

Article

Multi-Step Peak Passenger Flow Prediction of Urban Rail Transit Based on Multi-Station Spatio-Temporal Feature Fusion Model

Jianan Sun ¹, Xiaofei Ye ^{1,2,*} , Xingchen Yan ³ , Tao Wang ⁴ and Jun Chen ⁵ 

¹ Faculty of Maritime and Transportation, Ningbo University, Fenghua Road 818#, Ningbo 315211, China; canaan040712@163.com

² Jiangsu Province Collaborative Innovation Center of Modern Urban Traffic Technologies, Southeast University, Road #2, Nanjing 211189, China

³ College of Automobile and Traffic Engineering, Nanjing Forestry University, Nanjing 210037, China; xingchenyan.acad@gmail.com

⁴ School of Architecture and Transportation, Guilin University of Electronic Technology, Jinji Road 1#, Guilin 541004, China; wangtao@guet.edu.cn

⁵ School of Transportation, Southeast University, Nanjing 211189, China; chenjun@seu.edu.cn

* Correspondence: yexiaofei@nbu.edu.cn

Abstract: Accurate prediction of station passenger flow is crucial for optimizing rail transit efficiency, but peak passenger flow in urban rail transit (URT) is often disrupted by random events, making predictions challenging. In this paper, in order to solve this challenge, the Bi-graph Graph Convolutional Spatio-Temporal Feature Fusion Network (BGCSTFFN)-based model is introduced to capture complex spatio-temporal correlations. A combination of a graph convolutional neural network and a Transformer is used. The model separately inputs land use (point of interest, POI) and station adjacency information as features into the BGCSTFFN model, using the Pearson correlation coefficient matrix, which is evaluated on real passenger flow dataset from 1 to 25 January 2019 in Hangzhou. The results showed that the model consistently provided the best prediction results across different datasets and prediction tasks compared to other baseline models. In addition, in tasks involving predictions with different combinations of inputs and prediction steps, the model showed superior performance at multiple prediction steps. Its practical application is validated by comparing the results of passenger flow prediction for different types of stations. In addition, the impact of these features on the prediction accuracy and the generalization ability of the model were verified by designing ablation experiments and testing on different datasets.

Keywords: urban rail transit; short-term passenger flow prediction; transformer; POI; multi-step prediction



Academic Editor: Mahyar Amirgholy

Received: 24 December 2024

Revised: 21 January 2025

Accepted: 31 January 2025

Published: 3 February 2025

Citation: Sun, J.; Ye, X.; Yan, X.; Wang, T.; Chen, J. Multi-Step Peak Passenger Flow Prediction of Urban Rail Transit Based on Multi-Station Spatio-Temporal Feature Fusion Model.

Systems **2025**, *13*, 96. <https://doi.org/10.3390/systems13020096>

Copyright: © 2025 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

With the rapid development of information technology, China's urban rail transit (URT) has achieved large-scale and networked operations. However, the passenger flow at URT stations presents complex spatial and temporal characteristics, which pose a major challenge to optimizing system operations and improving the passenger travel experience. Therefore, it is of great significance to conduct a forecasting study of passenger flow at URT stations from passenger flow data.

With the wide application of automatic fare collection (AFC) systems, one can easily access and record a large amount of historical travel data. As a result, many scholars have performed rail passenger flow forecasting based on URT passenger flow data collected by

AFC systems using machine learning and deep learning models. Not only is it important to consider the impact of feeding historical passenger flow data collected by the AFC system into the prediction model so that it can learn from past events and become more adaptive, but also the impact of a number of factors such as weather, weekdays, the topology of the rail network, and the nature of the land use (e.g., POI). As a result, many studies have shown significant inconsistencies between predicted and actual passenger flow, and the errors in some of the predictions have been quite large. Only a few studies have considered multiple influencing factors such as weather and land use data in URT passenger flow prediction models.

Therefore, in order to solve the problem of the accuracy of URT passenger flow prediction decreasing due to the influence of these external factors, this study explores the spatial and temporal distribution of rail traffic and its influencing factors by using a variety of data sources of rail traffic, and proposes a novel short-term passenger flow prediction method for rail traffic, so that when predicting the short-term inbound passenger flow of URT, we can comprehensively take into account the characteristics of the passenger flow of URT and the external influencing factors, so as to accurately grasp the changes in passenger flow.

