

Research Article

Attention Mechanism with Spatial-Temporal Joint Deep Learning Model for the Forecasting of Short-Term Passenger Flow Distribution at the Railway Station

Zhicheng Dai^(b),^{1,2} Dewei Li^(b),^{1,2} and Shiqing Feng³

¹School of Traffic and Transportation, Beijing Jiaotong University, No. 3 Shangyuancun, Haidian, Beijing 100044, China
 ²Frontiers Science Center for Smart High-Speed Railway System, Beijing Jiaotong University, No. 3 Shangyuancun, Haidian, Beijing 100044, China

³Urban and Transportation Planning and Design Institute, Tianjin Municipal Engineering Design and Research Institute, Haitai South Road No. 30, Tianjin Binhai High-Tech Industrial Development District, Tianjin 300392, China

Correspondence should be addressed to Dewei Li; lidw@bjtu.edu.cn

Received 7 September 2023; Revised 8 February 2024; Accepted 29 February 2024; Published 11 March 2024

Academic Editor: Jianbin Xin

Copyright © 2024 Zhicheng Dai et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Accurate understanding of passenger flow distribution is crucial for effective station crowd management. However, due to the complexity and randomness of passenger flow and the unclear spatial-temporal correlation between functional areas within the station, predicting the spatiotemporal distribution dynamics of inflow and future short-term distribution trends is challenging. Emerging deep learning models offer valuable insights for accurately predicting passenger flow distribution. Thus, we propose a deep learning architecture, named "ST-Bi-LSTM," which combines a bidirectional long short-term memory network with a spatial-temporal attention mechanism. Initially, we outline the methodologies of Bi-LSTM, the DeepWalk-based spatial attention mechanism, and the temporal attention mechanism. The spatial attention mechanism is employed to extract station spatial network topology information and enhance the representation of passenger flow characteristics in highly correlated areas during the forecasting process. Simultaneously, the temporal attention Bi-LSTM is utilized for capturing temporal correlations. The architecture comprises four branches dedicated to station real-time video monitoring data, spatial network topology, function area attributes, and train timetables. Subsequently, leveraging in-station CCTV data, passenger travel behavior data, and train timetables, we apply the architecture to the Tianjin West High-Speed Railway Station. We conduct a comparative analysis of the prediction performance and time complexity of the proposed architecture against existing baseline models, demonstrating superior performance and robustness exhibited by the ST-Bi-LSTM model (achieving a reduction in RMSE of over 10%). This study facilitates the transition of station management from passive response to active prediction of station passenger flow dynamics.

1. Introduction

With its attributes of speed, exceptional comfort, high safety standards, and environmental friendliness, global highspeed railway operations face significant demand for travel. Railway stations serve as the central hubs of the highspeed railway network, responsible for the pivotal task of gathering and dispersing passengers. For instance, Berlin Central Station and Shanghai Hongqiao Station each experience daily passenger inflows of 250,000 and 126,000, respectively. Given the substantial influx of passengers into railway stations, frequent peaks in passenger activity occur within the station premises, particularly driven by train operation events. These peaks lead to a significant concentration of passengers in specific areas within a short timeframe, generating substantial traffic demand. Such scenarios can exert considerable pressure on public security and may result in various negative incidents, including stampedes.

Predicting the real-time dynamics of in-station passenger flow is essential for ensuring efficient and safe station crowd management. However, railway stations primarily estimate passenger flow distribution through manual and video monitoring, relying on empirical methods to assess crowd tendencies and implementing reactive control measures accordingly. Without the utilization of mathematical programming and scientific methodologies, station management may demonstrate inadequate overall and dynamic performance. Therefore, it is crucial to explore new methods for accurately predicting the short-term spatial and temporal distribution of passenger flow in railway stations. Such advancements will enable proactive management strategies, enhancing efficiency and safety within the station environment.

The current methods for traffic flow forecasting can be broadly categorized into two fundamental classifications: parametric methods and nonparametric methods. Parametric models consist of linear regression model [1], autoregressive moving average (ARIMA) [2], historical average models (HA) [3], Kalman filter models [4], etc. These models are based only on historical flow data; therefore, the capacity of taking care of temperamental and nonlinear traffic flow information is restricted. Nonparametric model conquers such drawbacks, capable of capturing additional transient attributes from verifiable historical data. It consists of the K nearest neighbor algorithm [5], support vector machines (SVMs) [6], Bayesian networks [7], neural network models [8, 9], and so on. These models achieve more accurate predictions than parametric models, but their relatively simple architecture and consideration of a single variable of passenger flow limit the result accuracy.

The neural network model can realize the representation of complex high-dimensional functions through the stack of hidden layers, effectively capturing the dynamic characteristics of nonlinear traffic. Therefore, increasing research focused on utilization of the neural network model and its variants, including the Artificial Neural Networks (ANN) [10], the Radial Basis Function Networks (RBF) [11, 12], and the Recurrent Neural Networks (RNN) [13], to improve the accuracy of forecasting. Ma et al. [14], Huang et al. [15], and Li et al. [16] utilized the improved LSTM for effectively capturing nonlinear traffic dynamics. Fu et al. [17] used Gated Recurrent Units (GRUs) for predicting the short-term flow. Li et al. [18] applied LSTM to the prediction of the short-term departing passenger flow in railway stations and demonstrated the model's effectiveness. These studies only took into account the temporal features exhibited by passenger flow, ignoring the effect of spatial correlation on flow fluctuations.

Therefore, based on time series data, certain studies begin to include the spatial features exhibited by traffic data as one of the variables influencing the variation in traffic flow. Due to the effective spatial feature extraction capabilities in regular grids, Convolutional Neural Networks (CNNs) are commonly employed [19, 20]. However, due to CNNs' difficulty in applying to data with a non-Euclidean structure, Graph Convolutional Networks (GCNs) are proposed [21]. Therefore, for improved performance, numerous models make use of spatial and temporal deep learning models, following the spatial-temporal models. Zhao et al. [22] represented the entire urban route network as an undirected graph, respectively, capturing geographical and temporal dependence in the traffic flow data using GCN and GRU. Zhang et al. [23] proposed a deep learning architecture combining the residual network, GCN, and LSTM for predicting the short-term passenger inflow and outflow in the urban railway stations on a network scale. Yu et al. [24] offered a Spatial-Temporal Graph Convolutional Network (STGCN) deep learning framework (STGCN), for dealing with the time series prediction issue.

