CLSTAN: ConvLSTM-Based Spatiotemporal Attention Network for Traffic Flow Forecasting

Liyan Xiong, 1 Weihua Ding, 1 Xiaohui Huang, 1 and Weichun Huang 2

1 School of Information Engineering, East China Jiaotong University, Nanchang 330013, China
2 School of Software, East China Jiaotong University, Nanchang 330013, China

Correspondence should be addressed to Weihua Ding; weihua.ding@foxmail.com

Received 7 May 2022; Revised 21 June 2022; Accepted 24 June 2022; Published 11 July 2022

Traffic flow forecasting is the essential part of intelligent transportation system (ITS), which can fully protect traffic safety and improve traffic system management capability. Nevertheless, it is still a challenging problem, which is influenced by many complex factors, including regional distribution and external factors (e.g., holidays and weather). To combine various factors to forecast traffic flow, we presented a novel neural network structure called ConvLSTM-based Spatiotemporal Attention Network (CLSTAN).

Specifically, our proposed model is composed of four modules: a preliminary feature extraction module, a spatial attention module, a temporal attention module, and an information fusion module. The spatiotemporal attention module can efficiently learn the complex spatiotemporal patterns of traffic flow through the attention mechanism. The spatial attention module uses a series of initial traffic flow maps as input and obtains the weights of the various regions through a ConvLSTM. The temporal attention module uses the spatially weighted traffic flow map as input and acquires the complex spatiotemporal patterns of traffic flow by a ConvLSTM that introduces an attention mechanism. Finally, the information fusion module integrates spatiotemporal information from multiple time dimensions to forecast future traffic flow. Moreover, to confirm the validity of our method, our experiments were conducted extensively on the TaxiBJ and BikeNYC datasets, and ultimately, CLSTAN performed better than other baseline experiments.

1. Introduction

Intelligent transportation system [1] is very important for the construction and development of modern cities. Traffic flow forecasting [2, 3], as an indispensable part of the intelligent transportation system, can be used as an index to evaluate the road state. Through traffic flow forecasting, the government can better conduct urban management [4] as well as social security management. According to related reports, the average monthly speed in Shanghai, China, is 23 km/h. Severe traffic congestion often causes a lot of inconveniences for people to travel. If traffic flow can be forecasted, early warnings can be made based on the results of traffic flow forecasting and measures can be taken in advance, thus avoiding the occurrence of city-wide or even intercity congestion. Therefore, in recent years, traffic flow forecasting has attracted extensive research interest from academia and industry.

Although this work has been extensively studied in recent years, however, trying to accurately forecast traffic flow is still a challenging task since it is often affected by many factors, such as region distribution, weather, traffic accident, and other external factors.

The main challenges are as follows:

1) Temporal correlation: as shown in Figure 1, the temporal correlation is composed of sequentiality and periodicity. In detail, sequentiality indicates that the traffic flow changes smoothly between adjacent time intervals, as shown in Figure 1(a). For instance, the traffic flow changes smoothly from 2 pm to 3 pm. And then, periodicity indicates that traffic flow usually repeats at a certain frequency. For instance, the traffic flow at 11 am today is similar to the flow at 11 am yesterday, as shown in Figure 1(b).
Spatial correlation: spatial correlation refers to the characteristic that traffic flow data changes may show different trends in different regions. As shown in Figure 2, the red area has serious traffic congestion due to more vehicles, while the green area has good traffic conditions. Therefore, compared with other areas, the red area traffic flow may change more drastically.

Existent factors: in a transportation system, different environmental factors have different degrees of influence to the traffic flow, such as holidays, weather, and events.

To resolve the challenges mentioned above, we propose an innovative deep learning model CLSTAN to combine various factors to forecast traffic flow. This model consists of four main modules: (1) the preliminary feature extraction module, (2) the spatial attention module, (3) the temporal attention module, and (4) the information fusion module. In subsequent sections, we will describe the implementation details of each module.

In conclusion, our main contributions are as follows:

1. We propose the spatial attention module and the temporal attention module based on ConvLSTM to enhance the capture of spatial and temporal correlations in spatiotemporal data.

