Hindawi
Journal of Advanced Transportation
Volume 2020, Article ID 4656435, 12 pages
https://doi.org/10.1155/2020/4656435

Research Article

A Hybrid Spatiotemporal Deep Learning Model for Short-Term Metro Passenger Flow Prediction

Hao Zhang,[1,2,3,4] Jie He,[1,2,3] Jie Bao,[5] Qiong Hong,[6] and Xiaomeng Shi[1,2,3]

[1]*Jiangsu Key Laboratory of Urban ITS, Southeast University, Nanjing, China*
[2]*Jiangsu Province Collaborative Innovation Center of Modern Urban Traffic Technologies, Nanjing, China*
[3]*School of Transportation, Southeast University, 2 Dongnandaxue Rd., Nanjing, Jiangsu 211189, China*
[4]*School of Traffic Engineering, Huaiyin Institute of Technology, Meicheng East Road #1, Huaian, Jiangsu 223001, China*
[5]*Civil Aviation College, Nanjing University of Aeronautics and Astronautics, Jiangjun Road #29, Nanjing, Jiangsu 211106, China*
[6]*Business School, Huaian Vocational College of Information Technology, Meicheng East Road #3, Huaian, Jiangsu 223003, China*

Correspondence should be addressed to Jie He; hejie@seu.edu.cn

Received 8 December 2019; Accepted 25 February 2020; Published 30 May 2020

Academic Editor: Giulio E. Cantarella

The primary objective of this study is to predict the short-term metro passenger flow using the proposed hybrid spatiotemporal deep learning neural network (HSTDL-net). The metro passenger flow data is collected from line 2 of Nanjing metro system to illustrate the study procedure. A hybrid spatiotemporal deep learning model is developed to predict both inbound and outbound passenger flows for every 10 minutes. The results suggest that the proposed HSTDL-net achieves better prediction performance on suburban stations than on urban stations, as well as generating the best prediction accuracy on transfer stations in terms of the lowest MAPE value. Moreover, a comparative analysis is conducted to compare the performance of proposed HSTDL-net with other typical methods, such as ARIMA, MLP, CNN, LSTM, and GBRT. The results indicate that, for both inbound and outbound passenger flow predictions, the HSTDL-net outperforms all the compared models on three types of stations. The results suggest that the proposed hybrid spatiotemporal deep learning neural network can more effectively and fully discover both spatial and temporal hidden correlations between stations for short-term metro passenger flow prediction. The results of this study could provide insightful suggestions for metro system authorities to adjust the operation plans and enhance the service quality of the entire metro system.

1. Introduction

The metro system constitutes an important supplementation to urban transportation systems, providing travellers with sustainable, reliable, and efficient mobility, reducing the number of trips made by private vehicles, and leading to reduced traffic congestion and vehicle emissions in urban areas. However, due to the effects of fluctuating spatial and temporal travel demand, the metro system has suffered from a series of problems in recent years, such as overcrowded platforms and poor service levels [1]. Accordingly, an accurate passenger flow prediction has become an important component of metro transportation system. The prediction results can be applied to support metro system management such as operation planning and station passenger crowd regulation [2–5].

Over the past few decades, considerable efforts have been devoted to predicting the short-term metro passenger flow [2–4, 6–9]. Wang et al. proposed a support vector machine combined online model for short-term metro ridership forecasting. The results indicated that the proposed model could better capture the periodicity and nonlinearity characteristics of metro ridership with the collected data from Nanjing metro system [6]. Zhang and Liang used an improved Kalman filter model to forecast short-term passenger flow in Beijing metro system [9]. Hao et al. proposed a sequence to sequence model embedded with the attention mechanism to make the multistep and network-wide

predictions for short-term passenger flow. The developed model has shown great abilities to capture the long-range dependencies and achieved more scalable and robust prediction performance than other baselines [4].

Most previous studies have considered the short-term metro passenger flow prediction only as a typical time-series problem and failed to incorporate the hidden spatial correlations between stations to enhance the prediction accuracy [2, 7, 8]. For example, station #A and station #B may exhibit similar passenger flow patterns because they are both close to college campus (see Figure 1). In addition, station #C and station #D are less than 1 km apart and locate in the same business district and may both have high passenger flow during the same peak period. Considering the spatiotemporal nature, it is essential to integrate both spatial and temporal characteristics into the short-term metro prediction models, which has great potential for improving the prediction performance in practical applications [10].

During the past few years, some researchers have proposed various hybrid models for considering both spatial and temporal dependencies in metro passenger flow prediction [2, 7, 8, 11]. For example, Zhu et al. developed an ARIMA-Wavelet model for forecasting the daily passenger flow of Beijing urban rail transit [7]. Li et al. combined linear ARIMA model and nonlinear symbolic regression to model the complex relationship beneath the passenger flow dataset. The results suggest that the developed hybrid model outperforms the single model with the real dataset from Xi'an metro line 1 [8]. Wei and Chen proposed a hybrid EMD-BPN approach for short-term passenger flow prediction of Taipei metro system with three stages [2].

More recently, deep learning models have been widely used in various transportation researches because they have great power to extract the hidden nonlinear relationships with distributed and hierarchical feature representations [3, 12–15]. Liu et al. proposed an end-to-end deep learning framework based on long short-term memory neural network (LSTM) for short-term metro passenger flow prediction [3]. The results suggest that multiple LSTM layers could better capture the temporal dependencies in the passenger flow data and exhibit great prediction performances. However, the LSTM neural network can only capture the temporal dependencies but fail to extract the spatial correlations among different stations. To address this limitation, this study proposes a hybrid spatiotemporal deep learning model based on convolutional LSTM neural network for short-term metro passenger flow prediction. In recent years, the convolutional LSTM neural network has been used to solve various spatiotemporal transportation prediction problems, such as taxi demand forecasting [14], crash risk prediction [15], and bus travel time prediction [16]. To the best of the authors' knowledge, this paper is one of the first attempts to employ convolutional LSTM neural network for short-term metro passenger flow prediction. The main contributions of this paper could be summarized as follows.

(a) The proposed hybrid spatiotemporal deep learning neural network (HSTDL-net) has great abilities to learn both spatial and temporal dependencies among the passenger flows of metro stations

(b) The proposed hybrid spatiotemporal deep learning architecture could integrate both convolutional filters and recurrent component in one end-to-end structure, which could learn the spatiotemporal patterns of passenger flows more efficiently

(c) Validated by the real dataset provided by Nanjing metro system, the proposed hybrid spatiotemporal deep learning model outperforms the selected benchmark methods, including conventional time-series models and several state-of-the-art machine learning approaches

The rest of the paper is organized as follows. Section 2 discusses the methodology of the convolutional LSTM neural network and the structure of proposed HSTDL-net. Section 3 describes the data source. Section 4 presents the results of data analysis and compares the predictive performance between the proposed approach and the benchmark models. Finally, conclusions are drawn and future research directions are indicated in Section 5.

2. Methodology

In this section, we construct a hybrid spatiotemporal deep learning neural network for predicting short-term metro passenger flow. The proposed HSTDL-net integrates both LSTM neural network and convolutional LSTM neural network into an end-to-end deep learning architecture. The used methods are briefly discussed in this section.