Therefore, in order to solve the problem of the accuracy of URT passenger flow prediction decreasing due to the influence of these external factors, this study explores the spatial and temporal distribution of rail traffic and its influencing factors by using a variety of data sources of rail traffic, and proposes a novel short-term passenger flow prediction method for rail traffic, so that when predicting the short-term inbound passenger flow of the URT, we can take into account comprehensively the characteristics of the passenger flow of the rail traffic and the external influencing factors, so as to accurately grasp the changes in passenger flow.

The main contributions of this paper are as follows:

1. In this study, an innovative approach, a Bi-graph Graph Convolutional Spatio-Temporal Feature Fusion Network (BGCSTFFN) combining multi-graph convolutional and Transformer models, is proposed, aiming to effectively model complex spatio-temporal dependencies in sequence data. By introducing a Bi-Graph Convolutional Network (BGCN), the method is able to deal with the similarity of the adjacency and point of interest (POI) information of multiple stations in a rail transit system, and capture the potential patterns of passenger flow changes in different time periods. In terms of the fusion of multiple data features, the feature fusion module merges feature sequences from different sources through dynamic weighted summation, and this approach enhances the robustness and accuracy of the prediction model, providing more accurate technical support for future passenger flow prediction in rail transit.
2. In order to investigate whether there is a higher degree of correlation in passenger flows between stations characterized by POI information, this study introduces the Pearson correlation coefficient to compute the similarity matrix of POI information among different stations. Specifically, by utilizing Pearson's correlation coefficient, we are able to quantify the similarity of POI distributions across stations, thus providing more accurate spatial correlation information for the model. This is because different types of point of interest (POI) have different attraction characteristics. Even if two stations are not directly adjacent, stations with similar surrounding POI may exhibit similar traffic flow patterns. For example, stations located near commercial centers, cultural attractions, or transportation hubs may have similar traffic flow characteristics, even if they are geographically distant from each other. Therefore, by introducing such POI similarity features in the model training process, the intrinsic connection between POI features and passenger flow can be reflected more effectively, thus improving

the ability to capture passenger flow patterns and improving the prediction accuracy. Therefore, by introducing such POI similarity features in the model training process, the intrinsic connection between POI features and passenger flow can be reflected more effectively, thus enhancing the ability to capture passenger flow patterns and enhancing the prediction accuracy.

3. This study validates the advantages of BGCSTFFN in short-term passenger flow prediction at URT stations during peak hours based on a real dataset of URT passenger flow in Hangzhou. The experimental results show that BGCSTFFN consistently achieves outstanding and stable performance in the prediction tasks with different combinations of input and output step sizes, demonstrating its strong robustness and adaptability in multi-step prediction tasks. Feature ablation experiments are conducted to verify the effects of different features on the prediction accuracy of the model.

The subsequent parts of the paper are organized as follows: Section 2 provides an overview of existing studies. Section 3 outlines the definition of short-term passenger flow prediction in peak hour URT networks. Section 4 clarifies the structure and mathematical formulation of BGCSTFFN. Section 5 examines the prediction performance of the model on a real dataset in HCM City. Finally, Section 6 summarizes the work of this thesis. Metro passenger flow prediction helps to optimize metro operations and ensure rational use of stations and trains. Accurate forecasting can avoid congestion, improve transport efficiency, reduce waiting times, and increase passenger satisfaction. This helps to reduce traffic congestion and improve the overall efficiency of urban transport systems.

2. Literature Review

With the continuous development of rail transit systems and the maturity of new technologies, research on short-duration passenger flow prediction for rail transit is also increasing. The development of URT short-duration passenger flow prediction can be roughly divided into three stages. The first stage is the traditional model based on mathematical statistics, the second stage is the model based on machine learning, and the third stage is the model based on deep learning.