2. Based on the spatial attention module and temporal attention module, we then introduced the preliminary feature extraction module and the information fusion module to design an innovative neural network called CLSTAN. The model integrates temporal correlation, spatial correlation, and external factors for traffic flow forecasting.

3. We conducted numerous experiments on two publicly available datasets (TaxiBJ and BikeNYC). The results of the experiments demonstrate the effectiveness of our proposed model.

The outline of the remaining sections is as follows. Section 2 describes the related work. Section 3 describes the concepts related to traffic flow and defines traffic flow forecasting. Section 4 details our proposed model. Section 5 describes the study of our experiments. In the end, we summarize our work and look forward to the future in Section 6.

2. Related Works

2.1. Traffic Flow Forecasting. Traffic flow forecasting has been extensively investigated over the past, and researchers have achieved numerous results. The study of traffic flow forecasting is mainly divided into three areas. Statistical models used for traffic flow forecasting include HA, ARIMA, and VAR [5–7]. The most representative is the autoregressive integrated moving average (ARIMA) and its variants [8–10]. Alghamdi et al. [11] propose a method to model traffic congestion using ARIMA. Compared to other models such as ARCH [12] and its variants [13], ARIMA ignores the spatial dependence. Ding et al. integrated the ARIMA and GARCH algorithms to propose ARIMA-GARCH [14] to make short-term traffic forecasts. These methods require data that meet certain assumptions, but
traffic flow data are so complex that they do not meet these assumptions, so they often do not work well in practice.

Later, with the research breakthroughs in machine learning methods, it is widely used in various fields. For example, Hu et al. [15–18] proposed various machine learning models for iron ore sintering with process based on Fuzzy C-Means Clustering and Differential Evolution algorithms. These models are able to perform carbon efficiency prediction under different conditions and greatly improve the prediction of carbon efficiency. Therefore, machine learning methods were also applied to traffic flow forecasting, such as KNN [19] and SVM [20]. Castro-Neto et al. [21] proposed a machine learning model based on online support vector machines to make short-term traffic forecasts. Sun et al. [22] proposed a machine learning method based on the Bayesian network to make short-time traffic forecasts, and the traffic flow between adjacent road links to the traffic network is modeled as the Bayesian network. These methods can model more complex data, but these methods do not work well on larger datasets.

Recently, DNN-based methods are widely used in different areas with many achievements. Inspired by these studies, many researchers have tried to employ deep learning algorithms to solve traffic flow forecasting problems. For instance, Zhang et al. [23] proposed a DNN-based model DeepST for extracting spatiotemporal attributes from traffic flow data. It designed the spatiotemporal component to be able to model both spatial near and distant dependencies. And later, they improved a deep learning framework ST-ResNet [24] based on ResNet [25] and used various temporal attributes (proximity, periodicity, and trend) in traffic flow for city-wide traffic flow forecasting. Yao et al. [26] presented a deep multiview spatiotemporal network for traffic flow forecasting, which takes a similar approach to graph convolution to obtain the spatial dependency. Liu et al. [27] proposed a component called Attentive Traffic Flow Machine (ATFM) and was able to efficiently extract spatiotemporal information from the traffic flow. Lin et al. [28] proposed a model called SpAE-LSTM, which extracts spatial features by sparse autoencoder and temporal features by LSTM. Yao et al. [29] proposed a traffic gating mechanism to extract the dynamic correlation between different regions and proposed a periodic attention mechanism to handle long-term time-series data. Ma and Song et al. [30–33] proposed a series of deep learning model for daily traffic flow forecasting. These methods focus on mining the relationship between traffic flow patterns and contextual factors. Experiments demonstrate that methods combining contextual factors and traffic patterns can improve prediction performance. Although all of the above studies have yielded good achievements, they all have areas for improvement.

2.2. Deep Learning. Convolutional neural network (CNN) [34], as a classical deep learning method [35], can extract features of images by different receptive fields so that it can be used to extract spatial characteristics of traffic flow. However, it cannot be used for feature extraction of time series. Recurrent neural network (RNN) [36] can be used to extract time-series features. Based on the recurrent neural network, researchers have improved various variants, such as the long short-term memory network (LSTM) [37] and the gated recurrent unit (GRU) [38]. Experimentally, these variants were shown to be better model time series and used to explore the temporal relationship of traffic flow. Shi et al. [39] combined the above approaches and presented the convolutional long short-term memory network (ConvLSTM). And then, Xiong et al. [40] employed the convolutional long and short-term memory network for spatiotemporal modeling of traffic flow. It was demonstrated that the convolutional long and short-term memory network can extract the spatiotemporal information of traffic flow effectively.