Compared to the application of time series prediction models and graph processing models, the application of attention mechanisms in flow prediction started relatively late. An increasing number of studies have begun to deeply concern attentional mechanism in spatial feature extraction considering its advantages. Wang et al. [25] utilized a unique CNN design for this task, which adopts two attention levels for better recognizing spatial-temporal patterns. Du et al. [26] proposed an attention-based LSTM network and convolution network to identify the potential patterns among the time series to achieve classification. Cinar et al. [27] proposed an extended attention model for recurrent neural networks (RNNs) designed to capture periods in time series. Zhou et al. [28] put forward a wide-attention and deep-composite (WADC) model, adopting the self-attention mechanism for extracting global pivotal features from data flows. Guo et al. [29] utilized the spatial-temporal attention mechanism for capturing the traffic flow data spatialtemporal correlation. Wang et al. [30] established a technique for attention mechanism in the GCN for determining the role played by variables in each node.

Although many models have considered spatial and temporal information, they ignore the way external factors impact data flow in prediction process. Zhang et al. [23] considered weather conditions and air quality in the process of passenger inflow prediction and quantified their influences on prediction precision but did not take into account how train timetable affects passenger inflow. Similarly, Wang et al. [31] proposed a temporal graph attention convolutional neural network model (TGACN) for forecasting the passenger density at significant station regions but did not take into account external factors, like train timetable, events, or area functional attributes.

Based on the aforementioned summary of existing research, the majority of studies have focused on leveraging temporal-spatial correlation within various networks such as railway networks and urban road networks. However, an important consideration is that although passenger flows enter the station through entrances in different directions, the internal travel process remains consistent due to the influence of functional area distribution. Consequently, the micronetwork of the station, based on passenger circulation plans, exhibits a consistent local structure. As a result, traditional graph convolutional network methods, like GCN and GAN, which rely on structure awareness, are no longer applicable. To address this limitation, we propose a spatial feature extraction method based on DeepWalk to establish location awareness of network nodes, enabling the quantification of node information with similar attributes but varying location distributions. Additionally, utilizing data obtained from AFC systems, inbound or outbound

passenger flows within railway stations are primarily considered as the prediction target. However, to date, no relevant research has explored the use of in-station micronetworks for predicting dynamic passenger flow distribution. Therefore, there is a research demand for further investigation into integrating in-station spatial and temporal correlations, as well as assessing the applicability of current methodologies for predicting passenger flow dynamics at various locations within the railway station.

To address these challenges, this research proposes a novel deep learning architecture named ST-Bi-LSTM, which incorporates DeepWalk-based spatial-temporal attention mechanisms. This architecture utilizes in-station video monitoring statistics to predict short-term passenger flow distribution in key areas. In this model, the spatial dependency between nodes in the station network is represented by a directed and unprivileged graph. The Deep-Walk model effectively captures spatial correlations by mapping semantic correlations between nodes, overcoming the limitations of traditional graph convolutional networks in distinguishing identical local structures within the network. Furthermore, the extracted spatial information is integrated with passenger flow distribution data, train schedule data, and area information. These fused features are then fed into the prediction framework to achieve accurate short-term passenger flow distribution predictions. The proposed architecture makes three primary contributions:

- The study introduces a novel DeepWalk-based combined spatial-temporal short-term prediction model for passenger flow. Additionally, it implements comprehensive prediction of passenger flow distribution in railway stations using dynamic video monitoring.
- (2) The study enriches the deep learning model by incorporating both spatial and temporal attention mechanisms to improve the accuracy and effectiveness of passenger flow prediction. It has been demonstrated that the model outperforms other baseline approaches in terms of accuracy and performance, showcasing its superiority in flow prediction.
- (3) Considering that the behavior of in-station passengers is predominantly influenced by their travel purpose, train timetable, and in-station activities, these factors are integrated into the framework along with area location information. The effectiveness of incorporating these three factors in improving prediction accuracy is demonstrated through ablation experiments.

The remainder of the paper is organized into four sections. Section 2 introduces the ST-Bi-LSTM architecture. Section 3 outlines the methodologies of spatial and temporal attention mechanisms, as well as Bi-LSTM. Section 4 presents the analysis of the case study results, along with the main findings. Section 5 summarizes the current study and discusses its implications, limitations, and potential directions for future research.

2. Forecasting Architecture

In this article, we present the ST-Bi-LSTM model architecture, depicted in Figure 1, which comprises four data branches. These branches are represented by I_1 to I_4 . The input data are collected from time t - n to t, and output data are obtained at time t + 1. We denote the datasets for each branch as Branch 2.1 to Branch 2.4. Branch 2.1 constructs station spatial network topology information and extracts spatial features of different nodes using DeepWalk algorithms. Branch 2.2 represents passenger flow distribution data. The spatial-temporal passenger flow distribution dataset is obtained by combining the spatial correlation characteristics of the output from Branch 2.1 spatial attention model. Branch 2.3 accounts for the impact of train timetables and station operating parameters on prediction accuracy. Branch 2.4 utilizes passenger travel behavior data to classify the functional characteristics exhibited by various areas within the station. After completing the preprocessing stage of the four data branches, feature fusion is conducted in the prediction architecture. Furthermore, Bi-LSTM, in combination with the temporal attention model, is employed in the trunk to extract data from each branch. Sections 2.1-2.4 provide a comprehensive overview of the model architecture.

2.1. Spatial Network Topology. The spatial network topology has been shown to be important for the short-term passenger flow prediction [32]. In railway stations, the spatial structure of the station and the facility layout limit passenger movement in railway stations. Therefore, our study treats the station space as a directed spatial network structure consisting of functional areas (vertex) and passenger circulation line (edges). Define $G = (V_t, E_t)$ as the time-varying spatial network topology in the station at time *t*. We describe the treatment of the network topology in Section 3, using the network $G = (V_t, E_t)$ as input. For the network architecture remains constant throughout our investigation, we simply take into account the real-time pattern.

$$I_1 = (V_t, E_t). \tag{1}$$

2.2. Passenger Flow Distribution Data. Area flow prediction necessitates historical flow data. In this study, we utilize passing passenger flow data collected by CCTV to generate the experimental dataset based on the functional characteristics exhibited by various regions within the station. To obtain passenger flow data, we employ a pedestrian dynamic tracking method by slicing the real-time monitoring video. Given the crowded nature of the scene, there exists an occlusion problem between passengers in the station and the camera viewpoint. This occlusion issue may result in a higher rate of missed detections if pedestrians are detected as a whole. Therefore, our study employs the head-tracking method for passenger flow tracking, enabling accurate passenger flow statistics within the designated area. The resulting passenger flow distribution data series are represented by the following equation:



FIGURE 1: ST-Bi-LSTM model architecture.

$$X_{t} = \begin{pmatrix} x_{1,t-m} & x_{2,t-m} & x_{3,t-m} & \cdots & x_{n,t-m} \\ x_{1,t-m+1} & x_{2,t-m+1} & x_{3,t-m+1} & \cdots & x_{n,t-m+1} \\ x_{1,t-m+2} & x_{2,t-m+2} & x_{3,t-m+2} & \cdots & x_{n,t-m+2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ x_{1,t} & x_{2,t} & x_{3,t} & \cdots & x_{n,t} \end{pmatrix}, \quad (2)$$

where X_t is the set of passenger flows for each partition in the space at historical time steps t, $x_{n,t}$ represents passenger flow of the area at the corresponding time step, and n denotes the number of areas for each area. According to the number of each area in the network, the areas are arranged in columns.