The first stage in the development of short-term passenger flow forecasting is the traditional mathematical statistics-based model, in which it can be divided into two main categories, as follows: one of the most representative is the autoregressive differential moving average (autoregressive integrated moving average model, ARIMA) time series model, and the other is based on other mathematical and statistical methods. Researchers including Chen [1], Cao [2], and Yan [3], have proposed a short-term passenger flow prediction model based on an autoregressive moving average (ARMA), and Jiao [4] used the Kalman filtering method to predict the passenger flow of URT. These statistical models for short-term urban rail passenger flow prediction are based on historical time series data to predict future passenger flow, which are relatively simple in model structure and have a certain ability to learn and adapt to historical time series data, but with the increase in passenger flow, the accuracy of their prediction is limited in improvement, and they can no longer satisfy the surge in the accuracy of short-term passenger flow prediction requirements. Therefore, most of the models developed at this stage are no longer used.

With the development of machine learning, Li [5], Wei [6], Sun [7], Tang [8], and Roos [9] proposed a BP neural network, an SVM (Support Vector Machine) model, an SVR (Support Vector Regression) model, and a Dynamic Bayesian network, respectively, for short-term URT passenger flow prediction. The hybrid prediction model based on machine learning can model more complex dependencies and have higher prediction accuracy than the traditional prediction model based on mathematical statistics, but most of the models

at this stage cannot take into account the more complex spatial and temporal correlations between stations, and they can only make individual predictions for one or a few stations, but the prediction accuracy is also affected by a certain degree of accuracy for the whole rail transit network with hundreds of stations. However, most models at this stage cannot take into account the more complex spatial and temporal correlations between stations and can only predict one or a few stations individually.

Deep learning, an important branch of machine learning, has been developing rapidly in recent years, and its excellent prediction performance has brought about a dramatic change in short-term traffic flow prediction. Pratap et al. [10] constructed an artificial neural network (ANN) model for predicting passenger flow in the North Central Railway (NCR) region and obtained a good prediction accuracy. Nataša et al. [11] proposed a hybrid model based on the integration of Genetic Algorithms (GA) and artificial neural networks (ANN) for predicting the monthly passenger flow in Serbian Railways.

With the development of deep learning techniques, these models based on LSTM [12–15] and Gated Recursive Units (GRU) [16], etc., are gradually becoming mainstream due to their expertise in capturing dependencies in time series, e.g., Li et al. [17] selected the number of relevant stops, number of outbound stops, holidays, peak times, and weather as five features that affect passenger flow. Yang et al. [15] proposed an improved model based on an Enhanced Long Short-Term Feature Memory (ELF-LSTM) neural network. The proposed network enhanced the long-term time-dependent features embedded in the passenger flow data and combined the short-term features to predict the origin destination (OD) flow in the coming hour. Hao et al. [18] proposed an end-to-end framework for large-scale urban URT passenger flow prediction using a sequence-to-sequence model embedded with an attention mechanism. The model used a stacked bidirectional LSTM encoder and a unidirectional LSTM decoder and was validated on the Singapore Metro system dataset. Guo et al. [19] proposed a model based on the fusion of Support Vector Regression (SVR) and Long Short-Term Memory (LSTM) neural networks for predicting URT passenger flow. Du et al. [20] proposed a deep irregularity called a DST-ICRL convolutional residual LSTM network model for URT passenger flow prediction. Wang et al. [21] proposed a learning network based on the optimal passenger flow input information algorithm (MTFLN) method. The experimental results demonstrated that the method improved the training efficiency and prediction accuracy of traditional prediction models. In addition, hybrid models are beginning to be applied to capture complex spatio-temporal dependencies. Chen [22], Li [23], and Xion [13] proposed a parallel architectural prediction model combining CNN and LSTM, which successfully explored the spatio-temporal characteristics of the URT passenger flow and thus significantly improved the prediction accuracy. Wang [24] proposed the SR-GA adaptive station arrangement method to rearrange the line stations and combined it with GRU and Conv1d to construct a model of RS-Conv1dGRU for the short-term inbound passenger flow prediction of URT. Convolutional neural network (CNN) are widely used for the prediction and analysis of data in tasks dealing with Euclidean data, such as images and regular grids. However, when it comes to modeling graph-structured data or traffic network data, graph convolutional network (GCN) exhibit superior capabilities. GCN is able to effectively capture complex spatial dependencies based on graphs, and therefore, many researchers have introduced graph convolutional neural network (GCN) in combination with models such as LSTM to improve the capturing of spatio-temporal features in passenger flows. Ye et al. [25] used three combined modules of graph convolutional network (GCN) and LSTM, respectively, to capture spatio-temporal influences. Finally, the outputs of the two components were fused with different weights to predict the URT passenger flow. Zhang et al. [26] proposed a deep learning architecture combining residual network (ResNet), graph convolutional