2.3. Attention Mechanism. In recent studies, attention mechanisms have been widely used in different tasks such as natural language processing [41, 42], image caption [43], and speech recognition [44, 45]. Xu et al. [46] proposed two attention mechanisms in an image recognition task and used visualization to graphically demonstrate the effects of the attentional mechanisms. V elickovic et al. [47] presented a network structure with attention mechanism and experimented on graph-structured data, which showed that they could notice the most critical parts of the graph-structured data. Liang et al. [48] presented a multilevel attention mechanism network to predict time series with excellent results.

3. Preliminaries

In this section, we will introduce some relevant definitions of traffic flow forecasting.

3.1. Traffic Networks. In previous studies [49, 50], researchers have used a variety of methods to split a city into areas, such as zip codes or latitude and longitude. In this study, we will split the city into $H \times W$ square grid maps according to latitude and longitude. Each grid represents a different geographical location of the city. Specifically, we represent each grid as $g(i, j) \mid 1 \leq i \leq H, 1 \leq j \leq W$.

3.2. Traffic Flow Map. In the real world, we are able to obtain a large number of tracks of taxis and bicycles through cellphone signals and GPS signals. By using these tracks, we can get the amount of bikes or taxis entering and leaving a certain area in a certain time interval. In this study, we denote the amount of vehicles entering and leaving a given area in a given time interval as inflow and outflow. Specifically, we refer to the traffic flow map at time interval $t$ on day $d$ as $M_{d,t} \in \mathbb{R}^{H \times W}$, where the first channel is inflow and the second channel is outflow. Figure 3 shows an example of inflow and outflow.

3.3. External Factors. Zhang et al. [24] demonstrated that traffic flow is influenced by complicated external factors. For instance, an unexpected downpour will cause a sudden
traffic congestion in a certain area, or people may congregate in a busy commercial area on a holiday, causing a large increase in traffic flow in the area compared to normal. In this study, we mainly focus on the impact of weather and holidays on traffic flow forecasting. We encode the weather and holiday information by the One-Hot Encoding method and connect all the external factors to a tensor. Specifically, we denote the external factors for the time interval \( t \) on day \( d \) as \( E_d^t \).

### 3.4. Convolutional LSTM Network

Shi et al. first proposed the ConvLSTM and used it to make short-time precipitation forecasting. They defined the short-time precipitation forecasting as spatiotemporal sequence forecasting problem. However, LSTM needs to expand the spatio-temporal data into 1D vector when solving the spatio-temporal sequence forecasting problem, which makes the spatial information is lost. To solve this problem, Shi et al. replace the fully connected layer of each gate in LSTM with ConvLSTM is suitable for processing spatiotemporal data. Compared to LSTM, ConvLSTM not only can model the temporal relationship like LSTM but also can extract spatial features such as CNN. Experiments demonstrate that ConvLSTM can better capture spatiotemporal correlations compared to LSTM.

Traffic flow is also a typical spatiotemporal data, and ConvLSTM is suitable for processing spatiotemporal data. So, we propose our spatial attention module and temporal attention module based on ConvLSTM. The structure of ConvLSTM is similar to that of the traditional LSTM, we define the computational procedure of the ConvLSTM as follows.

For \( t \)-th ConvLSTM cell in a layer, the input consists of the past cell state \( C_{t-1} \), the past hidden state \( H_{t-1} \), and input \( X_t \). The output is the updated hidden state \( H_t \) and the updated cell state \( C_t \). The cell state is determined by the gating mechanism \( (i_t, f_t, o_t) \). The input gate \( i_t \) determines what degree new information is recorded into the cell state, the forget gate \( f_t \) determines how much the previous cell state \( C_{t-1} \) will be forgotten, and the output gate \( o_t \) determines what degree information about \( C_t \) is transferred to the hidden state \( H_t \). Then, the updating formulas for ConvLSTM are given below:

\[
\begin{align*}
i_t &= \sigma(W_{xi} \cdot X_t + W_{hi} \cdot H_{t-1} + W_{ci} \cdot C_{t-1} + b) \\
f_t &= \sigma(W_{xf} \cdot X_t + W_{hf} \cdot H_{t-1} + W_{cf} \cdot C_{t-1} + b) \\
c_t &= f_t \odot C_{t-1} + i_t \odot \tanh(W_{xc} \cdot X_t + W_{hc} \cdot H_{t-1} + b) \\
o_t &= \sigma(W_{xo} \cdot X_t + W_{ho} \cdot H_{t-1} + W_{co} \cdot C_t + b) \\
H_t &= o_t \odot \tanh(c_t)
\end{align*}
\]

where \( \cdot \) denotes the convolution operator and \( \odot \) denotes the Hadamard product and \( W_{xi}, W_{hi}, W_{ci}, b \in \{ i, f, o, c \} \) are all learnable parameters.

### 3.5. Traffic Flow Forecasting

Given a series of past traffic flow maps up to the time interval \( t \) on day \( d \) and external factors, our goal is to forecast the future traffic flow map \( M^{t+1}_d \) for the time interval \( t + 1 \) on day \( d \):

\[
M^{t+1}_d = F_d(M^{t-m-d}_d, E^{t-m-d}_d)
\]
4.1. Preliminary Feature Extraction Module. In this section, we describe in detail how this module performs feature extraction for traffic flow and external factors.

For the traffic flow maps, we use two convolutional layers and multiple residual network units to obtain feature embedding from a given set of traffic flow maps \( M_{ij} \in \mathbb{R}^{2 \times h \times w} \), as shown in Figure 5(a). Each residual unit consists of two convolutional layers; the specific structure is shown in Figure 5(b). By feeding the traffic flow map into the preliminary feature extraction module, we can obtain the extracted traffic flow features \( \hat{M}_{ij} \in \mathbb{R}^{2 \times h \times w} \).

For the external factors, we employ two fully connected layers to extract features from the given external factors \( E_{ij} \), as shown in Figure 6. Since the subsequent work needs to fuse the extracted traffic flow features \( \hat{M}_{ij} \) with external factors features \( \hat{E}_{ij} \), we reshape the obtained external factor features into \( \hat{E}_{ij} \in \mathbb{R}^{2 \times h \times w} \).

After the preliminary extraction of traffic flow and external factors' features, we fuse these two extracted features to generate a new feature \( \hat{F}_{ij} \in \mathbb{R}^{2 \times h \times w} \) and denote as follows:

\[
\hat{F}_{ij} = \hat{M}_{ij} + \hat{E}_{ij}, \tag{3}
\]

where \( \hat{F}_{ij} \) is a combination of traffic flow features and external factors at a particular time, which will be fed to subsequent modules for learning their features in spatial and temporal dimensions.

4.2. The Spatial Attention Module. For a specific time, traffic flow changes in different regions are different. For instance, during the morning rush hour, the change in traffic flow in residential areas and industrial parks is undoubtedly huge compared to other areas such as commercial areas. Therefore, as shown in Figure 7, we believe that, to make traffic flow forecasting more accurate, we need to assign higher weights to areas with more dramatic traffic flow changes. Therefore, we propose the spatial attention module for inferring the spatial weights of each region and assigning the obtained spatial weights to the original traffic flow data.

The specific structure of the spatial attention module is shown in Figure 8. The spatial attention module uses the hidden output state of a ConvLSTM with the current input to deduce the spatial weights of each region. And then, the obtained spatial weights of each region are assigned to the current input and used as the input of the temporal attention module.

Specifically, through ConvLSTM, we can obtain the future state of the traffic flow map \( \hat{H}_{ij} \):

\[
\hat{H}_{ij} = \text{ConvLSTM}(\hat{F}_{ij}). \tag{4}
\]
And then, combine the obtained future state $H_j^i$ with the current input $F_j^i$ to obtain the spatial weights of each region $W_j^i$:

$$W_j^i = \text{Conv}_{1\times1}(F_j^i \oplus H_j^i),$$  \hspace{1cm} (5)

where $\oplus$ denotes the concatenation operation and $\text{Conv}_{1\times1}$ denotes the convolution operation with the convolution kernel of $1 \times 1$.