2.3. Train Timetable. To our knowledge, previous research has not explored the impact of train timetables on the distribution of in-station passenger flows. However, the train timetable is a crucial factor influencing the distribution of passenger flow at stations. For example, as the departure time approaches and ticketing information is broadcasted in the station, passenger flow tends to gather in the ticketing area. Conversely, passenger flow in the waiting area may decline, as Chinese railway passengers typically wait for trains within the station premises.

The study defines "train departure time proximity" as a metric to measure the impact of timetable. Since the train departure time is determined based on the train timetable, taking the opening time of the ticket gate as T_0 , the time that corresponds to passenger flow change in the local area due to passenger behavior is T. Then the difference between the two time nodes is $\Delta T = T_0 - T$, and the different time span is used to judge the train departure time proximity P at different times, as shown in Table 1. The preprocessed input data for train timetable were obtained, as shown in the following equation:

$$I_{3} = (P_{1,t}, P_{2,t}, P_{3,t}, \dots P_{n,t}),$$

$$P_{k,t} = \begin{pmatrix} P_{1,t-m} & P_{2,t-m} & P_{3,t-m} & \cdots & P_{\omega,t-m} \\ p_{1,t-m+1} & P_{2,t-m+1} & P_{3,t-m+1} & \cdots & P_{\omega,t-m+1} \\ p_{1,t-m+2} & P_{2,t-m+2} & P_{3,t-m+2} & \cdots & P_{\omega,t-m+2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ p_{1,t} & p_{3,t} & p_{3,t} & \cdots & p_{\omega,t} \end{pmatrix},$$
(3)

where $P_{k,t}$ denotes the set of departure time differences of trains with different distances from each other at time t, k denotes the region number, which is used to represent the effect of the train's departure time on the passenger flow in the region, and ω denotes the train number of the station sorted by the train's departure time in a day.

TABLE 1: Approaching value of departure time.

Р	ΔT
0	0
4	(0, 5]
3	(5, 10]
2	(10, 20]
1	(20, +∞)

2.4. Area Information. As previously discussed, the entire graph is sampled into multiple node sequences and fed into DeepWalk. However, this process results in the disruption of many edges connected to nodes and the loss of some connection information. Such loss could potentially detrimentally affect the training and prediction of graph data. Therefore, to address this issue and mitigate information loss, data that encapsulate the attributes and characteristics of the area itself are utilized as input for Branch 2.4.

First is the area location information. According to the study by Liu et al. [33], the following 3 metrics should be used to determine a target node's node relevance in a traffic network:

- (1) The number of other nodes that are connected with the target node.
- (2) Weights of edges that are connected to the target node.
- (3) The significance possessed by other nodes that are connected with the target node.

Since this study constructs a directed unweighted network, metrics (1) and (3) are considered.

In addition to considering the impact of regional functional attributes on passenger flow fluctuations, we also explore the influence of area function on passengers' in-station behavior. To investigate this further, we conducted a survey on travelers' behavior, defining the concept of "Passenger Aggregation Factor" to quantify the degree of flow aggregation in different areas. The survey on travelers' behavior was disseminated as a questionnaire via the Internet, yielding 846 valid responses. Participants ranged in age from 20 to 65 years old (mean = 26.51, standard deviation = 6.83), with a male-tofemale gender ratio of 49.76% and 50.24%, respectively. All participants had prior experience with rail travel and were familiar with the process.

Based on the survey results, the ratio of travelers heading to a single area within the travel process to the total number of respondents was categorized into three classes: 0% to 40%, 40% to 60%, and greater than 60%. These classes corresponded to aggregation values of 1 to 3, respectively, with higher values indicating greater passenger flow fluctuation in the area, as shown in Table 2.

Hence, equations below explain the input.

$$I_{4} = (A_{1,t}, A_{2,t}, \dots, A_{n,t}),$$

$$A_{k,t} = \begin{pmatrix} a_{1,t-m} & a_{2,t-m} & a_{3,t-m} \\ \vdots & \vdots & \vdots \\ a_{1,t} & a_{2,t} & a_{3,t} \end{pmatrix},$$
(4)

where $A_{k,t}$ denotes information data of area k at time t and $a_{i,t}$ is value of type i information for the region k at time t.

To generate weighted indicators, the network flattens the preprocessed input data and incorporates them into the fully connected layer. Subsequently, for the first and second layers, a Bi-LSTM with 128 neurons is introduced. The network then transfers the results to the feature fusion section.

3. Methodology

The methodology design of the proposed model incorporates attention mechanisms and Bidirectional Long Short-Term Memory (Bi-LSTM). Therefore, the study provides a brief overview of the individual methodologies of each component.

3.1. DeepWalk-Based Spatial Attention Mechanism. Previous studies have indicated that, in contrast to the structure-aware feature of other graph neural networks, the position-aware feature of DeepWalk can effectively distinguish identical local structures within the network. Consequently, it has the capability to capture a broader range of graph structures and extract vertex information [34]. DeepWalk was introduced by Perozzi et al. [35] in 2014, comprising a random walk generator for node sequence sampling and a semantic model SkipGram for embedding representation of node information.

In this study, DeepWalk is adopted to take the node sequence obtained by random walk sampling as data input. Mapping the semantic correlation between nodes to spatial correlation by SkipGram, DeepWalk uses the probability distribution of co-occurrence between the target node v_i and other nodes v_j , $j \in [i - w, i + w]$, $j \neq i$ in the network as the spatial attention weight to distinguish the contribution made by passenger flow fluctuations in other areas of the network to the predicted area passenger flow.

Algorithms 1 and 2 show the main methods of Deep-Walk and SkipGram, where W_{v_i} is a sequence of nodes with v_i as root and $\Psi(v_j)$ denotes a mapping function, which maps the vertex v_j to representation vector. The objective function of SkipGram holds the objective function of maximizing the co-occurrence probability of neighbor nodes and target nodes in the sequence. Equation (5) explains the problem.

$$\min_{\Phi} -\log\Pr(\{v_{i-w}, \dots, v_{i-1}, v_{i+1}, \dots, v_{i+w}\} | \Phi(v_i)).$$
(5)

After this, soft attention is to calculate the probability distribution. The Hierarchical Softmax is chosen as the spatial attention distribution calculation function, taking the probability calculation of the input vector of node v_i as an example.

$$\lambda_{j} = \frac{\Pr(\nu_{j}|\Phi(\nu_{i}))}{\sum_{k=1}^{|V|} \exp\left(\Pr\left(\nu_{k}|\Phi(\nu_{i})\right)\right)},\tag{6}$$

where λ_j denotes attention distribution and $\Pr(v_j | \Phi(v_i))$ is the scoring function for attention model.