network (GCN), and long short-term memory (LSTM) (called “ResLSTM”) to predict the short-term passenger flow of URT. Zhao et al. [27] developed a spatio-temporal neural network called the time graph convolutional network (TGCN), which was a combination of GCN and GRU, and predicted urban traffic flow. Meanwhile, the attention model can capture global and dynamic spatio-temporal features, which helps to improve the accuracy of traffic flow prediction. Therefore, Xia [28] and Zhang et al. [29] introduced the model with an attention mechanism into the field of traffic flow prediction, thus improving the performance of complex traffic flow prediction. Zhang et al. [30] developed an advanced URT multi-step short-term passenger flow prediction model, which made full use of point of interest (POI) information as the graphical data and extracted the POI features through CNN. The model integrated an LSTM network based on the Transformer mechanism, an attention mechanism module, and a CNN network, which significantly improved the prediction accuracy and model performance. Liu et al. [31] addressed the challenge of predicting station passenger flow during peak hours in urban rail transit. They introduced the Multi-Sequence Spatio-Temporal Feature Fusion Network (MSSTFFN) model based on trend decomposition to capture the complex spatio-temporal correlations and realize accurate short-term passenger flow prediction.

A summary of the methods used in the deep learning modelling literature and their limitations and advantages is shown in Table 1.

Table 1. Methods used in the deep learning modelling literature and their limitations and advantages.

Literature Number	Method	Advantage	Limitation
[12,14,15,17,18,20,21]	LSTM	Long-term dependency handling Trend capturing	High computational cost Low sensitivity to short-term fluctuations
[16]	GRU	Efficient for sequential data Handles non-linear relationships well Faster training compared to LSTM	Low sensitivity to long-term dependencies Sensitive to hyperparameters
[19]	SVR-LSTM	Combines linear and non-linear strengths Captures both short-term and long-term dependencies	Complex model design Complicated debugging
[24]	CNN- GRU	Captures spatial and temporal patterns Robust to noise	Complex architecture High computational cost
[13,22,23,30]	CNN- LSTM	Captures spatial and temporal features Handles long-term dependencies	Requires large datasets and optimization Complex model structure
[25,26]	GCN-LSTM	Captures spatial and temporal dependencies Effective for graph-structured data	Complex model design Requires large datasets and optimization
[27]	GCN-GRU	Captures spatial and temporal dependencies Effective for graph-structured data	Complex architecture Sensitive to hyperparameter optimization
[28,29]	Attention mechanism	Focuses on important features Handles long-range dependencies	High computational cost Requires large datasets and optimization
[30,31]	Transformer	Captures long-range dependencies Parallelizable and scalable	High computational cost Requires large datasets and optimization

Overall, the current research trend is to increasingly employ a combination of deep learning methods. These methods are dedicated to improving the accuracy of short-term URT passenger flow prediction by integrating the extraction of multiple features related to passenger flow in URT stations. Although there have been studies that have made significant progress in this area, there are still some obvious limitations:

1. There is still a relative lack of specialized forecasting studies of passenger flows at URT stations during peak hours. While existing forecasting models and methods perform well during regular hours, it is often difficult for existing models to accurately capture these dynamics during peak hours due to the dramatic increase in passenger flow, pressure on station capacity, and the complexity of passenger behavior. The

specificity of peak periods requires more refined and targeted forecasting strategies to effectively address the challenges posed by passenger flow fluctuations. However, the current research on peak periods still appears to be insufficient, which limits the accurate prediction and management of passenger flows during peak periods.

2. There is still little research on incorporating point of interest (POI) into forecasting models for URT passenger flow forecasting. POI, such as commercial areas, residential areas, or office buildings, can have an impact on the trends of passenger flow changes, but many existing models tend to ignore these factors, and current research on the integration of POI is not deep enough, which limits the comprehensive understanding and effective prediction of URT passenger flows. Therefore, in order to investigate whether the station passenger flow between stations characterized by POI information has a higher degree of correlation, based on the distribution of POI information points around each station, the Pearson correlation coefficient between the distribution data is calculated to indicate the similarity of the surrounding POI information between each station, and is inputted into the prediction model so that the prediction model can more accurately capture the changes in passenger flow due to the geographic location and the type of activity, thus improving the accuracy and practicality of the prediction.