2.1. The Structure of HSTDL-Net. In this study, two different types of variables are incorporated in the short-term metro passenger flow prediction. The first type of variables is both spatially and temporally varied during the study period, such as the inbound passenger flow and outbound passenger flow variables. There exist both spatial dependencies and temporal dependencies in this type of variables. The second type of variables is only temporally varied but spatially static during the study period, such as the weather variables and air quality variables. There exist strong periodicity and only temporal dependencies in this type of variables.

In this study, to capture the spatial and temporal dynamics of the two types of variables, we propose a hybrid spatiotemporal deep learning neural network (HSTDL-net) for predicting short-term metro passenger flow. Figure 2 illustrates the structure of the proposed HSTDL-net. More specifically, stacked convolutional LSTM layers are developed to capture the spatiotemporal features in the first type of variables. The stacked LSTM layers are developed to extract the temporal features in the second type of variables. The extracted high-level features from the two components are further merged together and input into multiple fully connected layers to generate the final predicted passenger flow value. The used methods in each component are briefly explained as follows.

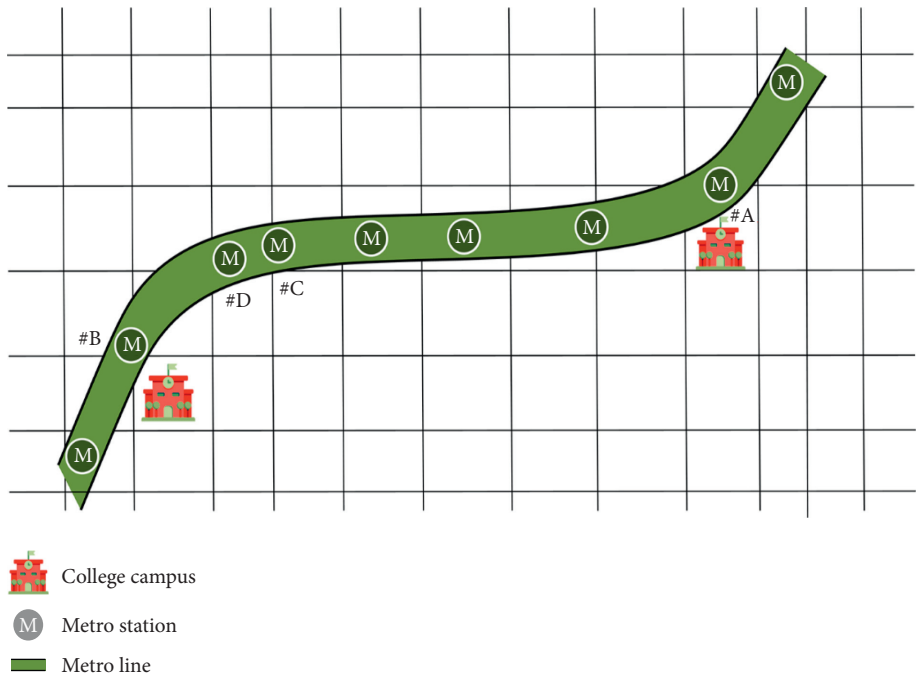

Figure 1: Illustrations of hidden spatial correlations between metro stations.

2.2. Long Short-Term Memory (LSTM) Neural Network. Long short-term memory (LSTM) neural network is a specific type of recurrent neural network (RNN), which has exhibited great performance in forecasting time-series datasets [13, 17, 18]. For example, Ma et al. have attempted to use LSTM neural network for capturing the temporal patterns in short-term traffic speed data with the collected remote microwave sensor data [13]. Wollmer et al. have applied LSTM neural network to online driver distraction detection with the driving and head tracking data [18]. The LSTM neural network could address the issues of gradient exploding and gradient vanishing, which are very prevalent in traditional RNN with large prediction time step [17]. The most important components in LSTM neural network are the three kinds of gates in the memory cell of hidden layer (see Figure 3). Specifically, the forget gate is designed to eliminate information that is not related to the prediction task. The input gate controls the information that can be considered in the prediction task. For time step t, the input gate $\boldsymbol{i}_t$, forget gate $\boldsymbol{f}_t$, and output gate $\boldsymbol{o}_t$ are calculated iteratively in the following ways ((1)–(5)):

$$i_t = \sigma\left(W_{xi}x_t + W_{hi}h_{t-1} + W_{ci} \odot c_{t-1} + bi\right), \tag{1}$$

$$f_t = \sigma\left(W_{xf}x_t + W_{hf}h_{t-1} + W_{cf} \odot c_{t-1} + b_f\right), \tag{2}$$

$$c_t = f_t \odot c_{t-1} + i_t \odot \tanh\left(W_{xc}x_t + W_{hc}h_{t-1} + b_c\right), \tag{3}$$

$$o_t = \sigma\left(W_{xo}x_t + W_{ho}h_{t-1} + W_{co} \odot c_t + b_o\right), \tag{4}$$

$$h_t = o_t \odot \tanh(ct), \tag{5}$$

where $\boldsymbol{x}_t$ indicates the second type of variables mentioned in Section 2.1, which are only temporally varied but spatially static. Specifically, in this study $\boldsymbol{x}_t$ represents the input historical weather variables in each time step such as temperature, precipitation, wind speed, and pressure; $\boldsymbol{c}_t$ indicates the activation vectors for each cell; $\boldsymbol{h}_t$ indicates the related predicted value. $\boldsymbol{W}_{xi}$ indicates the weight matrix between input weather variable and the output of input gate. $\boldsymbol{b}_i$ indicates the bias value of input gate. Similarly, $\boldsymbol{W}_{hi}$, $\boldsymbol{W}_{ci}$, $\boldsymbol{W}_{xf}$, $\boldsymbol{W}_{hf}$, $\boldsymbol{W}_{cf}$, $\boldsymbol{W}_{xc}$, $\boldsymbol{W}_{hc}$, $\boldsymbol{W}_{xo}$, $\boldsymbol{W}_{ho}$, and $\boldsymbol{W}_{co}$ indicate the weight matrixes connecting the vector of the first subscript to the vector of the second subscript. $\boldsymbol{b}_c$, $\boldsymbol{b}_f$, and $\boldsymbol{b}_o$ indicate the relevant bias values. $\odot$ represents the element-wise product between weights' matrix and bias matrix. σ and tan h represent the active functions in LSTM neural network with the following forms:

$$\sigma(x) = \frac{1}{1 + e^{-x}}, \qquad \tanh(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}}. \tag{6}$$

In many previous studies, multiple LSTM neural networks were structured in a stack form such that more complex temporal dependencies among variables could be learned and the prediction accuracy could be further improved [14, 15].

