.D. degree in information and communication engineering from Shanghai Jiao Tong University, China, in 2015. She is currently an Associate Professor with the College of Physics and Information Engineering, Fuzhou University, Fuzhou, China. Her research interests include dynamic spectrum sharing, incentive mechanisms design in vehicle networks, and network economies.

Youjia Chen (Member, IEEE) received the B.S. and M.S. degrees in communication engineering from Nanjing University, Nanjing, China, in 2005 and 2008, respectively, and the Ph.D. degree in wireless engineering from The University of Sydney, Australia, in 2017. From 2008 to 2009, she worked at Alcatel-Lucent Shanghai Bell Company Ltd. She has been with the College of Photonic and Electronic Engineering, Fujian Normal University, China, since August 2009. In 2018, she joined the College of Physics and Information Engineering, Fuzhou University, China. Her current research interests include network traffic flow modeling, wireless caching, and machine learning.