TABLE 2: Attractiveness of different functional areas.

Passenger behavior	Area name	Aggregation factor		
Queuing up for ticket checking	Ticket checking area	3		
Waiting behavior in the station	Waiting area	3		
Ticket pickup	Ticket collection area	2		
Bathroom use behavior	Restrooms	2		
Shopping behavior	Business area	1		
Restaurant dining	Station Restaurant	1		

Input: Directed nonweight graph <i>G</i> (<i>V</i> , <i>E</i>)
window size w
embedding size d
walks per set vertex γ
walk length ϕ
Output: Vertex representation matrix $\Psi \in \mathbb{R}^{ V \times d}$
(1) Initialization: Sliding window sampling Ψ from $u^{ V \times d}$
(2) Construct a binary Tree T from V
(3) for $i = 0$ to γ do
(4) $O = $ Shuffle (V)
(5) for each $v_i \in O$ do
(6) $W_{v} = \text{RandomWalk}(G, v_i, \phi)$
(7) SkipGram (Ψ, W_{ν}, w)
(8) end for
(9) end for

ALGORITHM 1:DeepWalk (G, w, d, γ, ϕ) .

```
(1) for each v_j \in W_{v_i} do

(2) for each u_k \in W_{v_i}[j - w: j + w] do

(3) J(\Psi) = -\log \Pr(u_k | \Psi(v_j))

(4) \Psi = \Psi - \alpha * (\partial J / \partial \Psi)

(5) end for

(6) end for
```

ALGORITHM 2:SkipGram (Ψ, W_{ν_i}, w) .

3.2. Bidirectional-LSTM. Due to the spatial-temporal correlation of passenger flow and the large, complex, and variable data involved, prediction process involves handling high volumes of data.. We establish a Bi-LSTM deep learning model to capture spatial-temporal features of the passenger flow. Figure 2 shows the structure of each neuron in LSTM network, which consists of three parts: forgetting gate f_t , input gate i_t , and output gate O_t , which can be represented as follows:

$$f_{t} = \sigma \Big(W_{f} \cdot \Big[h_{t-1}^{l} \cdot x_{t} \Big] + b_{f} \Big),$$

$$i_{t} = \sigma \Big(W_{i} \cdot \Big[h_{t-1}^{l} \cdot x_{t} \Big] + b_{c} \Big),$$

$$\widetilde{C}_{t} = \tanh \Big(W_{c} \cdot \Big[h_{t-1}^{l} \cdot x_{t} \Big] + b_{c} \Big),$$

$$C_{t} = f_{t} \cdot C_{t-1} + i_{t} \cdot \widetilde{C}_{t},$$

$$o_{t} = \sigma \Big(W_{o} \cdot \Big[h_{t-1}^{l}, x_{t} \Big] + b_{o} \Big),$$

$$h_{t}^{l} = o_{t} \cdot \tanh (C_{t}).$$

$$(7)$$

In above equations, W_f, W_i, W_c stands for the corresponding weight of different gates f, i, c, b_f, b_i, b_c denotes the bias term in different parts of gate in forgetting gate f_t , input gate i_t , and output gate $O_t, h_t^{(l)}$ represents *l*th hidden layer state at time *t*, and σ is the sigmoid activation function.

Bi-LSTM considers forward and backward LSTM simultaneously. For each moment of flow, the input is provided to two LSTMs in opposite directions, one participating A in the forward computation and another A' in the reverse. The final network output depends on the summation of the forward and reverse calculations, but the weights are not shared between the two directions. Figure 3 displays the Bi-LSTM structure.

$$\begin{aligned} H_t^{(1)} &= f \Big(W^{(1)} H_{t-1}^{(1)} + U^{(1)} x_t + b^{(1)} \Big), \\ H_t^{(2)} &= f \Big(W^{(2)} H_{t+1}^{(2)} + U^{(2)} x_t + b^{(2)} \Big), \\ h_t &= H_t^{(1)} \oplus H_t^{(2)}, \\ y_t &= g \big(V h_t \big), \end{aligned}$$
(8)



FIGURE 2: Neuronal structure of LSTM.



FIGURE 3: Bidirectional recurrent neural network structure expanded by time series.

where x_t refers to the input flow at time t, U, V represent different weight matrices in the network, and $W^{(1)}$, $W^{(2)}$ correspond to different weights from both the forward and reverse order directions.

3.3. Temporal Attention Mechanism. Prediction models for passenger flow distribution can be influenced by various factors such as area attributes, spatial network topology, passenger entries, and train timetables, resulting in a high level of complexity. Therefore, assigning weight scores solely based on recentness in the Bi-LSTM network might be inadequate. Given that LSTM combined with attention mechanisms has proven effective in traffic flow prediction, many researchers have incorporated it into short-term passenger flow prediction [36–38]. Consequently, we introduced a temporal attention mechanism to capture different feature weights.

To address the limitation of recentness-based weight score assignment by traditional attention mechanisms, we utilize a fully connected network to provide weights that can be graded based on the Bi-LSTM output. This approach builds upon earlier work by Zhang et al. [36]. As a result, the proposed model assigns scores to the output weights. Subsequently, we obtain the attentionbased hidden layer output H * by the following procedure.

$$H^{*} = \tanh\left(\operatorname{att}(H,q)\right),$$

$$\operatorname{att}(H,q) = \sum_{n=1}^{t} \alpha_{n}h_{n},$$

$$\alpha_{n} = \frac{\exp\left(s\left(h_{n},q\right)\right)}{\sum_{j=1}^{t}\exp\left(s\left(h_{j},q\right)\right)},$$

$$s\left(H,q\right) = v^{T} \tanh\left(W'x + U'q\right),$$

(9)

where *q* is the query vector, α_n denotes attention distribution, $s(h_n, q)$ denotes additive model, and W', U', and v^T are learned parameters.

4. Case Study

In this section, we provide an overview of the study scenario, detailing the model configuration and comparing it with other prediction models to assess the predictive effectiveness of the developed framework.

4.1. Types of Graphics. In our case study, we focus on Tianjin West High-Speed Railway Station, designed to accommodate 23.67 million passengers annually, with a maximum capacity of 5,000 people, making it the largest railway station in Tianjin. Our research scenario specifically targets the passenger waiting area of the Beijing-Shanghai high-speed railway on the second floor of Tianjin West Railway Station and its surrounding vicinity. The layout of this area is schematically depicted in Figure 4. The study area comprises 4 entrance areas (green areas), 5 passenger flow intermediate areas (gray areas) indicating ticket pickup machines and commercial areas, 4 ticket checking areas (blue areas), and 3 waiting areas (yellow areas). The topology of the station spatial network constructed for this study is illustrated in Figure 5.