Finally, we multiply the obtained spatial attention weights $W_j^i$ with the current input $F_j^i$, according to the element positions to get the spatially weighted traffic flow data $\tilde{F}_j^i$:

$$\tilde{F}_j^i = F_j^i \times W_j^i,$$  \hspace{1cm} (6)

where $\times$ denotes the Hadamard product.

4.3. The Temporal Attention Module. When temporal relationships need to be modeled, we advocate the use of LSTM as the main part of the temporal attention module, and considering the particularity of traffic flow maps, ConvLSTM is better able to perform this task. However, traditional ConvLSTM focuses on the extraction of temporal information, ignoring the importance of different time intervals is different for the subsequent time-series prediction. Therefore, we choose to introduce the attention mechanism into the traditional ConvLSTM as our temporal attention module.

Figure 9 illustrates the specific structure of the temporal attention module. It takes a series of spatially weighted traffic flow as input and then feeds them into ConvLSTM to obtain a series of outputs. Finally, this series of output is multiplied with the temporal attentive score to obtain the spatiotemporal weighted outputs.

Specifically, by feeding spatially weighted traffic flow $\{F_j^{i\in T}\}_{i=1}^T$ to ConvLSTM, we can obtain the series of outputs $\{H_j^{i\in T}\}_{i=1}^T$.
And then, we randomly initialized a series of vectors \( q_i \), in which \( q_i \) is the query vector of \( H_i \). With \( q_i \) and \( H_i \), we can get the corresponding attention score \( \alpha_i \). This step can be represented by the following equation:

\[
\alpha_i = \text{softmax}(s(H_i, q_i)) = \frac{\exp(s(H_i, q_i))}{\sum_{i=1}^{t-1} \exp(s(H_i, q_i))} \tag{8}
\]

Finally, the prediction \( P \) can be obtained by an operation of weighed sum \( \{\alpha_i\}_{i=1}^{t-1} \) and \( \{H_i\}_{i=1}^{t-1} \):

\[
P = \sum_{i=1}^{t-1} \alpha_i H_i. \tag{9}
\]

4.4. The Information Fusion Module. In the previous description, we believed that traffic flow forecasting is affected by periodicity and sequentiality in temporal dimensions. How accurately these two properties are weighed is important for the forecast performance. To address this issue, we introduce an information fusion module. Specifically, the structure of the information fusion module is shown in
Figure 10. This module can dynamically learn the weights of these two properties from external factors. These weights are used as fusion weights to fuse the information in two time dimensions and finally obtain the prediction results.

Specifically, we define the periodic prediction results and sequential prediction results as $P^p$, $P^s$, respectively. Through the information fusion module, we obtain the fusion weights of the two prediction results as $\omega^p$, $\omega^s$. The periodic and sequential predictions are fused according to the fusion weights to obtain the final prediction $P_f$, denoted as follows:

$$P_f = P^p\omega^p + P^s\omega^s.$$  \hfill (10)

5. Experiments

In this section, we verify the validity of CLSTAN on two publicly available datasets, TaxiBJ and BikeNYC. We will then describe the experiments in detail.

5.1. Dataset. In this study, we select two representative publicly available datasets for city-wide traffic forecasting, including the TaxiBJ dataset and the BikeNYC dataset. The two datasets are publicly accessible and different comparison algorithms can be fairly compared on the same dataset. The summary of TaxiBJ and BikeNYC is shown as follows.

TaxiBJ dataset: this dataset collected over 34,000 taxis’ GPS data and external factors for over 16 months from 2013 to 2016. External factors include holidays, temperature, weather, and wind speed. Specifically, the first fifteen months of data are divided into the training set and the remaining data are divided into the test set.

BikeNYC dataset: this dataset collected over 4,300 bicycles rental data and external factors from April to October 2014. The external factors include 20 types of holidays. Similar to the TaxiBJ dataset, specifically, the first 172 days of data are divided into the training set and the last 10 days of data are divided into the test set.