2.3. Convolutional LSTM Neural Network. In this study, the metro passenger flow not only exhibits significant temporal patterns for each station, but also shows great spatial patterns across different stations. For example, two stations with similar surrounding land use may exhibit similar passenger flow patterns. Due to this spatiotemporal nature, the standard LSTM neural network is not an ideal model for short-term

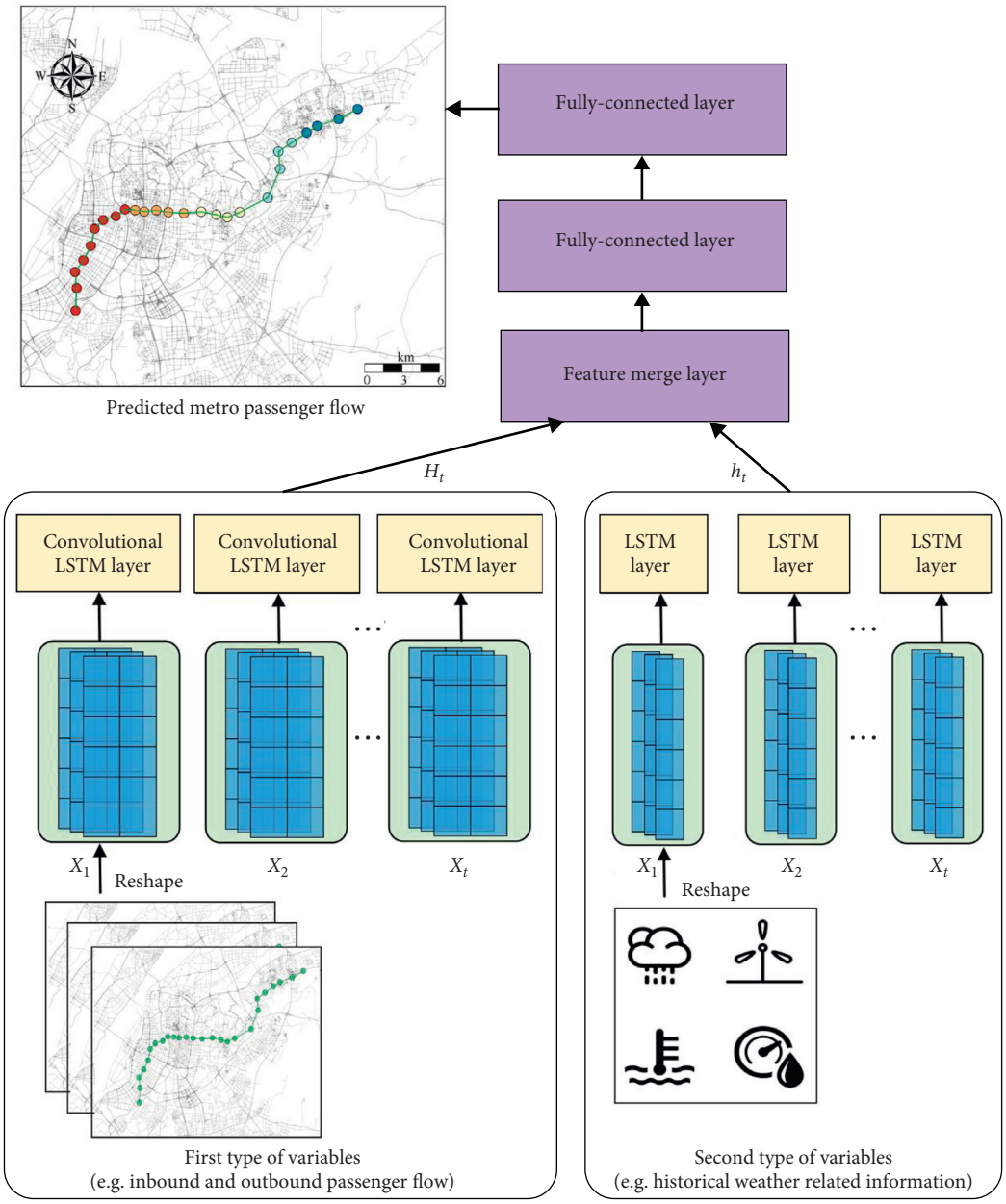

Figure 2: The structure of the proposed HSTDL-net.

metro passenger flow prediction because it cannot learn the spatial dependencies among variables [14, 16]. To overcome this limitation, Shi et al. innovatively combined the convolutional layers and LSTM layers into an end-to-end deep learning structure, the convolutional LSTM, which could model both spatial and temporal characteristics simultaneously [19]. The most important feature of convolutional LSTM neural network is the convolution operation between neighboring LSTM cells. Specifically, all the inputs, hidden states, and outputs of various gates are transformed to 3D tensors in convolutional LSTM neural network (see Figure 4(a)). Then, the convolutional filters are applied to the input-to-state and state-to-state transitions for a certain passenger flow grid cell (see Figure 4(b)).

For each time step t, the input gate I_t, forget gate F_t, and output gate O_t work in a similar way as the standard LSTM neural network ((7)–(11)):

$$I_t = \sigma\left(W_{xi} * X_t + W_{hi} * H_{t-1} + W_{ci} \odot C_{t-1} + b_i\right), \quad (7)$$

$$F_t = \sigma\left(W_{xf} * X_t + W_{hf} * H_{t-1} + W_{cf} \odot C_{t-1} + b_f\right), \quad (8)$$

$$C_t = F_t \odot C_{t-1} + \mathbf{I}_t \odot \tanh\left(W_{xc} * X_t + W_{hc} * H_{t-1} + b_c\right), \quad (9)$$

$$O_t = \sigma\left(W_{xo} * X_t + W_{ho} * H_{t-1} + W_{co} \odot C_t + b_o\right), \quad (10)$$

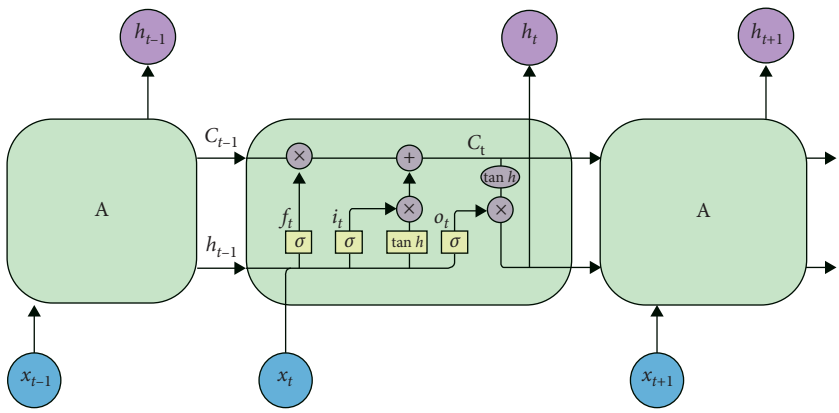

Figure 3: Illustrations of long short-term memory neural network.