3. Problem Statement

3.1. Passenger Flow Sequence Features

The URT morning peak hour traffic sequence extracts inbound traffic from historical AFC traffic data. The information fields of the AFC data include Passenger ID, Arrival Time, Arrival Station, Departure Time, and Departure Station. URT traffic Y_i^t represents the number of passengers that entered the station i during the time period t . $Y^t \in \mathbb{R}^{N \times 1}$ Indicates the volume of traffic at all URT stations during the time period t . N represents the number of all stations on the URT lines. The passenger flow data are normalized before inputting into the prediction model.

Data normalization is the process of eliminating the effect of differences in magnitude by transforming the original data into the interval $[0, 1]$ by applying a linear transformation. Equation (1) is as follows:

$$y = \frac{y_{raw} - y_{min}}{y_{max} - y_{min}} \quad (1)$$

where y_{raw} denotes the original data of the sample, y_{min} denotes the minimum value in the sample, and y_{max} denotes the maximum value in the sample.

3.2. Station Adjacency Relationship Features

In the URT network, there is a relationship between the physical proximity of individual stations and their passenger flow patterns. Stations that are geographically adjacent tend to exhibit similar patterns of passenger distribution, though this relationship may not hold universally across all hours due to the variability in passenger demand and other influencing factors. The adjacency relationships of some of the stations are shown in Table 2, where 0 indicates non-adjacent and 1 indicates adjacent. Therefore, the adjacency characteristics of URT stations are represented by a matrix Adj_A , where A_{ij} denotes the adjacency relationship between URT stations, and $A_{ij} = 0$ denotes that the stations are not adjacent, and $A_{ij} = 1$ denotes that the stations are adjacent.

3.3. POI Similarity Features

Urban rail passenger traffic will be related to the surrounding area's land use, road traffic facilities, employment units, and other transport travel attraction point characteristics.

POI data (point of interest), also known as point of interest data, are the abstraction of geographically located physical objects as points.

Table 2. Station adjacency relationships.

Station	0	1	2	...	80
0	0	1	0	...	0
1	1	0	1	...	0
2	0	1	0	...	0
3	0	0	1	...	0
...
80	0	0	0	...	0

This study uses the POI classification code list provided by the Gaode Map platform to classify the POI information points around the station. The specific classification attributes are shown in Table 3. The raw data of the POI information around each URT station are obtained by counting and organizing the number of POI information points within 800m around each URT station, as shown in Table 4.

Table 3. POI information classification codes and attributes.

Codes	010000	020000	030000	040000	050000
Attributes	Auto Service	Auto Dealers	Auto Repair	Motorcycle Service	Food and Beverages
Codes	060000	070000	080000	090000	100000
Attributes	Shopping	Daily Life Service	Sports and Recreation	Medical Service	Accommodation Service
Codes	110000	120000	130000	140000	150000
Attributes	Tourist Attraction	Commercial House	Governmental Organization and Social Group	Science/Culture and Education Service	Transportation Service
Codes	160000	170000	180000	190000	200000
Attributes	Finance and Insurance Service	Enterprises	Road Furniture	Place Name and Address	Public Facility
Codes	220000	970000	990000		
Attributes	Incidents and Events	Indoor Facilities	Pass Facilities		

Table 4. Raw data of POI category counts around some URT stations.

Station	010000	020000	030000	...	990000
0	18	0	2	...	0
1	26	10	8	...	0
2	61	34	24	...	0
3	35	14	8	...	0
...
80	33	1	2	...	0

In order to investigate whether the URT passenger flows between URT stations with POI information characteristics have a higher degree of correlation, based on the distribution of POI information points around each URT station, the Pearson correlation coefficient between the distribution data is calculated to represent the similarity coefficient of the POI information around each station, and the calculation process is shown in Equations (2) and (3):

$$Peason(X, Y) = \frac{E(XY) - E(X)E(Y)}{\sqrt{E(X^2) - E^2(X)}\sqrt{E(Y^2) - E^2(Y)}} \quad (2)$$

$$R_POI = \text{Pearson}(STA_POI_i, STA_POI_j) \quad (3)$$

where $E()$ represents the mathematical expectation; R_POI represents the information correlation degree of POI around different URT stations; and STA_POI_i represents the POI information sequence around the URT station.