Passenger flow data are collected and quantified in different areas using multiangle and multidirectional realtime monitoring video within the station. OpenCV and its classifier are employed to achieve dynamic people counting within delineated areas, as illustrated in Figure 6. This figure presents an example of dynamic tracking statistics from the waiting hall monitoring video at Tianjin West Station. Following data collection, one-day passenger flow data from each area within the station are obtained. Specifically, passenger flow and train timetable data from 16 areas within the hub are selected for analysis during the time period of 8: 00-19:10 on December 15, 2020. Figure 7 visualizes the distribution of passenger flow data across the 16 regions during the study period. To analyze the model's prediction performance at different time granularities, the passenger flow data are processed into 10-second, 20-second, 30second, and 60-second time intervals. For model calibration, the validation split rate is set at 0.2. Section 2 provides examples of data preprocessing.



FIGURE 4: Second floor layout of Tianjin West Passenger Transport Station.



FIGURE 5: Spatial network topology of Tianjin West Passenger Transport Station.

4.2. Model Configuration. The model is built using TensorFlow-2.12.0, Python 3.8, and Keras 2.3.1. All models are trained and tested on GeForce RTX 3060 Ti GPUs. In the Skip-gram model, the hidden layer comprises 128 neurons. Considering the typical flow of passengers into the station (enter the station, get tickets, other travel behaviors, wait for

the train, and take the train), the sequence length parameter ϕ should be greater than 5 to capture at least 5 nodes in the passenger flow network. The sliding window size must not exceed half of the number of inbound process nodes, and thus it is set to w = 2. We set the embedding dimension to 8 to maximize the retention of spatial feature information.



FIGURE 6: Station elevated floor CCTV passenger flow dynamic tracking and statistics.



FIGURE 7: Waterfall chart of passenger flow distribution data during the study period.

The Bi-LSTM consists of 64 neurons per layer for the deep learning model part. The network structure and specific parameters of the ST-Bi-LSTM network are visualized using the Graphviz-8.1.0 module, as shown in Figure 8.

We set the validation split rate at 0.2 for model calibration. With the number of iterations set to 200, the training process adopts model checkpoints and early stopping techniques to preserve the optimal model and prevent overfitting, thereby avoiding improper parameter initialization. Figure 9 illustrates the training and validation losses, which exhibit significant fluctuations during the first 175 epochs. However, after 175 epochs, both losses stabilize, and there are no further rises or falls in either the training or test set losses. Instead, only slight oscillations occur, indicating the robustness of our proposed model. 4.3. Baseline Models. The section primarily focuses on evaluating the effectiveness of various time series prediction models. Except for the ARIMA prediction model, all deep learning model optimizers used for comparison are uniformly selected from RAdam. RAdam is an optimizer that combines the advantages of SGD for fast convergence and Adam for fast training [39]. For the three variants of the prediction framework, we use the same parameters as the proposed model to ensure a fair comparison.

ARIMA: It is a common traditional time series data forecasting model. We use the minimum AIC principle to determine the model order to achieve the prediction for passenger flow data.

BPNN: the BPNN network includes 2 hidden layers, each of which has 64 neurons.



FIGURE 8: Network structure and detailed parameter configuration of ST-Bi-LSTM.



FIGURE 9: Variation for training loss and validation loss.

CNN: CNN networks contain convolutional layers with two average convolutional layers, with 32 and 64 neurons, and convolutional kernel size of 3 * 3.

RNN: the RNN consists of two hidden layers, also containing 64 neurons.

LSTM: the LSTM contains two hidden layers, each consisting of 64 neurons.

ST-Bi-LSTM (no graph): we deleted Branch 2.1.

ST-Bi-LSTM (1A): considering passenger flow data for only one area related to the forecast area.

ST-Bi-LSTM (no A): we deleted Branch 2.3.

ST-Bi-LSTM (no T): we deleted Branch 2.4.

ST-Bi-LSTM (no T&A): we only adopted Branch 2.1. and 2.2.

Adam ST-Bi-LSTM: the Adam optimizer is used for training, and other parameters are the same as the RAdam ST-Bi-LSTM, with a learning rate of 0.0001.

4.4. Loss Function and Evaluation Metrics. The end-to-end training serves for model optimization. The mean square error (MSE) is chosen for the prediction process to calculate the loss function. Meanwhile, for assessing the model performance, the following three metrics were employed: root mean square error (RMSE), mean absolute error (MAE), and mean absolute percentage error (MAPE).

Loss = MSE =
$$\frac{1}{n} \sum_{i=1}^{n} (y_i - \hat{y}_i)^2$$
,
RMSE = $\sqrt{\frac{1}{n} \sum_{i=1}^{n} (y_i - \hat{y}_i)^2}$, (10)
MAE = $\frac{1}{n} \sum_{i=1}^{n} |y_i - \hat{y}_i|$,
MAPE = $\frac{100\%}{n} \sum_{i=1}^{n} |\frac{y_i - \hat{y}_i}{y_i}|$.

In the formula, y_i is the true value, \hat{y}_i is the predicted value, and *n* is the number of training samples.

4.5. Results and Discussion

4.5.1. Prediction Error Evaluation Index. According to Table 3 and Figure 10, ST-Bi-LSTM significantly outperforms the basic mathematical statistical model and other deep learning models. The ARIMA model is the least effective in training due to its inability to capture the full range of nonlinear characteristics of passenger traffic. CNNs have the worst prediction results relative to other deep learning models, and this gap increases as the time granularity increases. Relative to traditional RNN, the LSTM has improved the accuracy of training effect for passenger flow data. BPNN training obtains slightly better results relative to