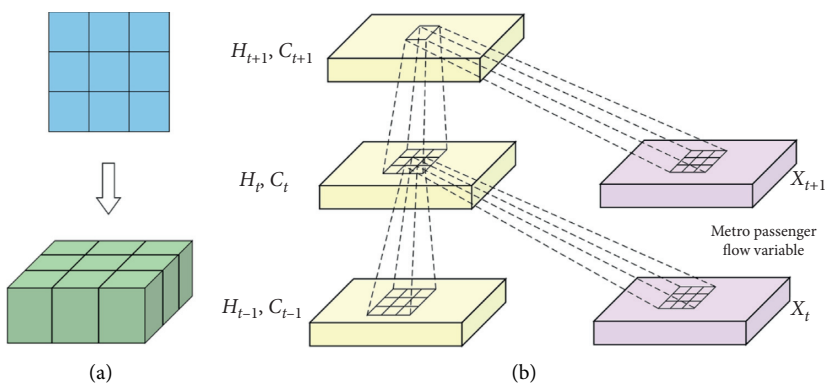

Figure 4: Illustrations of convolutional LSTM neural network. (a) The transformation from 2D matrix to 3D sensor. (b) The structure of convolutional LSTM neural network.

$$H_t = O_t \odot \tan h(C_t), \tag{11}$$

where the operator * refers to the convolution operator, which is the greatest difference between convolutional LSTM and standard LSTM neural network. X_t refers to the tensor of input metro passenger flow for time step t. X_t is a 3D tensor and can be regarded as a temporal stack of the passenger flow map for all the stations. I_t, H_t, C_t, and F_t refer to the input gate tensor, hidden tensor, cell output tensor, and forget gate tensor, respectively. Here, $\boldsymbol{W}_{xi}$, $\boldsymbol{W}_{hi}$, $\boldsymbol{W}_{xf}$, $\boldsymbol{W}_{hf}$, $\boldsymbol{W}_{xc}$, $\boldsymbol{W}_{hc}$, $\boldsymbol{W}_{xo}$, and $\boldsymbol{W}_{ho}$ serve as convolutional filters, which are replicated across the tensors with shared weights, and thus explore the spatial local correlations. Zero padding technique is applied to ensure a consistent spatial dimension during the convolutional operation.

The recurrent structure and the convolution component in convolutional LSTM neural network make both spatial and temporal patterns of short-term metro passenger flow which could be better learned. Similar to standard LSTM neural network, multiple convolutional LSTM neural networks could also be structured in a stack form for building a more robust and accurate predictor. In this study, the input metro passenger flow of all the stations for each time step is first reshaped to a tensor structure $X = (X_1, X_2, \ldots, X_t)$ (see Figure 4). Then, through a two-layer convolutional LSTM neural network, the input metro passenger flow tensors are mapped to a sequence of hidden tensors $H = (H_1, H_2, \ldots, H_t)$.

2.4. Feature Merge Layer and Fully Connected Layer. The temporal features captured from LSTM neural network and the spatiotemporal features captured from convolutional LSTM neural network are then concatenated into a dense vector in the feature merge layer. Finally, the dense vector is input into several fully connected layers to obtain the final predicted value of metro passenger flow.

$$\hat{y}_t = W_{\text{LSTM}} X_t^{\text{LSTM}} + W_{\text{ConvLSTM}} X_t^{\text{ConvLSTM}} + b_t, \tag{12}$$

where X_t^{LSTM} and X_t^{ConvLSTM} indicate the extracted features by LSTM and convolutional LSTM layers at t time step, respectively. W_{LSTM}, W_{ConvLSTM}, and b_t indicate the related weights and bias.

2.5. Objective Function. The objective of the short-term metro passenger flow prediction model is to minimize the error between the real and predicted passenger flows for each metro station at every time step. During the whole training process, the objective function is given as follows:

$$\mathrm{RMSE} = \sqrt{\frac{1}{n_p}\sum_{i=1}^{n}\sum_{j=1}^{m}\left(y_{(i,j)} - \hat{y}_{(i,j)}\right)^2}, \tag{13}$$

where $y_{(i,\ j)}$ and $\hat{y}_{(i,\ j)}$ stand for the real and predicted passenger flow for metro station i at time step j, respectively. n refers to the number of metro stations and m refers to the number of total predicted time steps, and $n_p = n \times m$.

2.6. Evaluation Metrics. Three statistics are used to evaluate the performance of the proposed metro passenger flow prediction model, including root mean squared error (RMSE), mean absolute error (MAE), and mean absolute percentage error (MAPE), as formalized in (13)–(15).

$$\mathrm{MAE} = \frac{1}{n_p}\sum_{i=1}^{n}\sum_{j=1}^{m}\left|y_{(i,j)} - \hat{y}_{(i,j)}\right|, \tag{14}$$

$$\mathrm{MAPE} = \frac{1}{n_p}\sum_{i=1}^{n}\sum_{j=1}^{m}\frac{\left|y_{(i,j)} - \hat{y}_{(i,j)}\right|}{y_{(i,j)}}. \tag{15}$$

3. Data Source

In this study, the metro passenger flow dataset is collected from Nanjing metro line 2, which is maintained by the Nanjing Metro Corporation (NMC). Nanjing city is the economic and cultural center of Eastern China with more than 8 million people [20]. The Nanjing metro network began in 2005 and has expanded rapidly over the past decade due to the growth of urban travel demand. The metro line 2 was completed and operated in 2010, connecting the city's east and west areas [21]. As is shown in Figure 5, Nanjing metro line 2 has 26 stations with a total length of 38 km, including 2 terminal stations, 2 transfer stations, and 22 regular stations.

The Nanjing metro line 2 is available from 6:00 am to 11:00 pm. In the Nanjing metro system, passengers need to tap in and tap out smart cards when they enter and exit the station. Thus, each trip record provides the following information: smart card id, card type, total amount, the timestamp of enter or exit, and the station name. In this study, the short-term metro passenger flow is defined as the number of passengers that enter into a station for every 10 minutes. The study period covers a total of 66 weekdays from July 1st to September 30th, 2014. For the purpose of model training, the first 50 days are split as the training dataset and the remaining 16 days are used for test. The weather variables are collected from the Nanjing Meteorological Bureau which provides hourly aggregated weather information through local meteorological stations. The obtained weather variables include the hourly aggregated temperature, precipitation, wind speed, and pressure. The considered variables in the present study are described in Table 1.

Figure 6 further illustrates the temporal distribution of daily passenger flows for three typical station types. It can be found that the passenger flow patterns of different types of metro stations vary greatly. For the regular stations, their passenger flows show an obvious morning peak and evening peak period. For the terminal stations, their passenger flows only show a morning peak. By contrast, the passenger flows of transfer stations remain relatively high in the daytime and may still have a rapid growth around late-night time period. In general, the passenger flow patterns have great diversity among various types of metro stations, leading great challenges to the short-term metro passenger flow prediction.

4. Results of Data Analysis

4.1. Model Construction. A series of parameters should be determined in the process of model construction. In this study, uniform random search is applied to select the optimized value. More specifically, the input passenger flow of 26 metro stations was distributed in the form of a grid map with a length of 2 and a width of 13. For the convolution filter, the filter size was set as (2×2), which is the common setting in many previous studies [9, 22, 23]. The filter length was optimized from 10 to 40 (see Table 2). For the recurrent component, the number of time steps and the number of hidden units in the LSTM cell were chosen by the results of uniform random search. For the component of optimizer, four widely used optimizers were compared, including Adam, Nadam, RMSprop, and SGD [24]. The RMSprop was selected as the best one, and the learning rate was set from 0.001 to 0.01. For the training related parameters, the batch size was set from 20 to 80, and the number of training epochs was set from 50 to 200. In addition, to address the overfitting issue, the dropout layer was applied in this study [25]. The uniform random search interval and optimal parameter values are listed in Table 2.

The loss function of the training model is the root mean square error (RMSE). All the input variables were standardized by min-max normalization before conducting the model construction. The proposed convolutional LSTM model was implemented on the basis of Keras framework with TensorFlow as backend [26]. All the experiments were conducted with Python 3.5.2 in a Windows 10 system. To accelerate the efficiency of model training, a GTX 1060 Graphics Processing Card is used. The results of the optimal parameters are shown in Table 2.