3.4. Weather and Time Label Features

Weather can be categorized into six categories: sunny, cloudy, overcast, drizzle, showers, and moderate rain. The categories are denoted by (0~5, respectively), while in terms of whether it is a working day time label, a holiday is considered as a non-working day, and 0–1 is used to denote whether it is a working day, where 0 denotes a non-working day and 1 denotes a working day. The above time label features of weather and whether it is a working day or not, respectively, are input into the prediction model by generating one-hot codes.

3.5. Description of the Problem

The task of forecasting short-duration passenger flow at URT stations during peak hours is to utilize historical passenger flow data collected by the AFC and certain exogenous variables to forecast the inbound passenger flow at each station within the URT network for future time intervals. The short-term passenger flow forecasting problem for peak hour URT stations can be formulated as follows:

$$\hat{Y}^{t+k} = \underset{Y^{t+k}}{\text{optimal}} \left(F \left(\left[Y^{t-m}, Y^{t-m+1}, \dots, Y^t \right], \text{Adj_A}, R_POI, [TW_{t-m}, TW_{t-m+1}, \dots, TW_t] \right) \right) \quad (4)$$

where \hat{Y}^{t+k} is the predicted URT short-term passenger flow at the time, and $[Y^{t-m}, Y^{t-m+1}, \dots, Y^t]$ represents the historical URT passenger flow sequence for the m time interval from the time $t - m$ to the time t . TW denotes the weather and time labels matrix. $F()$ denotes the function of prediction, and $\text{optimal}()$ denotes the prediction function with optimal parameters.

4. Methodology

4.1. BGCSTFFN Model

The framework of the proposed spatio-temporal Transformer model based on bi-graph convolutional fusion for the task of urban rail transit passenger flow prediction is shown in Figure 1.

The model consists of a spatial feature extraction module, a feature fusion module, and a Transformer module. Firstly, the URT passenger flow distribution sequences and spatial features of different station connections are input into the spatial feature extraction module to explore spatial features, which consists of two graph convolutional network (GCN) layers with independent parameters. Secondly, the two URT passenger flow distribution sequences with spatial features extracted by the GCN layer are inputted into the feature fusion module, which will fuse multiple external spatial feature URT passenger flow sequences weighted and summed into a single sequence containing the extracted multi-feature URT passenger flow sequences, and then the sequence is inputted into the URT passenger flow Transformer module, in which the input of the decoder layer not only includes the output of the encoder layer, but also a time and weather labeling feature matrix that has been processed by a convolutional layer and has the same size as the prediction step. The entire Transformer module consists of six stacked encoder and decoder layers and a feedforward neural network. The results are finally summed up and the predictions are output through a linear layer.