	1	1	1			0						
Time granularity		10 s			20 s	;		30 s			60 s	
Indicators	RMSE	MAE	MAPE (%)	RMSE	MAE	MAPE (%)	RMSE	MAE	MAPE	RMSE	MAE	MAPE (%)
ARIMA	16.282	10.281	55.1	43.069	23.602	58.7	63.227	35.636	63.2%	120.975	63.829	88.1
BPNN	3.973	2.975	14.1	8.843	5.877	18.5	14.824	8.627	23.9%	44.443	28.725	26.6
CNN	5.961	3.239	18.4	10.604	8.685	22.3	31.238	14.611	27.4	47.961	25.614	28.6
RNN	5.626	5.879	13.8	9.365	6.174	18.3	30.962	12.962	23.9%	46.716	22.606	23.6
LSTM	5.203	5.808	13.7	8.722	5.114	17.5	28.806	10.589	23.9%	31.320	17.701	22.4
ST-Bi-LSTM (no graph)	2.913	2.214	12.7	7.047	4.961	17.2	14.878	8.534	23.3%	29.499	17.782	21.7
ST-Bi-LSTM (1A)	2.927	2.023	12.9	6.744	4.297	16.3	13.696	8.759	20.9%	34.015	22.683	31.4
ST-Bi-LSTM (no A)	2.849	2.094	12.6	6.614	4.291	15.9	13.247	8.191	20.1%	25.461	14.874	19.4
ST-Bi-LSTM (no T)	2.931	2.145	13.1	6.761	4.306	16.2	13.600	8.35	20.8%	28.103	16.912	20.7
ST-Bi-LSTM (no T&A)	2.986	2.966	13.5	6.651	4.395	16.5	13.855	8.527	20.2%	30.216	17.411	22.1
Adam ST-Bi-LSTM	2.889	2.018	13.0	6.640	4.309	16.2	13.377	8.346	19.7%	26.513	14.821	19.4
ST-Bi-LSTM	2.716	1.921	12.1	6.605	4.282	15.7	13.215	8.126	18.4%	25.133	15.036	18.9

TABLE 3: Comparison of prediction performances obtained using different TGs in different baseline models.



FIGURE 10: Comparison of prediction performances obtained for different models and TGs.

RNN and LSTM. The passenger flow prediction effect is significantly optimized when comparing ST-Bi-LSTM with other deep learning and mathematical statistical models optimized considerably. Obviously, the architecture in the study is highly robust, i.e., even if a branch is removed from the framework, the prediction results do not show a large error change as a result.

Among ST-Bi-LSTM and its five variants, due to numerous characteristics, the full ST-Bi-LSTM functions, including spatial-temporal correlation, train timetable, area attributes, and other factors, are fully considered in the framework. The results indicate that spatial correlation has the greatest impact on prediction outcomes, followed by inputting train timetable and area information. This suggests that passenger flow is influenced bytrain departure times, as passengers tend to gather both at waiting areas and ticket gates. Our model helps to obtain the passenger flow fluctuation degree in the area. After quantifying these two effects, based on 30s timegranularity passenger flow data, the RMSE, MAE, and MAPE decreased from 13.855, 8.527, and 20.2% to 13.215, 8.126, and 18.4%, respectively, for the model considering schedule and area attractiveness compared to ST-Bi-LSTM (no T&A).

Figure 11 compares the prediction performance exhibited by ST-Bi-LSTM with Bi-LSTM and LSTM by displaying their prediction residuals for 30 s timegranularity passenger flow data. The three models have the same peak residuals at 16:44 and 18:30, and the fitting effect is weak for large passenger flows. The residual peaks of the three models remain the same at 16:44. Still, the residual peaks of the 18:30 passenger flow reflect large differences, with the residuals of the LSTM ranging from -100 to 150, the residuals of the temporal attention Bi-LSTM and ST-Bi-LSTM being significantly smaller, mostly in the range of -100 to 50, and the residuals of the ST-Bi-LSTM having the smallest overall fluctuations and the lowest peaks. Therefore, the residuals of the ST-Bi-LSTM fusion model exhibit a stronger stability and the absolute value is lower, relative to the other two models.

4.5.2. Fusion Model Prediction Performance Evaluation. For more clearly testing the model prediction performance, scatter plots and linear fits of the actual and predicted values for the cases of time granularity of 10 s, 20 s, 30 s, and 60 s are plotted in Figure 12. The black line is the linear fit target.



FIGURE 11: Passenger flow residuals for LSTM, temporal attention Bi-LSTM, and ST-Bi-LSTM.



FIGURE 12: Scatter plot of predicted and actual values at different time granularities.

The calculated Pearson's correlation coefficients at different time granularities were 0.99482, 0.98614, 0.97466, 0.97972, and the standard deviations of the residuals were 2.6827, 6.7084, 14.2942, 25.8434.. As the time granularity increases, the prediction effect decreases slightly, especially for some of the peak passenger flow locations. As the time granularity increases for the same study period, the dataset's data decrease, and the learning effect decreases slightly. However,

the overall model can still fit the passenger flow data more accurately at different time granularities.

The prediction output of the framework and its residuals for different time granularity data are analyzed, and Figure 13 shows the passenger flow data and residuals for different time granularity data. It can be seen that the residual peaks appear in similar areas for all four-time granularities, indicating that the points of passenger flow



FIGURE 13: Passenger flow residuals at different time granularities (10 s, 20 s, 30 s, and 60 s).

	Method	RMSE
1	ST-Bi-LSTM	6.605
2	GCN ² -Bi-LSTM (16)	6.755
3	GCN ² -Bi-LSTM (32)	6.617
4	GCN ² -Bi-LSTM (64)	7.061
5	GCN ² -Bi-LSTM (96)	6.645
6	GCN ² -Bi-LSTM (128)	7.033
7	GCN ³ -Bi-LSTM (16)	6.477
8	GCN ³ -Bi-LSTM (32)	6.420
9	GCN ³ -Bi-LSTM (64)	7.007
10	GCN ³ -Bi-LSTM (96)	7.020
11	GCN ³ -Bi-LSTM (128)	6.627

TABLE 4: Comparison of GCN and DeepWalk-based method prediction errors for different parameters.



FIGURE 14: RMSE comparison for GCNs and DeepWalk-based model.



FIGURE 15: Predicted and actual value comparison for check-in area.



FIGURE 16: Predicted and actual value comparison for passenger waiting area.

anomalies are similar for the prediction of different time granularity data. As the time granularity increases, the residual peak increases, but the peak range can be kept in the range of -100 to 200, and the residual values of other time periods are kept in the range of -50 to 50.

4.5.3. Comparison of Graph Feature Extraction Performance. For verifying the prediction effectiveness of the network features extracted by DeepWalk graph neural network compared with GCN, the study compares the prediction error results for two and three layers of GCN with different neurons, respectively, without considering the influence of exogenous factors in the process.

Table 4 lists the passenger flow prediction errors of the GCN with different parameters and plots the spline connectivity as in Figure 14, which shows that the prediction errors of the two-layer GCN (blue line) are overall higher than those of the DeepWalk (black dashed line) for the same prediction using Bi-LSTM. In comparison, the prediction errors of the three-layer GCN are only slightly

lower than those of the DeepWalk when the neurons are taken as 16 and 32. However, comparison of their time complexities $(O_{\text{DeepWalk}}(d \cdot \phi \cdot |V| \cdot \log(|V|)), O_{\text{GCN}}(|V| \cdot D^2) D$ is the feature dimension), DeepWalk-based spatial feature extraction method has 81% lower time complexity than the GCN-based method in the application scenario of this study.

