2019 20th IEEE International Conference on Mobile Data Management (MDM)

STCNN: A Spatio-Temporal Convolutional Neural Network for Long-Term Traffic Prediction

Zhixiang He
Department of Computer Science
City University of Hong Kong
Hong Kong
zhixiang.he@my.cityu.edu.hk

Chi-Yin Chow
Department of Computer Science
City University of Hong Kong
Hong Kong
chiychow@cityu.edu.hk

Jia-Dong Zhang
Department of Computer Science
City University of Hong Kong
Hong Kong
jzhang26@cityu.edu.hk

***Abstract*—As many location-based applications provide services for users based on traffic conditions, an accurate traffic prediction model is very significant, particularly for long-term traffic predictions (e.g., one week in advance). As far, long-term traffic predictions are still very challenging due to the dynamic nature of traffic. In this paper, we propose a model, called Spatio-Temporal Convolutional Neural Network (STCNN) based on convolutional long short-term memory units to address this challenge. STCNN aims to learn the spatio-temporal correlations from historical traffic data for long-term traffic predictions. Specifically, STCNN captures *the general spatio-temporal traffic dependencies* and *the periodic traffic pattern*. Further, STCNN integrates both traffic dependencies and traffic patterns to predict the long-term traffic. Finally, we conduct extensive experiments to evaluate STCNN on two real-world traffic datasets. Experimental results show that STCNN is significantly better than other state-of-the-art models.**

***Index Terms*—Spatio-temporal dependencies, periodic traffic patterns, convolutional neural network, long-term traffic predictions.**

I. Introduction

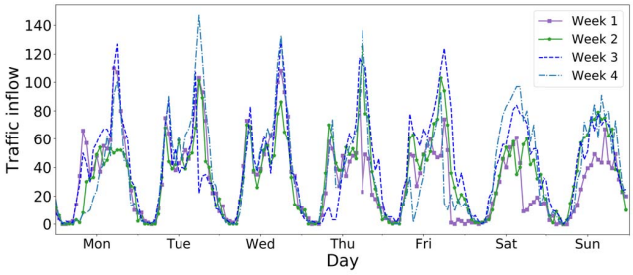
Traffic predictions are important for many location-based applications, e.g., intelligent transportation systems, trip planning, and city planning. Accurate traffic predictions will improve traffic conditions and alleviate travel delays because of higher utilization of the underlying road network capacity. For example, travelers can make timely and proper travel decisions according to predicted traffic. In general, the traffic prediction problem includes prediction for any traffic related problems, such as travel time, travel speed, traffic volume, and traffic flow. There are many studies for short-term traffic predictions, but their performance on the long-term prediction is still limited. To this end, this paper focuses on solving the problem of long-term traffic predictions.

Existing traffic prediction models mainly include data-driven statistical models and machine learning models. Among data-driven models, Auto-Regressive Integrated Moving Average (ARIMA) [1], Kalman Filters (KF) [2], [3], and aggregating methods [4], [5] are popularly used for traffic predictions. Machine learning models such as Random Forest (RF) [6], K-Nearest Neighbors (KNN) [7], Support Vector Regression (SVR) [8], and Artificial Neural Network (ANN) models [9], [10] have received much attention for traffic predictions.

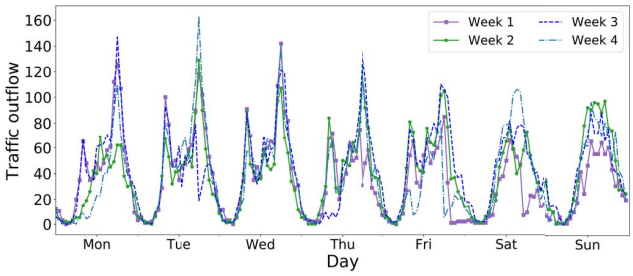
In recent years, deep learning techniques are successfully used in fields such as computer vision [11] and nature language processing [12], which encourage researchers to utilize deep learning techniques for traffic prediction problems. It is well known that convolutional neural networks (CNNs) and recurrent neural networks (RNNs) are two most popular kinds of deep neural networks. As far, there are a number of deep learning-based models applied in traffic prediction problems. For example, a deep neural network with CNNs for traffic speed prediction is proposed in [13]. RNNs are utilized to learn the temporal dependencies for predicting the traffic speed of a single location from historical traffic time series [14]. Furthermore, most studies leverage traffic time series data with spatial information for traffic predictions [15]–[18], which focus on exploring the spatio-temporal dependencies from the historical traffic series. For instance, Zhang et al. design a residual network (ST-ResNet) to discover spatio-temporal dependencies for traffic flow predictions [17], which take advantages of residual learning that makes neural networks have a very deep structure [19]. He et al. [18] propose a spatio-temporal attention neural network for traffic speed predictions, in which a spatial attention mechanism is designed for capturing spatial correlations and a temporal attention mechanism is proposed to explore the reliable temporal dependencies. Unfortunately, these studies do not consider long-term traffic predictions (e.g., predict traffic for a week in advance).

Generally, deep learning models have much better performance than traditional models; however, current deep learning models mainly focus on short-term prediction. There are few works for long-term traffic prediction [20], [21]. They are based on convolutional long short-term memory (ConvLSTM) [22] units which can simultaneously capture spatial correlations and temporal dependencies. For example, the work [20] designs a Double Spatio-Temporal Neural network (D-STN) for predicting traffic volume, but it does not make use of the periodic data and directly combines the prediction with weekly empirical mean as the final prediction, which leads to bad performance for long-term predictions (e.g., one week). The work [21] designs a Multiscale Spatiao-Temproal Feature Learning Networks (MSTFLN) that considers the one-day traffic speeds with different sizes of time intervals to predict the traffic speeds for the next future day. Although

2375-0324/19/$31.00 ©2019 IEEE
DOI 10.1109/MDM.2019.00-53



(a) Traffic inflow to the region.



(b) Traffic outflow from the region.

Fig. 1: Traffic flow data over four consecutive weeks (from 19 May, 2014 to 16 June, 2014) in a region of New York City.

it takes inputs as weekly periodic data of the day before the target predicted day in different scales, it ignores the temporal dependency from the traffic of recent previous days that have significant impacts on long-term traffic predictions.

In this paper, we propose a model, called Spatio-Temporal Convolutional Neural Network (STCNN) for predicting traffic flow of one week in advance. This work is motivated by the periodic patterns from the traffic data. For example, Fig. 1 shows the traffic flow data during four consecutive weeks in a region of New York City, USA. Fig. 1a and Fig. 1b show the traffic inflow and outflow that indicate the number of bikes with one hour-interval go to and go out from the region, respectively. As the results show clear day-period and week-period patterns, we consider periodic data (e.g., day and week) which provides significant periodic characteristics for improving the accuracy of traffic predictions.