5.2. Evaluation Metric. For evaluating the model, we compare Root Mean Square Error (RMSE) between the baseline methods and our methods, which is calculated as follows:

$$RMSE = \sqrt{\frac{1}{Q} \sum_{i=1}^{Q} (Y_i - \bar{Y}_i)^2},$$  \hfill (11)

where $Y_i$ and $\bar{Y}_i$ denote the true and predicted values of traffic flow, respectively. $Q$ denotes the amount of samples employed for validation.

5.3. Methods for Comparison. HA: historical average method forecasts traffic flow by using historical average.

ARIMA: autoregressive integrated moving average (ARIMA) is a classical time-series forecasting model that combines moving average and autoregressive components to model time series.

SARIMA [51]: seasonal ARIMA is a variation of ARIMA, which considers seasonal effects.

VAR [52]: vector autoregression (VAR) is a classical random processing method that captures the linear relationship between several time series.

DeepST: DeepST is a model based on deep learning that is the first one to capture spatial information by convolution.

ST-ANN [53]: it extracts spatial features by the values of 8 adjacent regions and temporal features by the values of the prior 8 time intervals and uses the extracted spatiotemporal features for traffic flow forecasting.

ST-ResNet [53]: ST-ResNet is also a traffic forecasting method based on deep learning. This method combines density, period, trend data, and external factors for traffic flow forecasting.

VPN [54]: video pixel networks (VPN) are a model of probabilistic video for multiframe forecasting.

PredNet [55]: PredNet is a CNN-based method for modeling the dependencies among successive video inputs and subsequent frames.

PredRNN [56]: PredRNN is a method used to generate subsequent frames of video sequences by capturing spatiotemporal features of the input frames through recurrent neural networks.

5.4. Performance Comparison

5.4.1. Comparison with Baseline Methods. Table 1 shows the results of the experiment also demonstrate that our method improves the prediction accuracy and outperforms other methods.
5.4.2. Comparison with Different Variants of Our Model.
In our experiments, we conducted ablation experiments on our model to verify the effectiveness of different components on the TaxiBJ dataset and the BikeNYC dataset. Specifically, there are five types of our model and its variants:

- **CLSTAN**: the complete model we proposed.
- **CLSTAN + ConvGRU**: replace ConvLSTM in the CLSTAN model with ConvGRU.
- **CLSTAN-SA**: remove the spatial attention module in the CLSTAN model.
- **CLSTAN-TA**: remove the temporal attention module in the CLSTAN model.
- **CLSTAN-STA**: remove both spatiotemporal attention modules in the CLSTAN model.

We conducted experiments for the above five cases, and the results are shown in Table 2. Firstly, as can be seen in Table 2, CLSTAN with both spatiotemporal attention modules applied shows the best performance. Using both spatiotemporal attention modules, we can better capture the spatiotemporal relationship of traffic flow, thus improving the accuracy of forecasting. Then, using only the spatial attention module or only the temporal attention module, we can also reduce RMSE to some extent, which demonstrates the effectiveness of these two modules. Finally, as a variant of ConvLSTM, ConvGRU also achieves good results, but it is

---

**Table 1: Comparisons of RMSE on TaxiBJ and BikeNYC datasets.**

<table>
<thead>
<tr>
<th>Methods</th>
<th>BikeNYC</th>
<th>TaxiBJ</th>
</tr>
</thead>
<tbody>
<tr>
<td>HA</td>
<td>21.57</td>
<td>57.79</td>
</tr>
<tr>
<td>SARIMA</td>
<td>10.56</td>
<td>26.88</td>
</tr>
<tr>
<td>VAR</td>
<td>9.92</td>
<td>22.88</td>
</tr>
<tr>
<td>ARIMA</td>
<td>10.07</td>
<td>22.78</td>
</tr>
<tr>
<td>ST-ANN</td>
<td>—</td>
<td>19.57</td>
</tr>
<tr>
<td>DeepST</td>
<td>7.43</td>
<td>18.18</td>
</tr>
<tr>
<td>VPN</td>
<td>6.17</td>
<td>16.75</td>
</tr>
<tr>
<td>ST-ResNet</td>
<td>6.37</td>
<td>16.69</td>
</tr>
<tr>
<td>PredNet</td>
<td>7.45</td>
<td>16.68</td>
</tr>
<tr>
<td>PredRNN</td>
<td>5.99</td>
<td>16.34</td>
</tr>
<tr>
<td>CLSTAN (our)</td>
<td>5.65</td>
<td>15.23</td>
</tr>
</tbody>
</table>

---

**Figure 10: The structure of the information fusion module.**
slightly inferior to ConvLSTM. Therefore, we choose ConvLSTM as the main structure of the spatiotemporal attention module.