4.2. Results of the Proposed HSTDL-Net. In order to compare the model performances for different types of metro stations, four test datasets were predicted by the trained HSTDL-net. As is shown in Tables 3 and 4, the prediction results indicate that models of transfer stations usually generate the highest RMSE and MAE values for both inbound and outbound passenger flows. This is expected because transfer stations have large passenger volume and may result in larger predicted residuals. Moreover, the models of terminal stations show the worst performances in terms of the highest MAPE values for both inbound and outbound passenger flows. The passenger flow patterns of terminal stations are more unstable and complex, which brings great challenges to the prediction task.

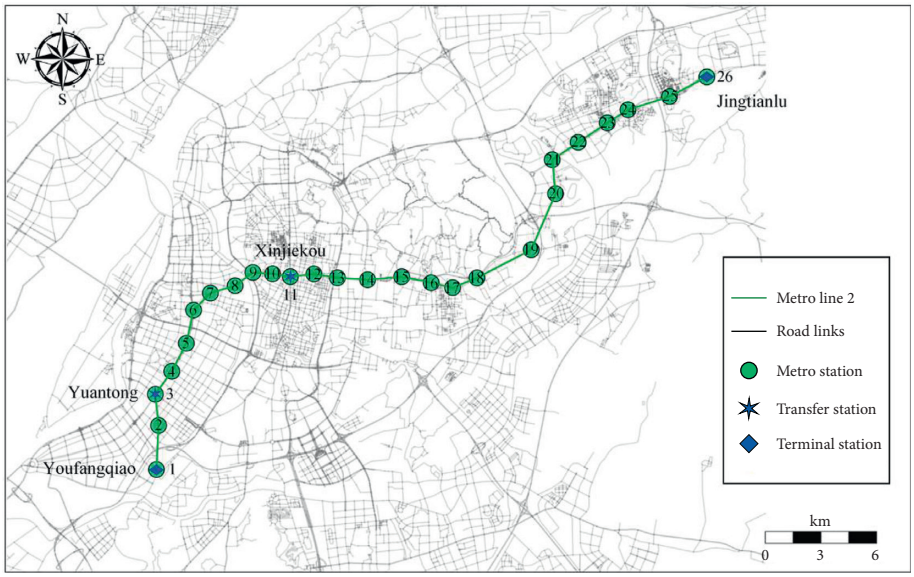

Figure 5: The study Nanjing metro line 2.

Table 1: Description statistics of considered variables.

Variables	Description	Min	Max	Mean	S.D.
Metro flow inbound	Number of passengers entering into station every 10 minutes	0	1020	83.16	188.62
Metro flow outbound	Number of passengers exiting from station every 10 minutes	0	1223	85.37	194.57
Temperature	The hourly average temperature during the time interval (°C)	15.3	37.3	24.321	2.935
Precipitation	The hourly average precipitation during the time interval (mm)	0	49.2	0.075	1.126
Wind speed	The hourly average wind speed during the time interval (m/s)	0	14.3	3.207	1.961
Pressure	The hourly average pressure during the time interval (hPa)	1000	1010.2	1005.21	2.102

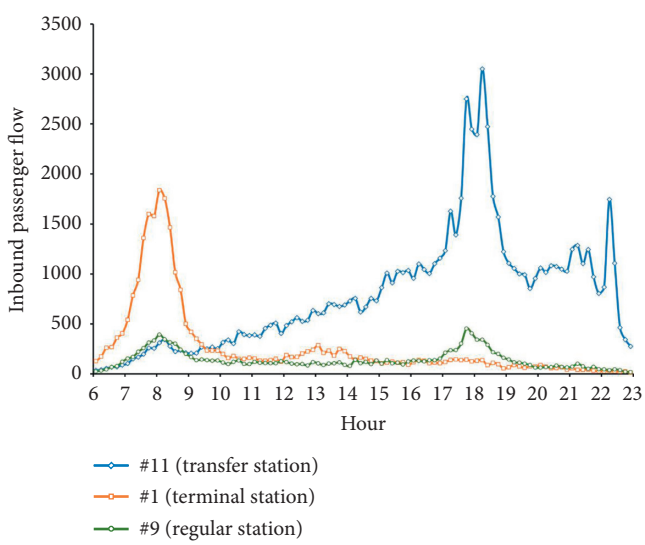

Figure 6: The temporal distribution of passenger flow for three typical station types.

Figure 7 further depicts the spatial distribution of predicted residuals for all the stations of Nanjing metro line 2. The residual errors of suburban stations are generally lower than those of urban stations. The suburban stations exhibit better prediction performance because they are mainly located around residual places with commuting as the primary trip purpose and accordingly exhibit a more stable passenger flow pattern. In addition, it can be found that stations near large shopping centers, such as Xinjiekou and Shanghailu station, show a relatively poor performance in terms of higher residuals. The reasons are mainly twofold. First, stations located in the central business areas may be also transfer stations involving multiple metro lines, resulting in greater fluctuations of passenger flow. Second, travels for shopping and entertainment are more flexible and irregular than commuting, making passenger flow prediction of these stations become more difficult.

4.3. Results of Model Comparison. The proposed HSTDL-net is further compared with other prevalent prediction models using the same dataset. Specifically, CNN, LSTM, ARIMA, multilayer perceptron (MLP), and the gradient boosting regression tree (GBRT) are selected as the benchmark models in this study. The selected models include statistical method, machine learning method, ensemble tree based method, and deep learning method, which ensure a fair and comprehensive comparison. Moreover, these selected methods have also been widely applied in many previous studies of short-term passenger flow prediction [3, 7, 8, 11, 27].

Table 2: Parameter settings in the model construction.

Parameter	Description	Uniform random search interval	Optimal value
Convolutional filter			
L	The length of convolution filter	(10, 40)	22
$(a \times b)$	The size of convolution filter	—	(2×2)
Recurrent component			
T	The number of predicted time steps in recurrent block	(5, 20)	6
H	The number of hidden units in the cells	(64, 128)	128
Optimizer			
O	The selected optimizer during model training	(Adam, Nadam, RMSprop, and SGD)	RMSprop
α	The learning rate	(0.001, 0.01)	0.01
Training setting			
D	The dropout rate	(0.2, 0.4)	0.2
B	The batch size for each training epoch	(20, 80)	35
E	The number of training epochs	(50, 200)	180

Table 3: Performance of the proposed HSTDL-net for inbound passenger flow.

	Total stations	Terminal stations	Transfer stations	Regular stations
RMSE	23.227	32.246	52.794	18.264
MAE	14.172	18.782	25.237	12.167
MAPE (%)	23.6	31.6	16.2	22.1

Table 4: Performance of the proposed HSTDL-net for outbound passenger flow.