Our proposed STCNN takes advantages of ConvLSTM units to capture spatio-temporal dependencies; thus, the one-day traffic data are presented as a spatio-temporal matrix with both spatial and temporal information, in which the first dimension denotes the time steps in a day and the second dimension presents the locations in the transportation network. STCNN adopts a general encoder-decoder architecture [12]. Specifically, the encoder consists of two components: (1) We first learn the spatio-temporal traffic dependencies using ConvLSTM over the general spatio-temporal traffic matrix series on consecutive days. (2) As simply stacking multiple ConvLSTM units cannot model the long-term temporal dependencies (e.g., period), we design a Skip-ConvLSTM model that applies ConvLSTM over skipped spatio-temporal traffic matrix series to explore the useful periodic traffic patterns for discovering long-term temporal dependencies. Note that skipped spatio-temporal traffic matrix series are the periodical time series, e.g., a series of traffic matrices in a number of consecutive historical Mondays. In the decoder, another ConvLSTM is applied to decode the spatio-temporal dependencies from the encoder to generate the future spatio-temporal hidden states which are concatenated with the periodic traffic patterns for making final long-term traffic predictions.

The contributions of this paper can be summarized as follows:

- The Skip-ConvLSTM is proposed to leverage the periodic characteristics from skipped history traffic matrix series for capturing periodic traffic patterns which make great contributions for long-term traffic predictions.
- STCNN considers the general spatio-temporal dependencies and the periodic traffic patterns to make long-term traffic flow predictions (e.g., for a week in advance).
- We conduct comprehensive experiments on two public available datasets to evaluate the performance of STCNN . Experimental results show the superiority of our STCNN compared to the state-of-the-art models.

The rest of this paper is organized as follows. Section II highlights the related work. Section III formally describes the long-term traffic prediction problem. Section IV presents the proposed model STCNN for long-term traffic predictions. Section V shows the experimental results and analysis. Finally, the conclusion is given in Section VI.

II. Related work

This section first introduces some traditional models for traffic data predictions, and then we review the deep learning-based models for spatio-temporal series prediction problems.

A. Traditional models for traffic predictions

Traffic data prediction problem is a time series prediction problem. Classical time series models (e.g., ARIMA [1] and KF [2], [3]) dominated by linear operations are popularly used for traffic predictions. For machine learning models, Sun et al. apply KNN algorithm [7] to make traffic prediction, which searches the most nearest patterns from historical data. SVR [8] is exploited for travel time predictions to further improve the prediction accuracy by transforming the traffic data into a higher dimensional feature space to make them linearly separable. Most of these models only depend on the traffic series data and estimate the traffic at an individual location; and hence, they are not able to consider the dynamic spatial correlations of the traffic data in the transportation network. ANN models are applied to traffic prediction problem (e.g., [9], [10]), due to their great abilities for modeling non-linear properties, but not enough for modeling linear properties [23]. These traditional models mainly study the short-term traffic prediction problem. The study [24] proposes a regression model with latent factors that represents the periodic features cross roads for long-term traffic flow predictions.

However, it ignores local periodic features generating from individual roads.

B. Deep learning-based spatio-temporal series predictions

Lv et al. propose a stacked autoencoder model which is a deep learning architecture to learn the potential traffic flow features, such as the non-linear spatial and temporal dependencies from the traffic data [25]. Some researchers make use of CNNs for traffic predictions (e.g., [13], [15], [17], [26]). In [13], CNNs explores spatial dependencies over an image which is converted from the transportation network traffic data. Another model based on CNNs is designed for traffic flow predictions [26]. It takes a traffic matrix with three channels (i.e., flow, speed, and occupancy) as inputs of the model to capture spatio-temporal features. Zhang et al. [17] use the ST-ResNet model based on residential CNNs to capture spatio-temporal correlations, in which ST-ResNet has a very deep structure by leveraging the residential learning. Specifically, it considers three components (i.e., closeness, period, and trend) modeling by convolutional residential networks. Yu et al. propose the spatio-temporal graph convolutional neural network which performs convolutions on graph structured traffic series to explore spatio-temporal dependencies for short-term traffic speed predictions [15].

Due to the native properties of CNNs for grid-based data and RNNs for sequential data, several deep learning neural networks separately employing CNNs for spatial correlations and RNNs for temporal correlations have been designed to make predictions [16], [27]–[29]. For example, the work [29] combines CNN and RNN for correlated time series forecasting. Li et al. [16] propose a diffusion convolutional neural network which explores the spatial dependency by modeling dynamics of the traffic flow as a diffusion process and captures the temporal dependency via an encoder-decoder architecture. CNNs are used to capture spatial features and then return them as the input to the LSTM that extracting the spatial-temporal features for traffic flow predictions [27]. The work [28] also employees an encoder-decoder architecture for exploring temporal dependencies whereas it directly utilizes the graph convolution neural network to learn spatial features on the graph of transportation network. Unfortunately, these studies do not focus on the long-term traffic prediction problem.

Few studies develop ConvLSTM based models that simultaneously capture spatial and temporal dependencies for long-term traffic predictions, e.g., D-STN [20] and MSTFLN [21]. ConvLSTM is originally proposed for precipitation nowcasting problem, in which full connections have been replaced by convolutional operations in both the input-to-state and state-to-state transitions [22]. D-STN [20] explores temporal dependencies using ConvLSTM units and catches dynamic spatial correlations using 3D-ConvNet units [30]. However, D-STN with multiple stacking ConvLSTM units cannot model the long-term spatio-temporal dependencies. MSTFLN [21] takes ConvLSTM units over weekly traffic series from multiple scales in a day to explore multiscale spatio-temporal dependencies for predicting the traffic speed of elevated highways in the next future day; however, it does not consider the traffic of recent days that has significant impacts on the future traffic predictions.

In this paper, STCNN also leverages the ability of ConvLSTM for modeling spatio-temporal dependencies. STCNN can distinguish itself from previous works as we introduce the Skip-ConvLSTM over the skipped spatio-temporal traffic matrix series for exploring periodic traffic patterns that are significant for long-term traffic predictions. In short, our main contribution is that we simultaneously consider ConvLSTM to model the general spatio-temporal dependencies and Skip-ConvLSTM to model periodic traffic patterns for long-term traffic flow predictions.

III. Preliminaries

In this section, we first give some useful definitions and the problem addressed in this paper.

1) Regions: The transportation network is segmented into $\mathtt{N}$ grid-based regions $\{\mathcal{R}_1, \mathcal{R}_2, \cdots, \mathcal{R}_{\mathtt{N}}\}$ based on the longitude and latitude.