4.5.4. Prediction Performance on Individual Area. The study selects three typical areas within the Tianjin West Hub for examining whether the forecasting framework can effectively predict the passenger flow in different types of areas with varying characteristics of fluctuation in space. The first is Area 6, a ticket gate in the Tianjin West Hub, which shows significant fluctuations in passenger flow over time. The second is Area 9, a commercial area within the hub, adjacent to the waiting area and ticket gates. The last one is Area 13, a waiting area between Area 7 and 16 ticket gates, which is connected to the commercial area, the entry gates, and the ticket gates.



FIGURE 17: Predicted and actual value comparison for commercial area.

According to Figure 15, for the ticket gate area 6, the predicted values at the four time granularities always agree with the actual values during both peak and off-peak periods (refer to the residual plot in Figure 11), showing the strong robustness of ST-Bi-LSTM. Because the ticket gate is the end of the passenger circulation line, which converges the passenger flows from different flow lines and is a significant area influenced by the train schedule, the regularity is evident, which helps the prediction effect.

As an essential transition area for the passenger waiting and ticketing process in the hub, the waiting area, i.e., Area 13 passenger flow, exhibits multipeak characteristics, and the prediction framework maintains good performance at all three time granularities, particularly during the peak passenger flow period, as shown in Figure 16.

Unlike the ticketing and waiting areas, the commercial area within the hub, i.e., Area 9, shows low regularity and significant variation in passenger inflow. However, as shown in Figure 17, the prediction framework we designed still better captures the regional passenger flow trend. Moreover, the fitting effect of the framework does not deteriorate as the time granularity increases from 10 to 60 seconds, indicating the reliability of the prediction performance of the framework.

In summary, the feasibility of the passenger flow forecasting framework constructed by the study for accurate forecasting is further verified by demonstrating the forecasting results for different areas within the station.

5. Conclusion

The paper proposes ST-Bi-LSTM, which incorporates a DeepWalk-based spatial attention model, temporal attention model, and Bi-LSTM for predicting short-term passenger flow distribution. It makes three key contributions:

(1) Capture of Spatiotemporal Characteristics. The ST-Bi-LSTM model effectively captures spatiotemporal characteristics of passenger flows, network topological data, area information, and train timetable influences on prediction accuracy.

- (2) Superior Prediction Accuracy. Utilizing dynamic passenger flow distribution data, ST-Bi-LSTM outperforms other benchmark models, achieving favorable prediction accuracy. Metrics such as RMSE, MAE, and MAPE demonstrate significant improvements compared to individual models. Additionally, when compared with a GCN-based deep learning prediction framework, ST-Bi-LSTM excels in spatial information extraction.
- (3) Robustness and Versatility. The ST-Bi-LSTM model exhibits strong robustness under different ablation experiments. Prediction outcomes remain consistent even when individual branches are removed (ST-Bi-LSTM (no graph), ST-Bi-LSTM (no T), ST-Bi-LSTM (no A), and ST-Bi-LSTM (no T&A)), showcasing the framework's applicability and reliability in railway station environments.

However, due to the limited availability of in-station monitoring video data and railway operation data obtained in this research, there is ample room for further improvement and dissemination of the research findings. Future research endeavors will focus on investigating the impact of heterogeneous traveler behavior on passenger flow fluctuations resulting from regional functional differences, with the aim of significantly enhancing prediction accuracy and extending forecast duration.

Regarding experimental data, the collection of passenger flow data through dynamic pedestrian monitoring from surveillance videos in the hub poses challenges in processing. Consequently, only a portion of the station area is selected for this study to validate the reliability of the prediction framework. We intend to expand the study area in future research initiatives.

Additionally, from the results of the scatter plot test, it is observed that prediction accuracy slightly decreases as time granularity increases, particularly for peak predictions. Future studies are advised to address these limitations and focus on mitigating the aforementioned challenges.

Data Availability

Some or all data, models, and codes that support the findings of this study are available from the corresponding author upon reasonable request.

Conflicts of Interest

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

Acknowledgments

This work was supported by the Fundamental Research Funds for the Central Universities (grant no. 2022JBQY006) and Search Project of China Railway Beijing Group Co. Ltd. (grant no. 2023BK01).

References

- H. Sun, C. Zhang, and B. Ran, "Interval prediction for traffic time series using local linear predictor," in *Proceedings of the 7th International IEEE Conference on Intelligent Transportation Systems*, pp. 410–415, IEEE Cat. No.04TH8749, Washington, DC, USA, March 2004.
- [2] C. Han, S. Song, and C. Wang, "Real-time adaptive prediction of short-term traffic flow based on ARIMA model," *Journal of System Simulation*, vol. 16, no. 7, pp. 1530–1532, 2004.
- [3] L. Jing and W. Guan, "A summary of traffic flow forecasting methods," *Journal of Highway and Transportation Research* and Development, vol. 21, no. 3, pp. 82–85, 2004.
- [4] 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.
- [5] J. W. C. Van Lint and C. P. I. J. Van Hinsbergen, "Short-term traffic and travel time prediction models," *Artificial Intelligence Applications to Critical Transportation Issues*, vol. 22, no. 1, pp. 22–41, 2012.
- [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] S. Sun, C. Zhang, and G. Yu, "A bayesian network approach to traffic flow forecasting," *IEEE Transactions on Intelligent Transportation Systems*, vol. 7, no. 1, pp. 124–132, 2006.
- [8] H. Shao and B.-H. Soong, "Traffic flow prediction with long short-term memory networks (LSTMs)," in *Proceedings of the* 2016 IEEE Region 10 Conference (TENCON), pp. 2986–2989, Singapore, November 2016.
- [9] T. Bogaerts, A. D. Masegosa, J. S. Angarita-Zapata, E. Onieva, and P. Hellinckx, "A graph CNN-LSTM neural network for short and long-term traffic forecasting based on trajectory data," *Transportation Research Part C: Emerging Technologies*, vol. 112, pp. 62–77, 2020.
- [10] N. Glišović, M. Milenković, N. Bojović, L. Švadlenka, and Z. Avramović, "A hybrid model for forecasting the volume of passenger flows on Serbian railways," *Operational Research*, vol. 16, no. 2, pp. 271–285, 2015.
- [11] Y. Li, X. Wang, S. Sun, X. Ma, and G. Lu, "Forecasting shortterm subway passenger flow under special events scenarios using multiscale radial basis function networks," *Transportation Research Part C: Emerging Technologies*, vol. 77, pp. 306–328, 2017.
- [12] H. Li, Y. Wang, X. Xu, L. Qin, and H. Zhang, "Short-term passenger flow prediction under passenger flow control using a dynamic radial basis function network," *Applied Soft Computing*, vol. 83, Article ID 105620, 2019.
- [13] Y. Liu, Y. Qin, J. Guo, C. Cai, Y. Wang, and L. Jia, "Short-term forecasting of rail transit passenger flow based on long shortterm memory neural network," in *Proceedings of the 2018 International Conference on Intelligent Rail Transportation* (*ICIRT*), pp. 1–5, Singapore, December 2018.
- [14] X. Ma, Z. Tao, Y. Wang, H. Yu, and Y. Wang, "Long shortterm 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] H. Huang, T. Wang, J. Liu, and S. Xie, "Predicting urban rail traffic passenger flow based on LSTM," in *Proceedings of the* 2019 IEEE 3rd Information Technology, Networking, Electronic

and Automation Control Conference (ITNEC), pp. 616–620, Chengdu, China, March 2019.