	Total stations	Terminal stations	Transfer stations	Regular stations
RMSE	26.268	34.368	55.176	20.005
MAE	16.825	21.794	26.301	14.711
MAPE (%)	25.7	33.6	18.2	23.7

ARIMA is the most conventional regression method for modeling time-series dataset. ARIMA method integrates the moving average model, the autoregressive component, and the moving average part [28]. MLP is a typical architecture of feedforward neural network, which consists of multiple fully connected layers. The hidden layers in MLP could capture the complex nonlinear relationships in the time-series dataset [23]. CNN is an emerging hot method, which has achieved great successes in the fields of image recognition and signal processing [22, 23]. More recently, many researchers have also applied it for solving various transportation problems [14, 15, 29]. The most important feature of CNN is that the neurons are connected to the preceding layer through a moving patch with the same weight values. Thus, CNN based methods are good at extracting the spatial dependencies among predicted variables. LSTM is a particular type of recurrent neural network, which could account for the gradient exploding and gradient vanishing problems [17, 30]. LSTM has been adopted to forecast the short-term passenger flow in previous studies [3, 18]. The mechanism has been discussed in previous section. GBRT is built on the basis of the core idea of ensemble tree, which aims to improve the model performance and model robustness by integrating the prediction results from multiple week regression trees [27]. After achieving great successes in other fields, GBRT has become more and more prevalent in many transportation studies [27, 31].

To ensure a fair comparison, all the compared models are fine-tuned under the same training dataset and number of training epochs. However, the structure of the input data for HSTDL-net cannot be directly applied to other compared models. Some data processing work should be conducted to satisfy the requirements of model input. Specifically, for the ARIMA, MLP, and GBRT models, the input passenger flow of all stations in the past T time steps is reshaped as a matrix in the form of (*batch size*, T). For CNN model, the input passenger flow data are reshaped as a 4D tensor in the form of (*batch size*, 2, 13, T). T indicates the channel of the image and (2, 13) indicates the size of the image. For LSTM model, the input passenger flow data are reshaped as a 3D tensor in the form of (*batch size*, T, 1).

Figure 8 visualizes the results of model comparisons in terms of MAPE value. Overall, for both inbound and outbound passenger flow predictions, all the models achieve the best performance on transfer stations but show the worst performance on terminal stations. In addition, the HSTDL-net has the lowest MAPE values compared to the other models, indicating that the proposed model could fully capture both spatial and temporal dependencies among passenger flow variables. Tables 5 and 6 further list the prediction results of all compared models. It can be found that ARIMA model achieves the lowest prediction accuracy in terms of the highest RMSE, MAE, and MAPE values for all the three types of stations. Moreover, for regular stations the

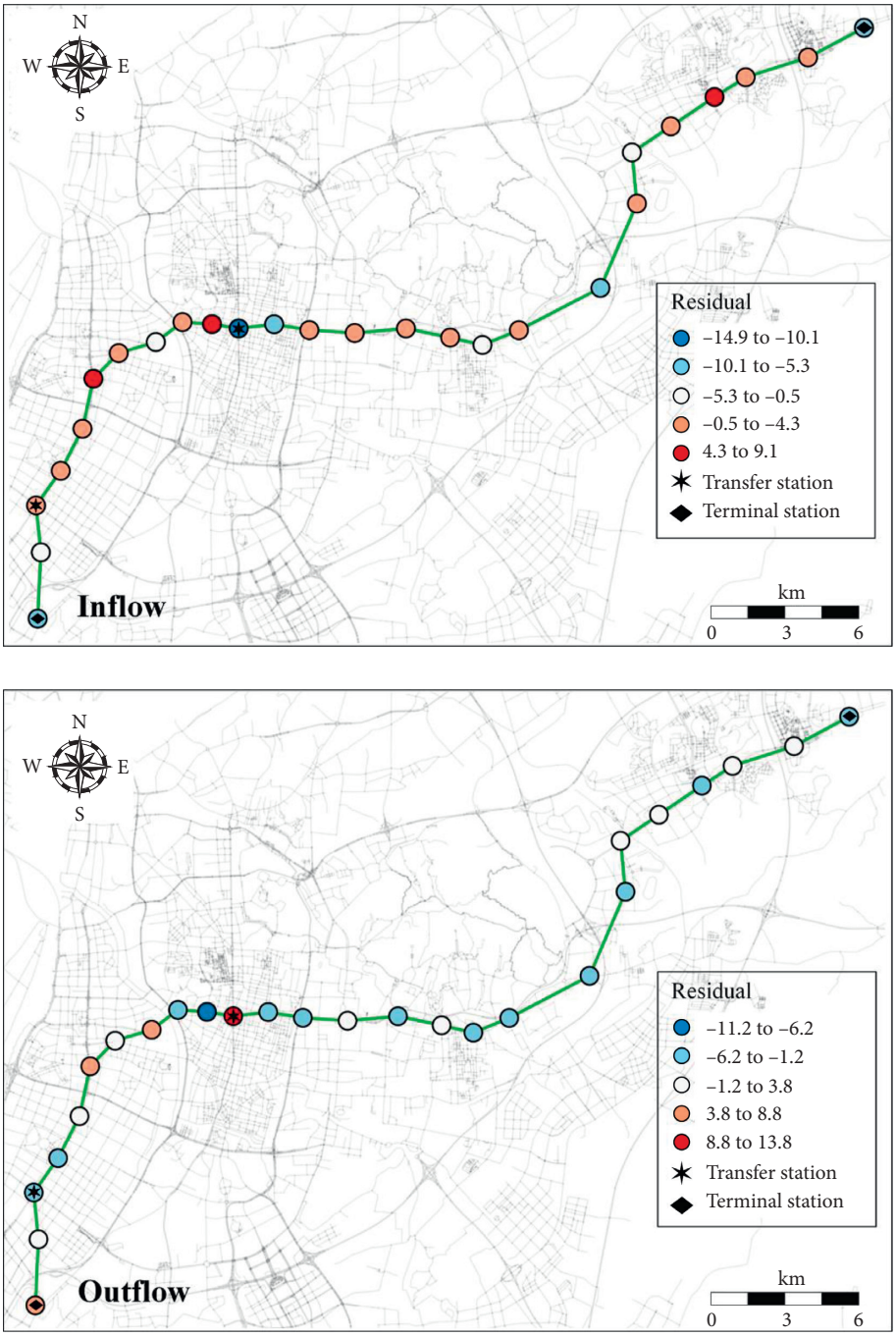

Figure 7: The spatial distribution of prediction errors for each station.

MAPE values of MLP model are both more than 40% for inbound and outbound passenger flow prediction, respectively. The compared results indicate that the two simple models have relatively poor performances in short-term metro passenger flow prediction. By contrast, the MAPE values of CNN and LSTM are both less than 40% for regular stations, which indicates that these two models could better capture the spatial dependencies or temporal dependencies among passenger flow variables.