2) Traffic inflow/outflow [31]: Let $\mathbb{P}_i$ be a set of trajectories at i-th time interval. For a region $\mathcal{R}_j$, the traffic inflow and outflow at i-th time interval are defined respectively as,

$$x_{ij}^{in} = \sum_{Tr \in \mathbb{P}_i} |\{k > 1 | v_{k-1} \notin \mathcal{R}_j \text{ and } v_k \in \mathcal{R}_j\}| \quad (1)$$

$$x_{ij}^{out} = \sum_{Tr \in \mathbb{P}_i} |\{k \geq 1 | v_k \in \mathcal{R}_j \text{ and } v_{k+1} \notin \mathcal{R}_j\}| \quad (2)$$

where $Tr: \ v_1 \rightarrow v_2 \rightarrow \cdots \rightarrow v_{|Tr|}$ is a trajectory in $\mathbb{P}_i$, v_k is the geospatial coordinate, $v_k \in \mathcal{R}_j$ means the location v_k is within region $\mathcal{R}_j$, and $|\cdot|$ is the cardinality of a set.

3) Spatio-temporal traffic matrix: Due to dynamic traffic, traffic flow can be considered as a type of time-dependent data. To achieve high prediction accuracy, the spatial information of the transportation network should also be considered. To this end, the one-day traffic flow data with their corresponding spatial information are jointly represented by a spatio-temporal matrix, $\mathbf{X} \in \mathbb{R}^{\mathtt{M} \times \mathtt{N} \times \mathtt{D}}$, where $\mathbf{x}_{ij} \in \mathbb{R}^{\mathtt{D}}$ indicates the traffic data of region $\mathcal{R}_j$ at i-th time interval in a day, $\mathtt{D}$ is the dimension of the traffic measure (we set $\mathtt{D} = 2$, $x_{ij}^0 = x_{ij}^{in}$ and $x_{ij}^1 = x_{ij}^{out}$), $\mathtt{M}$ is the total number of time steps in a day (e.g., there are 48 time steps for a day with 30-minute interval), and $\mathtt{N}$ is the number of regions in the transportation network.

4) General spatio-temporal traffic matrix series: Let $\mathcal{X} = [\mathbf{X}_1; \ldots; \mathbf{X}_t; \ldots; \mathbf{X}_{\mathtt{T}}] \in \mathbb{R}^{\mathtt{T} \times \mathtt{M} \times \mathtt{N} \times \mathtt{D}}$ be a series of traffic matrices for $\mathtt{T}$ consecutive days, where $\mathbf{X}_t \in \mathbb{R}^{\mathtt{M} \times \mathtt{N} \times \mathtt{D}}$ presents the spatio-temporal traffic matrix in t-th day.

5) Skipped spatio-temporal traffic matrix series set: To split $\mathcal{X}$ into $\mathtt{P}$ sub-sequences following the day-of-week, the length of window $\mathtt{T}$ must satisfy $\mathtt{T} = \mathtt{P} \times \mathtt{L}$, where $\mathtt{P}$ is the value of periodic pattern, and $\mathtt{L}$ is the length of consecutive weeks. For example, $\mathtt{T} = 28$, $\mathtt{P} = 7$, and $\mathtt{L} = 4$ in our experiments. After splitting, we have the set of skipped series, $\mathcal{Z} = \{\mathcal{Z}_1, \ldots, \mathcal{Z}_{t'}, \ldots, \mathcal{Z}_{\mathtt{P}}\}$, where $\mathcal{Z}_{t'} = [\mathbf{X}_{0 \times \mathtt{P}+t'}; \ldots; \mathbf{X}_{l \times \mathtt{P}+t'}; \ldots; \mathbf{X}_{(\mathtt{L}-1) \times \mathtt{P}+t'}] \in \mathbb{R}^{\mathtt{L} \times \mathtt{M} \times \mathtt{N} \times \mathtt{D}}$ is a skipped spatio-temporal traffic matrix series for $\mathtt{L}$ consecutive

weeks. In other words, $\mathcal{Z}_{t'}$ denotes the series of t'-th day in each historical week. For example, $\mathcal{Z}_1$ denotes the traffic series of historical L Mondays.

A. Long-term traffic prediction problem

We formulate the long-term traffic prediction problem as a spatio-temporal sequence prediction problem that can be solved under the general encoder-decoder architecture [12]. Specifically, let $\mathcal{X} = [\mathbf{X}_1; \ldots; \mathbf{X}_t; \ldots; \mathbf{X}_{\mathtt{T}}] \in \mathbb{R}^{\mathtt{T}\times\mathtt{M}\times\mathtt{N}\times\mathtt{D}}$ be the general spatio-temporal traffic series which consists of T days of the historical traffic data. Its corresponding skipped spatio-temporal traffic matrices series set is $\mathcal{Z} = \{\mathcal{Z}_1, \ldots, \mathcal{Z}_{t'}, \ldots, \mathcal{Z}_{\mathtt{P}}\}$. Our traffic prediction problem is to design a model over $\mathcal{X}$ and $\mathcal{Z}$ with periodic characteristics for predicting the traffic data of the following future $\mathtt{T}'$ days ($\mathtt{T}' = \mathtt{P}$ in this work), namely, $\mathcal{Y} = [\mathbf{Y}_1; \ldots; \mathbf{Y}_t; \ldots; \mathbf{Y}_{\mathtt{T}'}] \in \mathbb{R}^{\mathtt{T}'\times\mathtt{M}\times\mathtt{N}\times\mathtt{D}}$; in other words, we will predict the traffic for $\mathtt{T}'\times\mathtt{M}$ future time steps after the last time step of the current day.

IV. The Spatio-Temporal Network

In this section, we present the proposed model, including its architecture in Section IV-A, encoder for exploring traffic dynamics in Section IV-B, decoder for predicting long-term traffic in Section IV-C, and training procedure in Section IV-D.

A. Architecture

The success of ConvLSTM over spatio-temporal series data in precipitation nowcasting and video application motivates us to apply them for exploring traffic prediction problem [22], [32]. The proposed STCNN is a general encoder-decoder architecture based on ConvLSTM units, as illustrated in Fig. 2. (1) In the encoder, STCNN exploits ConvLSTM over $\mathcal{X}$ to explore the general spatio-temporal dependencies $\mathbf{H}^{conv}_{\mathtt{T}}$. Meanwhile, STCNN applies P Skip-ConvLSTMs over $\mathcal{Z}$ to capture the periodic traffic patterns, namely, the t'-th Skip-ConvLSTM takes the input of $\mathcal{Z}'_t$ and outputs its final skipped hidden state $\mathbf{H}^{skip}_{(\mathtt{L}-1)\times\mathtt{P}+t'}$, for $t' = 1, 2, \ldots, \mathtt{P}$. (2) In the decoder, STCNN utilizes the spatio-temporal hidden states from the encoder for the next week traffic predictions $(\hat{\mathbf{Y}}_1, \ldots, \hat{\mathbf{Y}}_{\mathtt{P}})$ which are sequentially generated by ConvLSTM over $\mathbf{H}'^{cs}_{t'}$ that is the concatenation of $\mathbf{H}'^{conv}_{t'}$ and $\mathbf{H}^{skip}_{(\mathtt{L}-1)\mathtt{P}+\mathtt{t}'}$, where $\mathbf{H}'^{conv}_{t'}$ is the spatio-temporal hidden state in t'-th step of the decoder. The details of STCNN are described as follows.