- [16] H. Li, Y. Zhao, C. Ma, K. Wang, X. Huang, and W. Zhang, "Short-term passenger flow prediction of urban rail transit based on SDS-SSA-LSTM," *Journal of Advanced Transportation*, vol. 2022, Article ID 2589681, 11 pages, 2022.
- [17] R. Fu, Z. Zhang, and L. Li, "Using LSTM and GRU neural network methods for traffic flow prediction," in *Proceedings of the 2016 31st Youth Academic Annual Conference of Chinese Association of Automation (YAC)*, pp. 324–328, Wuhan, China, November 2016.
- [18] J. Li, P. Huang, and Y. Yang, "Passenger flow prediction of high speed railway based on lstm deep neural network," in *Rail Norrköping 2019. 8th International Conference on Railway Operations Modelling and Analysis (ICROMA)*, pp. 723–739, Norrköping, Sweden, May 2019.
- [19] K. Wang, C. Ma, Y. Qiao, X. Lu, W. Hao, and S. Dong, "A hybrid deep learning model with 1DCNN-LSTM-Attention networks for short-term traffic flow prediction," *Physica A: Statistical Mechanics and Its Applications*, vol. 583, Article ID 126293, 2021.
- [20] 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, p. 818, 2017.
- [21] X. Zan and J. S. L. Lam, "Multibranch adaptive fusion graph convolutional network for traffic flow prediction," *Journal of Advanced Transportation*, vol. 2023, Article ID 8256907, 13 pages, 2023.
- [22] L. Zhao, Y. Song, C. Zhang et al., "T-GCN: a temporal graph convolutional network for traffic prediction," *IEEE Transactions on Intelligent Transportation Systems*, vol. 21, no. 9, pp. 3848–3858, 2020.
- [23] J. Zhang, F. Chen, Z. Cui, Y. Guo, and Y. Zhu, "Deep learning architecture for short-term passenger flow forecasting in urban rail transit," *IEEE Transactions on Intelligent Transportation Systems*, vol. 22, no. 11, pp. 7004–7014, 2021.
- [24] B. Yu, H. Yin, and Z. Zhu, "Spatio-temporal graph convolutional networks: a deep learning framework for traffic forecasting," 2017, https://arxiv.org/abs/1709.04875.
- [25] L. Wang, Z. Cao, G. de Melo, and Z. Liu, "Relation classification via multi-level attention CNNs," in *Proceedings of the* 54th Annual Meeting of the Association for Computational Linguistics, vol. 1, pp. 1298–1307, Berlin, Germany, October 2016.
- [26] Q. Du, W. Gu, L. Zhang, and S.-L. Huang, "Attention-based LSTM-CNNs for time-series classification," in *Proceedings of* the 16th ACM Conference on Embedded Networked Sensor Systems, pp. 410-411, Shenzhen, China, November 2018.
- [27] Y. G. Cinar, H. Mirisaee, and P. Goswami, "Position-based content attention for time series forecasting with sequence-tosequence RNNs," 2017, https://arxiv.org/abs/1703.10089.
- [28] J. Zhou, H.-N. Dai, H. Wang, and T. Wang, "Wide-attention and deep-composite model for traffic flow prediction in transportation cyber–physical systems," *IEEE Transactions on Industrial Informatics*, vol. 17, no. 5, pp. 3431–3440, 2021.
- [29] S. Guo, Y. Lin, N. Feng, C. Song, and H. Wan, "Attention based spatial-temporal graph convolutional networks for traffic flow forecasting," *Proceedings of the AAAI Conference* on Artificial Intelligence, vol. 33, no. 01, pp. 922–929, 2019.
- [30] J. Wang, Q. Chen, and H. Gong, "STMAG: a spatial-temporal mixed attention graph-based convolution model for multidata flow safety prediction," *Information Sciences*, vol. 525, pp. 16–36, 2020.

- [31] Y. Wang, Y. Qin, J. Guo, Z. Cao, and L. Jia, "Multi-point short-term prediction of station passenger flow based on temporal multi-graph convolutional network," *Physica A: Statistical Mechanics and Its Applications*, vol. 604, Article ID 127959, 2022.
- [32] Y. Han, S. Wang, Y. Ren, C. Wang, P. Gao, and G. Chen, "Predicting station-level short-term passenger flow in a citywide metro network using spatiotemporal graph convolutional neural networks," *ISPRS International Journal of Geo-Information*, vol. 8, no. 6, p. 243, 2019.
- [33] R. Liu, L. Gao, and J. Wu, "Key nodes mining in transport networks based in PageRank algorithm," in *Proceedings of the* 2009 Chinese Control and Decision Conference, pp. 4413–4416, Guilin, China, June 2009.
- [34] J. You, R. Ying, and J. Leskovec, "Position-aware graph neural networks," 2019, https://arxiv.org/abs/1906.04817.
- [35] B. Perozzi, R. Al-Rfou, and S. Skiena, "DeepWalk: online learning of social representations," 2014, https://arxiv.org/ abs/1403.6652.
- [36] Z. Zhang, M. Li, X. Lin, Y. Wang, and F. He, "Multistep speed prediction on traffic networks: a deep learning approach considering spatio-temporal dependencies," *Transportation Research Part C: Emerging Technologies*, vol. 105, pp. 297–322, 2019.
- [37] Q. Liu, B. Wang, and Y. Zhu, "Short-term traffic speed forecasting based on attention convolutional neural network for arterials," *Computer-Aided Civil and Infrastructure Engineering*, vol. 33, no. 11, pp. 999–1016, 2018.
- [38] Y. Wu, H. Tan, L. Qin, B. Ran, and Z. Jiang, "A hybrid deep learning based traffic flow prediction method and its understanding," *Transportation Research Part C: Emerging Technologies*, vol. 90, pp. 166–180, 2018.
- [39] L. Liu, H. Jiang, and P. He, "On the variance of the adaptive learning rate and beyond," 2019, https://arxiv.org/abs/1908. 03265.