Compared with CNN and LSTM models, the MAPE values of GBRT models for regular stations are only 27.4% and 29.0% for inbound and outbound passenger flow prediction. The GBRT model can greatly improve the prediction performance because the integration of numerous regression trees can fully capture the complex nonlinear relationships among passenger flow variables. In general, the proposed HSTDL-net achieves the best prediction performance for all three types of stations. The HSTDL-net is the only model with MAPE values less than 35% for both inbound and outbound passenger flow predictions. More specifically, for inbound passenger flow prediction, the MAE values of HSTDL-net on terminal stations have decreased by 7.17%, 33.92%, 29.57%, 43.42%, and 63.31% for each

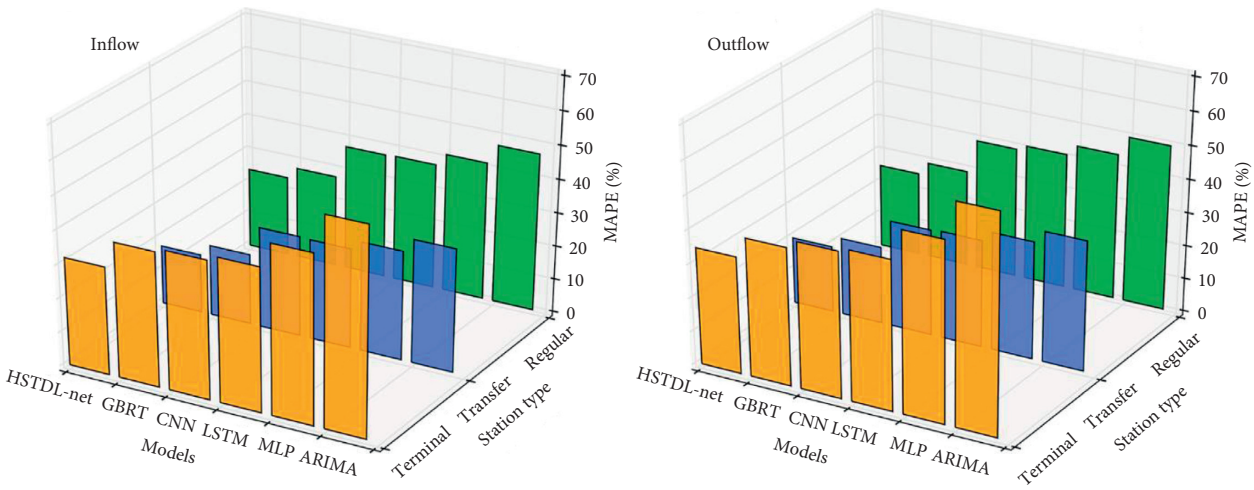

Figure 8: Visualization of model comparison results.

Table 5: The comparison of different models (inbound passenger flow).

Model	Terminal stations			Transfer stations			Regular stations		
	RMSE	MAE	MAPE (%)	RMSE	MAE	MAPE (%)	RMSE	MAE	MAPE (%)
HSTDL	**32.246**	**18.782**	**31.6**	**52.794**	**25.237**	**16.2**	**18.264**	**12.167**	**22.1**
GBRT	38.39	20.233	39.9	78.106	43.088	21.1	25.678	16.077	27.4
CNN	61.347	28.421	41.0	89.013	44.962	29.9	37.907	20.022	37.2
LSTM	49.951	26.667	42.6	85.329	43.066	29.7	35.913	19.995	37.5
MLP	63.998	33.195	50.0	98.637	47.86	32.5	50.338	25.531	41.2
ARIMA	77.124	51.186	61.3	125.434	56.093	36.6	60.691	36.057	47.0

Table 6: The comparison of different models (outbound passenger flow).

Model	Terminal stations			Transfer stations			Regular stations		
	RMSE	MAE	MAPE (%)	RMSE	MAE	MAPE (%)	RMSE	MAE	MAPE (%)
HSTDL	**34.368**	**21.794**	**33.6**	**55.176**	**26.301**	**18.2**	**20.005**	**14.711**	**23.7**
GBRT	39.778	24.563	41.0	81.581	45.176	23.2	27.521	18.077	29.0
CNN	64.037	30.055	43.5	91.401	46.753	31.8	39.956	22.364	39.1
LSTM	51.882	27.418	44.5	88.553	44.977	32.1	37.726	21.022	39.6
MLP	65.037	36.833	53.7	99.441	49.841	35.1	53.071	28.111	43.8
ARIMA	79.998	54.011	64.9	127.112	57.051	38.9	63.339	39.126	49.5

compared model (GBRT, CNN, LSTM, MLP, and ARIMA), respectively. The MAE values of HSTDL-net on transfer stations have decreased by 41.43%, 43.87%, 41.40%, 47.27%, and 55.01% for each compared model, respectively. The MAE values of HSTDL-net on regular stations have decreased by 24.32%, 39.23%, 39.15%, 52.34%, and 66.26% for each compared model, respectively. The results of comparative analysis indicate that the proposed HSTDL-net can more effectively and fully discover both spatial and temporal hidden correlations between stations for short-term metro passenger flow prediction.

5. Conclusions and Discussions

The present study investigates the short-term metro passenger flow prediction with the advanced deep learning technology. A total of three-month trip records are obtained from the metro line 2 of Nanjing metro system. The passenger flow patterns have great diversity across different types of metro stations, which have led to great difficulties in short-term passenger flow prediction. In this study, a hybrid spatiotemporal deep learning neural network (HSTDL-net) is proposed for predicting both inbound and outbound passenger flows on three types of stations for every 10 minutes. The developed HSTDL-net innovatively incorporates the convolution operation between LSTM cells, which could capture both spatial and temporal dependencies among passenger flow variables.

In general, the proposed HSTDL-net achieves greater prediction performance on suburban stations than on urban stations. For each metro station type, the model of transfer stations exhibits the best prediction accuracy and the model of terminal stations performs the worst. Moreover, a comparative analysis is conducted to compare the prediction performance between the proposed HSTDL-net and other

typical models, such as ARIMA, MLP, CNN, LSTM, and GBRT. The results suggest that, for both inbound and outbound passenger flow predictions, the proposed HSTDL-net outperforms all the compared models on three types of stations. More specifically, for inbound passenger flow prediction, the MAE values of HSTDL-net on terminal stations have decreased by 7.17%, 33.92%, 29.57%, 43.42%, and 63.31% for each compared model (GBRT, CNN, LSTM, MLP, and ARIMA), respectively. The MAE values of HSTDL-net on transfer stations have decreased by 41.43%, 43.87%, 41.40%, 47.27%, and 55.01% for each compared model, respectively. The MAE values of HSTDL-net on regular stations have decreased by 24.32%, 39.23%, 39.15%, 52.34%, and 66.26% for each compared model, respectively.

The results suggest that the proposed HSTDL-net can more effectively and fully discover both spatial and temporal hidden correlations between stations for short-term metro passenger flow prediction. The results of short-term passenger flow prediction could provide insightful suggestions for decision makers of metro systems. More specifically, the accurate prediction results can help the metro system authorities to dynamically modify the operation plans according to the fluctuation of passenger flow, such as adjusting the headway and train dispatching schedule to ensure the service quality of the entire metro system. In addition, with the predicted passenger flow in the next multiple time steps, the passenger congestion can be identified in advance such that timely crowd regulation plan and emergency response plan can be made. For example, the metro system management can assign extra trains and add more volunteers for relieving the passenger congestion and improving the service level of metro system.