B. Encoder for exploring traffic dynamics

1) General spatio-temporal dependencies: RNN is often used to process sequential data, but it has the problem of gradient vanish with the increasing length of the sequence. Its variants such as LSTM and gated recurrent units (GRU) prevent this problem using the gated mechanisms. Similar to LSTM, ConvLSTM has sigmoid gates – the input gate, forget gate, and output gate and cell state. However, ConvLSTM takes convolutional operations instead of full connections in LSTM, while reducing significantly the number of parameters in the model and enhances its power to mining spatio-temporal correlations from the inputs. Given a sequence of three-dimensional inputs, following the formulations in [22], the updating procedure of ConvLSTM can be formulated as follows,

$$\mathbf{i}^{conv}_t = \sigma(\mathbf{W}^{conv}_{xi} * \mathbf{X}_t + \mathbf{W}^{conv}_{hi} * H^{conv}_{t-1} + \mathbf{W}^{conv}_{ci} \odot \mathbf{C}^{conv}_{t-1} + \mathbf{b}^{conv}_i), \tag{3}$$

$$\mathbf{f}^{conv}_t = \sigma(\mathbf{W}^{conv}_{xf} * \mathbf{X}_t + \mathbf{W}^{conv}_{hf} * \mathbf{H}^{conv}_{t-1} + \mathbf{W}^{conv}_{cf} \odot \mathbf{C}^{conv}_{t-1} + \mathbf{b}^{conv}_f), \tag{4}$$

$$\mathbf{C}^{conv}_t = f_t \odot \mathbf{C}^{conv}_{t-1} + i_t \odot \tanh(\mathbf{W}_{xc} * \mathbf{X}_t + \mathbf{W}_{hc} * \mathbf{H}^{conv}_{t-1} + \mathbf{b}^{conv}_c), \tag{5}$$

$$\mathbf{o}^{conv}_t = \sigma(\mathbf{W}^{conv}_{xo} * \mathbf{X}_t + \mathbf{W}^{conv}_{ho} * \mathbf{H}^{conv}_{t-1} + \mathbf{W}^{conv}_{co} \odot \mathbf{C}^{conv}_t + \mathbf{b}^{conv}_o), \tag{6}$$

$$\mathbf{H}^{conv}_t = \mathbf{o}^{conv}_t \odot \tanh(\mathbf{C}^{conv}_t), \tag{7}$$

where $*$ denotes convolutional operation and $\odot$ denotes element-wise product; σ and $\tanh$ are logistic sigmoid function and hyperbolic tangent function, respectively; $\mathbf{W}^{conv}$ and $\mathbf{b}^{conv}$ denote weights and biases learning from the training model. Note that the input traffic data $\mathbf{X}_t$, the output spatio-temporal hidden state $\mathbf{H}^{conv}_t$, the input gate $\mathbf{i}^{conv}_t$, forget gate $\mathbf{f}^{conv}_t$, cell state $\mathbf{C}^{conv}_t$, and output gate $\mathbf{o}^{conv}_t$ are three-dimensional matrices. The first two dimensions are temporal steps and spatial regions, respectively. The last dimension is the traffic measure. In such a case, ConvLSTM outputs the final hidden state $\mathbf{H}^{conv}_{\mathtt{T}}$ that simultaneously catches the spatial and temporal dependencies from the historical data.

2) Periodic traffic patterns: As mentioned in Section IV-B1, RNNs with GRU or LSTM are designed to alleviate the vanishing gradient problem for long-term dependency learning. Unfortunately, they still may not be sufficient to capture the long-term dependency from the long sequential data. We attempt to simplify this issue through our Skip-ConvLSTM which performs ConvLSTM over the skipped time sequences with periodic characteristics. Fig. 1a and Fig. 1b show the curves of traffic inflow and outflow data in four consecutive weeks in New York City, in which we can observe that there are similar traffic patterns among these four weeks and there are salient periodic traffic patterns in the fixed day of each week. For instance, if we consider one week as the size of period, the traffic measures at time step t in historical Mondays have a significant impact on the prediction of traffic data at time step t in the future Mondays. Therefore, skip correlations are constructed between the current hidden state and the previous consecutive periodic hidden states. The Skip-ConvLSTM over t'-th skipped spatio-temporal traffic matrix series $\mathcal{Z}_{t'}$ is formulated as follows,

$$\mathbf{i}^{skip}_{(l\times\mathtt{P})+t'} = \sigma(\mathbf{W}^{skip}_{xi} * \mathbf{X}_{(l\times\mathtt{P})+t'} + \mathbf{W}^{skip}_{hi} * \mathbf{H}^{skip}_{(l-1)\times\mathtt{P}+t'} + \mathbf{W}^{skip}_{ci} \odot \mathbf{C}^{skip}_{(l-1)\times\mathtt{P}+t'} + \mathbf{b}^{skip}_i), \tag{8}$$

$$\mathbf{f}^{skip}_{(l\times\mathtt{P})+t'} = \sigma(\mathbf{W}^{skip}_{xf} * \mathbf{X}_{(l\times\mathtt{P})+t'} + \mathbf{W}^{skip}_{hf} * \mathbf{H}^{skip}_{(l-1)\times\mathtt{P}+t'} + \mathbf{W}^{skip}_{cf} \odot \mathbf{C}^{skip}_{(l-1)\times\mathtt{P}+t'} + \mathbf{b}^{skip}_f), \tag{9}$$

$$\mathbf{C}^{skip}_{(l\times\mathtt{P})+t'} = \mathbf{f}^{skip}_{(l\times\mathtt{P})+t'} \odot \mathbf{C}^{skip}_{(l-1)\times\mathtt{P}+t'} + \mathbf{i}'_t \odot \tanh(\mathbf{W}^{skip}_{xc}$$