5.4.3. Comparison with Different Attentional Time Step Sizes.

Furthermore, we also investigated the effect of different attentional time steps on the final prediction results. In our experiments, we set from 0 time steps to 6 time steps, and the results of the experiment are presented in Figure 11.

Figures 11(a) and 11(b) show the prediction results of different attentional time steps on the two datasets. As can be seen from the figures, our proposed model obtains the best prediction performance when the attention time step size is set up to 6. Compared with the worst results, the RMSE of the model with an attentional time step of 6 decreases by 2.2% on BikeNYC and 1.2% on TaxiBJ.

Table 2: RMSE of different variants on TaxiBJ and BikeNYC dataset.

<table>
<thead>
<tr>
<th>Methods</th>
<th>BikeNYC</th>
<th>TaxiBJ</th>
</tr>
</thead>
<tbody>
<tr>
<td>CLSTAN</td>
<td>5.65</td>
<td>15.23</td>
</tr>
<tr>
<td>CLSTAN + ConvGRU</td>
<td>5.67</td>
<td>15.36</td>
</tr>
<tr>
<td>CLSTAN-STA</td>
<td>5.85</td>
<td>16.02</td>
</tr>
<tr>
<td>CLSTAN-SA</td>
<td>5.71</td>
<td>15.69</td>
</tr>
<tr>
<td>CLSTAN-TA</td>
<td>5.83</td>
<td>15.76</td>
</tr>
</tbody>
</table>

Figure 11: The RMSE of different attentional time step size: (a) experiment on BikeNYC dataset; (b) experiment on TaxiBJ dataset.

Figure 12: The RMSE of CLSTAN: (a) experiment on BikeNYC dataset; (b) experiment on TaxiBJ dataset.
4.88\% on the BikeNYC dataset and by 5.17\% on the TaxiBJ dataset. Therefore, we set the attentional time step sizes to 6.

5.4.4. Training Process. The change in RMSE of our model on the BikeNYC dataset and the TaxiBJ dataset is presented in Figure 12. As the amount of training epochs increases, the accuracy of the model steadily improves and eventually stabilizes. When the epoch approaches 150 in BikeNYC and 400 in TaxiBJ, the RMSE change decreases slowly and stabilizes. Therefore, we set the number of early stopping steps to 20 in the BikeNYC dataset and 50 in the TaxiBJ dataset to avoid the overfitting problem.

6. Conclusion and Discussion

We present an innovative spatiotemporal attention neural network based on ConvLSTM for traffic flow forecasting. Our approach focuses on designing a temporal attention module and a spatial attention module. These two modules dynamically capture the complicated spatiotemporal relationships within the traffic data to better forecast the future traffic flow. Specifically, the spatial attention module aims to explore those areas where future traffic flow changes will be more dramatic so that our model can focus more on these areas when predicting. And the temporal attention module aims to discover those time intervals that will have more impact on future traffic flow changes so that our model can focus more on those time intervals when predicting. We conducted experiments on both the TaxiBJ dataset and the BikeNYC dataset, and the experimental results showed that CLSTAN outperformed other baseline experimental methods. Furthermore, the ablation experiment once again validated the performance of the proposed spatiotemporal attention module.

In recent years, GCN-based methods have begun to be extensively employed in traffic flow forecasting, and they can better extract spatial relationships. In the future work, we will combine our model with GCN-based methods and design a suitable network to achieve more accurate prediction results.

Data Availability

The datasets for this research were obtained from the study “Deep Spatio-Temporal Residual Networks for Citywide Crowd Flows Prediction.”

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This work was supported by the National Natural Science Foundation of China (nos. 62067002, 61967006, and 62062033), the Science and Technology Project of Transportation Department of Jiangxi Province (nos. 2021X0011 and 2022X0040), and the Natural Science Foundation of Jiangxi Province (no. 2021BAB202008).

References


12 Mathematical Problems in Engineering