This study is the first step towards exploring the short-term passenger flow prediction with advanced deep learning technology. Our future work will focus on incorporating other data sources into short-term metro passenger flow prediction. For example, the real-time weather information and the land use pattern surrounding each metro station may show close relationships with the passenger flow pattern. Moreover, the proposed HSTDL-net has achieved great performance on a single metro line. However, the model performance on an entire metro network with multiple metro lines should be further tested. The authors recommend future studies that could focus on these issues.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon the request.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

Acknowledgments

This research was supported by the National Natural Science Foundation of China (Grant nos. 71874067 and 51778141), the "Six Industry Talent Peak Project" fund of Jiangsu Province (Grant nos. RJFW-049 and JNHB-115), the "Green and Blue Project" fund of Jiangsu Province (Grant no. 2017SJB1641), the philosophy and social science fund for colleges and universities in Jiangsu Province (Grant no. 2018SJA1649), the Natural Science Fund for Colleges and Universities in Jiangsu Province (Grant no. 17KJB58001), and the fund of Huaian Vocational College of Information Technology (Grant no. HXYC2017003).

References

[1] H. Niu and X. Zhou, "Optimizing urban rail timetable under time-dependent demand and oversaturated conditions," *Transportation Research Part C: Emerging Technologies*, vol. 36, pp. 212–230, 2013.

[2] Y. Wei and M.-C. Chen, "Forecasting the short-term metro passenger flow with empirical mode decomposition and neural networks," *Transportation Research Part C: Emerging Technologies*, vol. 21, no. 1, pp. 148–162, 2012.

[3] Y. Liu, Z. Liu, and R. Jia, "DeepPF: a deep learning based architecture for metro passenger flow prediction," *Transportation Research Part C: Emerging Technologies*, vol. 101, pp. 18–34, 2019.

[4] S. Hao, D.-H. Lee, and D. Zhao, "Sequence to sequence learning with attention mechanism for short-term passenger flow prediction in large-scale metro system," *Transportation Research Part C: Emerging Technologies*, vol. 107, pp. 287–300, 2019.

[5] J. Wu, M. Liu, H. Sun, T. Li, Z. Gao, and D. Z. W. Wang, "Equity-based timetable synchronization optimization in urban subway network," *Transportation Research Part C: Emerging Technologies*, vol. 51, pp. 1–18, 2015.

[6] X. Wang, N. Zhang, Y. Zhang, and Z. Shi, "Forecasting of short-term metro ridership with support vector machine online model," *Journal of Advanced Transportation*, vol. 2018, Article ID 3189238, 13 pages, 2018.

[7] J. Zhu, W.-x. Xu, H.-t. Jin, and H. Sun, "Prediction of urban rail traffic flow based on multiply wavelet-ARIMA model," in *Green Intelligent Transportation Systems*, pp. 561–576, Springer, Berlin, Germany, 2018.

[8] L. Li, Y. Wang, G. Zhong, J. Zhang, and B. Ran, "Short-to-medium term passenger flow forecasting for metro stations using a hybrid model," *KSCE Journal of Civil Engineering*, vol. 22, no. 5, pp. 1937–1945, 2018.

[9] Z. Zhang and T. Liang, "Short-term forecasting of passenger flow on the metro platform using an improved Kalman filtering method," in *Proceedings of the 19th COTA International Conference of Transportation Professionals: Cictp 2019*, pp. 2789–2801, Nanjing, China, July 2019.

[10] L. Lin, Z. He, and S. Peeta, "Predicting station-level hourly demand in a large-scale bike-sharing network: a graph convolutional neural network approach," *Transportation Research Part C: Emerging Technologies*, vol. 97, pp. 258–276, 2018.

[11] X. Ma, J. Zhang, B. Du, C. Ding, and L. Sun, "Parallel architecture of convolutional bi-directional LSTM neural networks for network-wide metro ridership prediction," *IEEE Transactions on Intelligent Transportation Systems*, vol. 20, no. 6, pp. 2278–2288, 2019.

[12] Y. Lv, Y. Duan, W. Kang, Z. Li, and F. Wang, "Traffic flow prediction with big data: a deep learning approach," *IEEE Transactions on Intelligent Transportation Systems*, vol. 16, no. 2, pp. 865–873, 2015.

[13] 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.

[14] J. Ke, H. Zheng, H. Yang, and X. Chen, "Short-term forecasting of passenger demand under on-demand ride services: a spatio-temporal deep learning approach," *Transportation Research Part C: Emerging Technologies*, vol. 85, pp. 591–608, 2017.

[15] J. Bao, P. Liu, and S. V. Ukkusuri, "A spatiotemporal deep learning approach for citywide short-term crash risk prediction with multi-source data," *Accident Analysis & Prevention*, vol. 122, pp. 239–254, 2019.

[16] N. C. Petersen, F. Rodrigues, and F. C. Pereira, "Multi-output bus travel time prediction with convolutional LSTM neural network," *Expert Systems with Applications*, vol. 120, pp. 426–435, 2019.

[17] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Computation*, vol. 9, no. 8, pp. 1735–1780, 1997.

[18] M. Wollmer, C. Blaschke, T. Schindl et al., "Online driver distraction detection using long short-term memory," *IEEE Transactions on Intelligent Transportation Systems*, vol. 12, no. 2, pp. 574–582, 2011.

[19] X. Shi, Z. Chen, H. Wang, and D. Yeung, "Convolutional LSTM network: a machine learning approach for precipitation nowcasting," 2015, https://arxiv.org/abs/1506.04214.

[20] Nanjing Statistic Bureau (NSB), *Nanjing Statistic Yearbook 2018*, China Statistic Press, Beijing, China, 2018.

[21] Nanjing Institute of City and Transport Planning (NICTP), *Annual Report of Nanjing Transport*, Nanjing Institute of City and Transport Planning, Nanjing, China, 2018.

[22] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," *Communications of the ACM*, vol. 60, no. 6, pp. 84–90, 2017.

[23] L. Deng, "Deep learning: from speech recognition to language and multimodal processing," *APSIPA Transactions on Signal and Information Processing*, vol. 5, pp. 1–15, 2016.

[24] Google Research Team, "Tensorflow: large-scale machine learning on heterogeneous systems," 2015, https://arxiv.org/abs/1603.04467.

[25] W. S. Sarle, "Stopped training and other remedies for overfitting," in *Proceedings of the 27th Symposium on the Interface of Computing Science and Statistics*, pp. 352–360, Fairfax, VA, USA, 1995.

[26] F. Chollet, "Keras," 2015, https://github.com/fchollet/keras.

[27] Y. Xia and J. Chen, "Traffic flow forecasting method based on gradient boosting decision tree," in *Proceedings of the 2017 5th International Conference on Frontiers of Manufacturing Science and Measuring Technology (FMSMT 2017)*, pp. 413–416, Taiyuan, China, June 2017.

[28] G. E. P. Box and D. A. Pierce, "Distribution of residual autocorrelations in autoregressive-integrated moving average time series models," *Journal of the American Statistical Association*, vol. 65, no. 332, pp. 1509–1526, 1970.

[29] S. Dabiri and K. Heaslip, "Inferring transportation modes from GPS trajectories using a convolutional neural network," *Transportation Research Part C: Emerging Technologies*, vol. 86, pp. 360–371, 2018.

[30] J. W. C. Van Lint, S. P. Hoogendoorn, and H. J. Van Zuylen, "Freeway travel time prediction with state-space neural networks: modeling state-space dynamics with recurrent neural networks," *Transportation Research Record: Journal of the Transportation Research Board*, vol. 1811, no. 1, pp. 30–39, 2002.

[31] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proceedings of the IEEE*, vol. 86, no. 11, pp. 2278–2324, 1998.