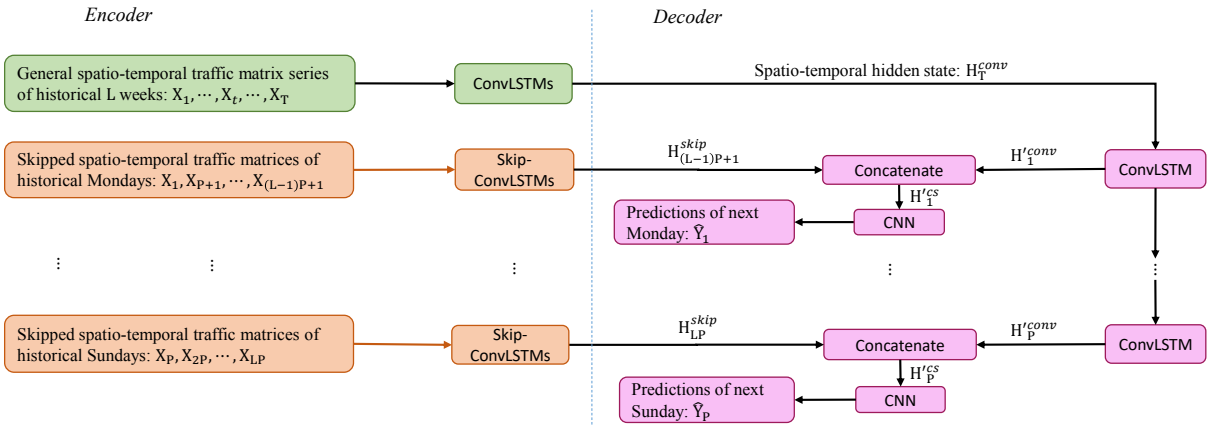


Fig. 2: Architecture of STCNN for one-week traffic predictions based on traffic data of L historical weeks. It consists of two parts: the encoder for modeling traffic dynamics and decoder for predicting long-term traffic. $\mathbf{H}_{\mathrm{T}}^{conv}$ denotes general spatio-temporal dependencies generated from ConvLSTM in the encoder; $(\mathbf{H}_{(\mathrm{L}-1)\mathrm{P}+1}^{skip}, \ldots, \mathbf{H}_{\mathrm{LP}}^{skip})$ denotes periodic traffic patterns generated from Skip-ConvLSTMs; $\mathbf{H}'^{conv}_{t'}$ is the spatio-temporal hidden state in t'-th step of the decoder; $\mathbf{H}'^{cs}_{t'}$ is the concatenation of $\mathbf{H}'^{conv}_{t'}$ and $\mathbf{H}^{skip}_{(\mathrm{L}-1)\mathrm{P}+\mathrm{t}'}$.

$$* \mathbf{X}_{(l\times \mathrm{P})+t'} + \mathbf{W}_{hc}^{skip} * \mathbf{H}_{(l-1)\times \mathrm{P}+t'}^{skip} + \mathbf{b}_c^{skip}), \tag{10}$$

$$\mathbf{o}_{(l\times \mathrm{P})+t'}^{skip} = \sigma(\mathbf{W}_{xo}^{skip} * \mathbf{X}_{(l\times \mathrm{P})+t'} + \mathbf{W}_{ho}^{skip} * \mathbf{H}_{(l-1)\times \mathrm{P}+t'}^{skip} + \mathbf{W}_{co}^{skip} \odot \mathbf{C}_{(l\times \mathrm{P})+t'}^{skip} + \mathbf{b}_o^{skip}), \tag{11}$$

$$\mathbf{H}_{(l\times \mathrm{P})+t}^{skip} = \mathbf{o}_{(l\times \mathrm{P})+t'}^{skip} \odot \tanh(\mathbf{C}_{(l\times \mathrm{P})+t'}^{skip}), \tag{12}$$

where $\mathbf{i}_{(l\times \mathrm{P})+t'}^{skip}$, $\mathbf{f}_{(l\times \mathrm{P})+t'}^{skip}$, $\mathbf{C}_{(l\times \mathrm{P})+t'}^{skip}$, and $\mathbf{o}_{(l\times \mathrm{P})+t'}^{skip}$ are the input gate, forget gate, cell state, and the output gate, respectively; $\mathbf{H}_{(l\times \mathrm{P})+t'}^{skip}$ is the skiped hidden state; $\mathbf{W}^{skip}$ and $\mathbf{b}^{skip}$ are weights and biases.

C. Decoder for predicting long-term traffic

To predict the traffic data of the future one week, we sequentially generate the predicted spatio-temporal matrix by the decoder that is another ConvLSTM. As the final hidden state $\mathbf{H}_{\mathrm{T}}^{conv}$ from the ConvLSTM in the encoder summaries the spatio-temporal dependencies of the whole input sequences, $\mathbf{H}_{\mathrm{T}}^{conv}$ is taken as the initial state for the decoder ConvLSTM. The updating steps of the decoder ConvLSTM are similar to (3)–(7). To distinguish them, we use $\mathbf{H}'^{conv}_{t'}$ to denote the hidden state at t'-th step in the decoder.

Note that the first two dimensions of $\mathbf{H}'^{conv}_{t'}$ and those of the skipped spatio-temporal hidden states $\mathbf{H}^{skip}_{(\mathrm{L}-1)\times \mathrm{P}+t'}$ from Skip-ConvLSTM are the same. We simply concatenate them together along the third dimension, written as,

$$\mathbf{H}'^{cs}_{t'} = [\mathbf{H}'^{conv}_{t'}; \mathbf{H}^{skip}_{(\mathrm{L}-1)\times \mathrm{P}+t'}], \tag{13}$$

where $\mathbf{H}'^{cs}_{t'}$ is the learned spatio-temporal features with both general spatio-temporal dependencies and periodic traffic patterns. After the encoder and decoder, we take $\mathbf{H}'^{cs}_{t'}$ as the input of CNNs to make the final prediction, given as,

$$\hat{\mathbf{Y}}_{t'} = f(\mathbf{W} * \mathbf{H}'^{cs}_{t'} + \mathbf{b}) \tag{14}$$

where $f(\cdot)$ is an activation function, $\mathbf{W}$ and $\mathbf{b}$ are weight and bias, respectively. In the convolutional operation, we set the filter size to the dimension of traffic measures with stride 1 and same padding.

D. Training

Since the proposed model can jointly handle the spatio-temporal traffic data, our architecture can be trained in a end-to-end way. The pseudo-code of training STCNN model is presented in Algorithm 1. During the training procedure, we use Adam optimizer [33] to train STCNN by minimizing the Mean Squared Error (MSE) between the predicted matrix $\hat{\mathcal{Y}}$ and the ground truth matrix $\mathcal{Y}$,

$$Loss(\Theta) = \frac{1}{n} \sum \|\mathcal{Y} - \hat{\mathcal{Y}}\|_2^2, \tag{15}$$

where n is the number of samples and Θ denotes the set of model parameters in STCNN.

V. Experiments

We first describe the datasets, evaluation metrics, compared models, and implementations in Section V-A. Then, we present the experimental results and analysis in Sections V-B, V-C, and V-D.

A. Experiment settings

Datasets. We conduct experiments over two public available datasets [17]. Each dataset includes the data type, time period, time interval, grid map size, average sampling rate, number of taxis or bikes, and available time interval. Their main statistics are depicted in TABLE I. In our experiments, we only take data from consecutive weeks and split each dataset into three parts: the first 80% of the dataset as training data, the following 10% as validation data, and the last 10% as testing data.

Algorithm 1: Pseudo-code for training procedure of STCNN

Input: Training data including historical spatio-temporal traffic matrix series $\mathcal{X}$, the set of the skipped traffic matrix series $\mathcal{Z}$, and future ground truth $\mathcal{Y}$;
Output: Learned STCNN model;
Initialization: All model parameters Θ in STCNN;
for *each epoch* **do**
  Shuffle training data;
  // Encoder for exploring traffic dynamics;
  for *each batch in training data* **do**
    $\mathbf{H}_1^{conv}, \mathbf{H}_t^{conv}, \ldots, \mathbf{H}_{\mathtt{T}}^{conv} \leftarrow$ Encoder ConvLSTMs;
  for $t' = 1$ *to* $\mathtt{P}$ **do**
    $\mathbf{H}^{skip}_{(\mathtt{L}-1)\mathtt{P}+t'} \leftarrow$ Skip-ConvLSTMs with $\mathcal{Z}_{t'}$;
  // Decoder for predicting future traffic;
  Initialize decoder hidden state: $\mathbf{H}'_0 = \mathbf{H}_{\mathtt{T}}^{conv}$;
  for $t' = 0$ *to* $\mathtt{T}'$ **do**
    $\mathbf{H}'^{conv}_{t'} \leftarrow$ Decoder ConvLSTMs ;
    $\mathbf{H}'^{cs}_{t'} \leftarrow$ Concatenate spatio-temporal hidden states by (13);
    $\hat{\mathbf{Y}}_{t'} \leftarrow$ CNNs with $\mathbf{H}'^{cs}_{t'}$ by (14);
  $\hat{\mathcal{Y}} = [\hat{\mathbf{Y}}_0; \ldots; \hat{\mathbf{Y}}_{t'}; \ldots; \hat{\mathbf{Y}}_{\mathtt{T}'}]$;
  Optimize Θ by minimizing (15);

TABLE I: Dataset statistics

Dataset	TaxiBJ	BikeNYC
Data type	Taxi GPS	Bike rent
Location	Beijing	New York
Time period	07/01/2013 - 10/30/2013 03/01/2014 - 06/30/2014 03/01/2015 - 06/30/2015 11/01/2015 - 04/10/2016	04/01/2014 - 09/30/2014
Time interval	30 minutes	1 hour
Gird map size	32×32	16×8
Average sampling rate	about 60 seconds	-
Number of taxis/bikes	$34,000+$	$6,800+$
Available time intervals	22,459	4,392

Evaluation metrics. Two standard metrics, namely, Root Mean Squared Error (RMSE) and Mean Absolute Error (MAE), are used to evaluate the performance of traffic predictions, i.e., $RMSE = \sqrt{\frac{1}{m}\sum_{i=1}^{m}(\hat{y}_i - y_i)^2}$ and $MAE = \frac{1}{m}\sum_{i=1}^{m}\|\hat{y}_i - y_i\|_1$, where m is the size of the testing set, y_i and $\hat{y}_i$ are the ground truth and the predicted value, respectively. For these two metrics, smaller values indicate better performance.

Compared models. To evaluate the performance of the proposed STCNN, we compare it with eight models, including three classical machine learning-based models and five state-of-the-art deep learning-based models. The compared models are listed below:

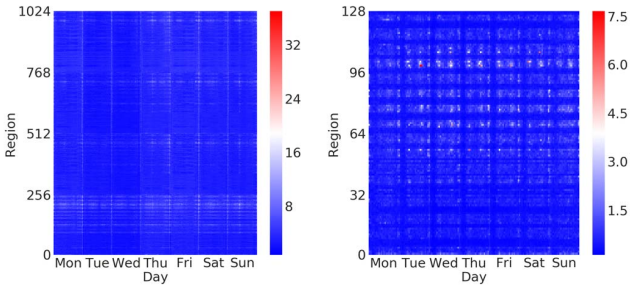
- SVR [8] is a well-known machine learning method.
- RF [6] is a machine learning-based ensemble method which is used for regression to make traffic predictions.
- ARIMA [1] is a clsssical time series prediction model.
- Seq2seq [12] is the general encoder-decoder architecture for sequential data.
- ST-ResNet [17] utilizes the residual neural network framework which can have a very deep structure to model the temporal closeness, period, and trend features of traffic flow data.
- D-STN [20] combines the prediction of a STN and the empirical mean of observed traffic volumes as the final prediction. The STN is composed of ConvLSTMs for temporal dependencies and three-dimensional ConvNet layers for local fluctuations [30].
- MSTFLN [21] is a deep learning network consisting of multiple ConvLSTMs and CNNs and is designed for long-term traffic speed predictions.
- Seq-ConvLSTM [22] is an encoder-decoder network based on ConvLSTM. ConvLSTM is a special RNN by using the convolutional operation in LSTMs to capture spatio-temporal correlations for precipitation nowcasting.
- **STCNN** is our proposed model that learns the general spatio-temporal traffic dependencies and the periodic traffic patterns from historical traffic data for long-term traffic predictions.

Implementations. We take the observations in four consecutive weeks (i.e., $\mathtt{L} = 4, \mathtt{P} = 7$) to predict traffic for seven days in advance. In SVR, we use Radial Basis Function kernel with the kernel coefficient 0.1 for training. In RF, we build 50 trees without the maximum depth constraint, the minimum number of samples for splitting an internal node is set to 128, and the random state is set to 2. All the compared deep learning-based models are implemented by TensorFlow framework. For Seq2seq, ST-ResNet, MSTFLN and D-STN, we tune hyperparameters by following the corresponding literature to get the optimal performances. As ST-ResNet is originally designed for one-step ahead prediction, we report its one-step result. MSTFLN can only make prediction for the next future day from the original setting, so we use its result of one day prediction for comparison. Since MSTFLN and D-STN cannot process traffic inflow and outflow data at the same model, we separately train the models on the traffic inflow or outflow data and take the average value of the two models as the final result. As D-STN does not converge on the TaxiBJ dataset, we do not report its performance for this dataset.

For Seq-ConvLSTM, we use two ConvLSTM layers with 128 and 32 units, both in the encoder and the decoder; three CNNs are used to make final predictions in the decoder and their embedding sizes are 256, 64, and 2. For **STCNN**, we also use the same setting for ConvLSTM layers and the three CNNs as Seq-ConvLSTM, both in the encoder and the decoder; and we use two Skip-ConvLSTM layers with 32 and 8 units in the encoder. The learning rate is set to 0.0001. During training, we take the early stop strategy by monitoring the value of valid loss with the maximal epoch 50.

TABLE II: The RMSE and MAE for one-week traffic predictions on the TaxiBJ and BikeNYC datasets

Dataset	TaxiBJ		BikeNYC	
Metric	RMSE	MAE	RMSE	MAE
SVR	76.34	65.03	34.09	31.37
RF	70.15	42.39	23.40	11.88
ARIMA	57.85	48.93	16.56	11.88
MSTFLN	51	29.80	22.29	10.52
D-STN	–	–	22.19	10.29
Seq2seq	17.39	10.33	9.00	4.21
Re-STNet	17.47	9.86	6.32	2.46
Seq-ConvLSTM	4.45	3.46	1.74	1.27
STCNN	**4.08**	**3.18**	**1.36**	**0.92**



(a) TaxiBJ (b) BikeNYC

Fig. 3: Heatmaps of MAEs over the day-of-the-week on TaxiBJ and BikeNYC datasets.

B. Overall results

TABLE II presents the RMSE and MAE of all the compared models for one-week traffic predictions (i.e., traffic predictions for seven days in advance) based on the BikeNYC and TaxiBJ datasets. The proposed **STCNN** consistently achieves the best accuracy among all the compared models. SVR has the worst performance and its RMSE and MAE are higher about 20 times than those of **STCNN**. RF, ARIMA, MSTFLN, and D-STN have better performance, whereas they are still much worse than Seq-ConvLSTM and **STCNN**. The key reason for the bad performance of MSTFLN is that it ignores the significant spatio-temporal information from recent historical days. As for D-STN, it simply stacks the ConvLSTM units for exploring temporal dependencies and does not take the input as the spatio-temporal traffic matrix of a day that simultaneously contains spatial and temporal information. These issues lead to D-STN reluctantly to predict traffic flow for one week in advance. Seq2seq and Re-STNet are also worse than Seq-ConvLSTM and **STCNN**. As Seq2seq directly applies LSTM on the historical data, which is weak for exploring spatio-temporal correlations. Re-STNet applies residential learning that is not good at long-term prediction. As a result, **STCNN** outperforms the state-of-the-art models.

C. Evaluation on the days of the week

As shown in Section V-B, **STCNN** and Seq-ConvLSTM significantly outperform other compared models; and thus, we further compare the performance of these models in this section. We here evaluate the average RMSE and MAE of **STCNN** and Seq-ConvLSTM for one-week traffic predictions with respect to the days of the week, i.e., Monday, Tusday, Wednesday, Thursday, Friday, Saturday, and Sunday. TABLE III and TABLE IV present RMSE and MAE from Monday to Sunday on the TaxiBJ and BikeNYC, respectively. Both tables indicate that **STCNN** has lower RMSE and MAE than Seq-ConvLSTM for all the days of the week. For example, on Monday, **STCNN** reduces RMSE from 4.69 to 4.19 and MAE from 3.81 to 3.41 on the TaxiBJ dataset. Likewise, **STCNN** reduces RMSE from 1.64 to 1.17 and MAE from 1.28 to 0.92 on the BikeNYC dataset. These results show the important role of Skip-CovLSTM for improving prediction accuracy as it provides the periodic traffic patterns. In addition, we use the absolute errors, i.e., MAE, to indicate the gaps between predictions and ground truth. Fig. 3a and Fig. 3b depict the heatmaps of average MAE of inflow and outflow on the TaxiBJ and BikeNYC datasets, respectively. Blue implies smaller error and red color implies the larger error. We can see that most areas of both two heatmaps are in blue and only few points are in red. These results suggest that the predicted traffic of our proposed **STCNN** is very close to the actual traffic.

D. Discussion on the Effect of Skip-ConvLSTM

As **STCNN** and Seq-ConvLSTM perform much better than other compared models, we mainly focus on the comparison of these two models and verify the effectiveness of Skip-ConvLSTM designed for **STCNN**. From the last two rows in TABLE II, we observe that **STCNN** has at least 8.3% improvement on RMSE and 8.1% on MAE on the TaxiBJ dataset. Similarly, **STCNN** records at least 20% improvement compared to Seq-ConvLSTM in terms of RMSE and MAE on BikeNYC dataset. Since **STCNN** takes advantages of Skip-ConvLSTM for modeling periodic traffic patterns that Seq-ConvLSTM does not consider; it implies the important role of Skip-ConvLSTM in **STCNN** for improving the long-term traffic prediction accuracy. Likewise, the results in TABLE III and TABLE IV lead to the importance of Skip-ConvLSTM. Our explanation is that: **STCNN** exploits not only the general spatio-temporal traffic dependencies but also the periodic traffic patterns for long-term traffic predictions, which makes a great contribution for improving the prediction accuracy.

VI. Conclusion

In this paper, we proposed a Spatio-Temporal Convolutional Neural Network (STCNN) for long-term traffic predictions (i.e., for one week in advance). STCNN employs an encoder-decoder architecture. In the encoder, it firstly captures the general spatio-temporal dependencies by ConvLSTM over the general spatio-temporal traffic matrix series, and then explores periodic traffic patterns using Skip-ConvLSTM over the skipped spatio-temporal traffic matrix series. After that, the decoder utilizes another ConvLSTM to decode the spatio-temporal dependencies and combines them with the periodic

TABLE III: The average RMSE and MAE of one-week traffic predictions between **STCNN** and Seq-ConvLSTM for each day-of-the-week on the TaxiBJ dataset.

Model	Metric	Mon	Tue	Wed	Thu	Fri	Sat	Sun	Mean
Seq-ConvLSTM	RMSE	4.69	3.72	3.56	4.89	5.00	4.50	4.61	4.42
STCNN	RMSE	**4.19**	**3.28**	**3.18**	**4.72**	**4.55**	**4.27**	**4.18**	**4.05**
Seq-ConvLSTM	MAE	3.81	2.92	2.72	3.78	3.94	3.48	3.57	3.46
STCNN	MAE	**3.41**	**2.57**	**2.44**	**3.67**	**3.60**	**3.34**	**3.26**	**3.18**

TABLE IV: The average RMSE and MAE of one-week traffic predictions between **STCNN** and Seq-ConvLSTM for each of the days of the week on the BikeNYC dataset.

Model	Metric	Mon	Tue	Wed	Thu	Fri	Sat	Sun	Mean
Seq-ConvLSTM	RMSE	1.64	1.72	1.89	1.69	1.80	1.60	1.84	1.74
STCNN	RMSE	**1.17**	**1.23**	**1.41**	**1.27**	**1.33**	**1.18**	**1.31**	**1.27**
Seq-ConvLSTM	MAE	1.28	1.40	1.41	1.32	1.35	1.33	1.43	1.36
STCNN	MAE	**0.92**	**0.91**	**0.92**	**0.87**	**0.90**	**0.90**	**1.00**	**0.92**

traffic patterns for long-term traffic predictions. Experimental results on two public available datasets show that STCNN can achieve better prediction accuracy than the state-of-the-art models.

VII. ACKNOWLEDGMENTS

This research was partially supported by the Innovation and Technology Fund (ITF) under Grant No. UIM/334.

REFERENCES

[1] M. S. Ahmed and A. R. Cook, *Analysis of freeway traffic time-series data by using Box-Jenkins techniques*. Transportation Research Board, 1979, no. 722.

[2] I. Okutani and Y. J. Stephanedes, "Dynamic prediction of traffic volume through kalman filtering theory," *Transportation Research Part B: Methodological*, vol. 18, no. 1, pp. 1–11, 1984.

[3] C. Kuchipudi and S. Chien, "Development of a hybrid model for dynamic travel-time prediction," *Transportation Research Record: Journal of the Transportation Research Board*, no. 1855, pp. 22–31, 2003.

[4] J.-D. Zhang, J. Xu, and S. S. Liao, "Sampling methods for summarizing unordered vehicle-to-vehicle data streams," *Transportation Research Part C: Emerging Technologies*, vol. 23, pp. 56–67, 2012.

[5] ——, "Aggregating and sampling methods for processing GPS data streams for traffic state estimation," *IEEE TITS*, vol. 14, no. 4, pp. 1629–1641, 2013.

[6] L. Breiman, "Random forests," *Machine Learning*, vol. 45, no. 1, pp. 5–32, 2001.

[7] H. Sun, H. X. Liu, H. Xiao, R. R. He, and B. Ran, "Short term traffic forecasting using the local linear regression model," in *Processings of the Annual Meeting of the Transportation Research Board*, 2003.

[8] C.-H. Wu, J.-M. Ho, and D. T. Lee, "Travel-time prediction with support vector regression," *IEEE TITS*, vol. 5, no. 4, pp. 276–281, 2004.

[9] H. Dia, "An object-oriented neural network approach to short-term traffic forecasting," *European Journal of Operational Research*, vol. 131, no. 2, pp. 253–261, 2001.

[10] J. Van Lint, "Reliable real-time framework for short-term freeway travel time prediction," *Journal of Transportation Engineering*, vol. 132, no. 12, pp. 921–932, 2006.

[11] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *NIPS*, 2012.

[12] I. Sutskever, O. Vinyals, and Q. V. Le, "Sequence to sequence learning with neural networks," in *NIPS*, 2014.

[13] X. Ma, Z. Dai, Z. He, J. Ma, Y. Wang, and Y. Wang, "Learning traffic as images: A deep convolutional neural network for large-scale transportation network speed prediction," *Sensors*, vol. 17, no. 4, pp. 1–16, 2017.

[14] X. Ma, Z. Tao, Y. Wang, H. Yu, and Y. Wang, "Long short-term memory neural network for traffic speed prediction using remote microwave sensor data," *Transportation Research Part C: Emerging Technologies*, vol. 54, pp. 187–197, 2015.

[15] B. Yu, H. Yin, and Z. Zhu, "Spatio-temporal graph convolutional neural network: A deep learning framework for traffic forecasting," in *IJCAI*, 2018.

[16] Y. Li, R. Yu, C. Shahabi, and Y. Liu, "Diffusion convolutional recurrent neural network: Data-driven traffic forecasting," in *ICML*, 2018.

[17] J. Zhang, Y. Zheng, D. Qi, R. Li, X. Yi, and T. Li, "Predicting citywide crowd flows using deep spatio-temporal residual networks," *Artificial Intelligence*, vol. 259, pp. 147 – 166, 2018.

[18] Z. He, C. Chow, and J. Zhang, "STANN: A spatiotemporal attentive neural network for traffic prediction," *IEEE Access*, vol. 7, pp. 4795–4806, 2019.

[19] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *IEEE CVPR*, 2016.

[20] C. Zhang and P. Patras, "Long-term mobile traffic forecasting using deep spatio-temporal neural networks," in *ACM MobiHoc*, 2018.

[21] D. Zang, J. Ling, Z. Wei, K. Tang, and J. Cheng, "Long-term traffic speed prediction based on multiscale spatio-temporal feature learning network," *IEEE TITS*, pp. 1–10, 2018.

[22] X. Shi, Z. Chen, H. Wang, D. Yeung, W. Wong, and W. Woo, "Convolutional LSTM network: A machine learning approach for precipitation nowcasting," in *NIPS*, 2015.

[23] G. Zhang and M. Qi, "Neural network forecasting for seasonal and trend time series," *European Journal of Operational Research*, vol. 160, no. 2, pp. 501 – 514, 2005.

[24] M. Okawa, H. Kim, and H. Toda, "Online traffic flow prediction using convolved bilinear poisson regression," in *IEEE MDM*, 2017.

[25] Y. Lv, Y. Duan, W. Kang, Z. Li, F.-Y. Wang *et al.*, "Traffic flow prediction with big data: A deep learning approach," *IEEE TITS*, vol. 16, no. 2, pp. 865–873, 2015.

[26] D. Zang, Y. Fang, D. Wang, Z. Wei, K. Tang, and X. Li, "Long term traffic flow prediction using residual net and deconvolutional neural network," in *PRCV*, 2018.

[27] Y. Liu, H. Zheng, X. Feng, and Z. Chen, "Short-term traffic flow prediction with Conv-LSTM," in *IEEE WCSP*, 2017.

[28] B. Liao, J. Zhang, C. Wu, D. Mcilwraith, T. Chen, S. Yang, Y. Guo, and F. Wu, "Deep sequence learning with auxiliary information for traffic prediction," in *ACM SIGKDD*, 2018.

[29] R.-G. Cirstea, D.-V. Micu, G.-M. Muresan, C. Guo, and B. Yang, "Correlated time series forecasting using multi-task deep neural networks," in *ACM CIKM*, 2018.

[30] S. Ji, W. Xu, M. Yang, and K. Yu, "3D convolutional neural networks for human action recognition," *IEEE TPAMI*, vol. 35, no. 1, pp. 221–231, 2013.

[31] J. Zhang, Y. Zheng, D. Qi, R. Li, and X. Yi, "DNN-based prediction model for spatio-temporal data," in *ACM SIGSPATIAL*. ACM, 2016.

[32] Y. Tang, W. Zou, Z. Jin, and X. Li, "Multi-scale spatiotemporal conv-lstm network for video saliency detection," in *ACM ICMR*, 2018.

[33] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," in *ICLR*, 2015.