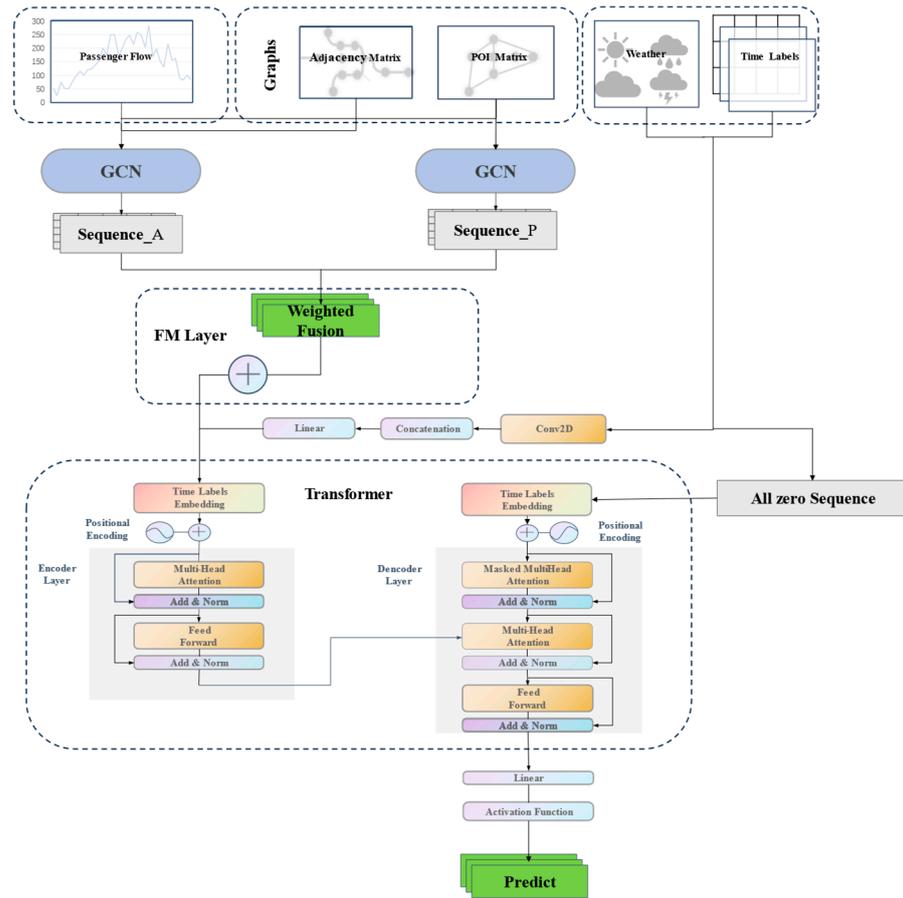


Figure 1. BGCSTFFN general framework.

4.2. Graph Convolutional Network (GCN)

A graph convolutional neural network (GCN) is an extension of a convolutional neural network to graph-structured data, which can aggregate the neighbor information of each node by performing convolutional operations on the graph and can utilize them for updating and learning node features. It also retains the importance of spatial relationships, thus effectively capturing spatial relationships in graph-structured data.

The two GCN layers include three inputs, the feature matrix $Y^T (Y^T \in \mathbb{R}^{N \times F^0})$, and the adjacency matrix $Adj_A (Adj_A \in \mathbb{R}^{N \times N})$, where N is the number of stations and F^0 is the number of station features.

The computation rule for the GCN layer is given in Equation (5):

$$H^l = f(D^{-\frac{1}{2}} \widetilde{Adj_A} D^{-\frac{1}{2}} H^{l-1} W^{l-1}) \quad (5)$$

where l denotes the number of layers; H^l denotes the output of l layer; D denotes the degree matrix of the nodes; Adj_A denotes the adjacency matrix of the topological graph; W^{l-1} denotes the weight parameter of the first layer; and $f()$ denotes the nonlinear activation function.

4.3. Spatial Feature Module

Spatial correlation exists in two ways: (1) Adjacency relationships, where adjacent stations tend to be similarly affected or show higher correlations. (2) Different stations with high POI information similarity will have similar temporal patterns, and some physically unconnected or neighboring stations will have similar trends in certain time periods, and this correlation will change over time. In this section, we construct a Bi-Graph Convolutional

tional (BGCN) layer to simultaneously extract adjacency correlations and the correlations of different stations with the same POI information similarity. Therefore, the spatial feature module includes three inputs, as follows: a sequence of passenger flows at N URT stations $Y_N = \{Y_1, Y_2, \dots, Y_i\}$, where each sequence of URT station flows includes a sequence of passenger flows at T time steps $Y^T = \{Y^1, Y^2, Y^3, \dots, Y^t\}$; a matrix characterizing the adjacency of different stations Adj_A , with dimension $N \times N$; and a matrix representing the similarity of the POI of different stations R_POI , where N is the number of stations. After inputting the three features into the two GCN layers, respectively, the passenger flow sequence with two spatial features and the station information is obtained. Finally, the two features are fused by the weighted summation of random parameters. The calculation process is shown in Equations (6)–(8), and the overall structure of the module is shown in Figure 2.

$$data_{adj} = H^l(Adj_A, Y_N) \quad (6)$$

$$data_{poi} = H^l(R_POI, Y_N) \quad (7)$$

$$data_{hybrid} = \alpha \odot data_{adj} + (1 - \alpha) data_{poi} \quad (8)$$

wherein $data_{hybrid}$ represents a passenger flow sequence having both an adjacency correlation and similarity information of the same POI information.

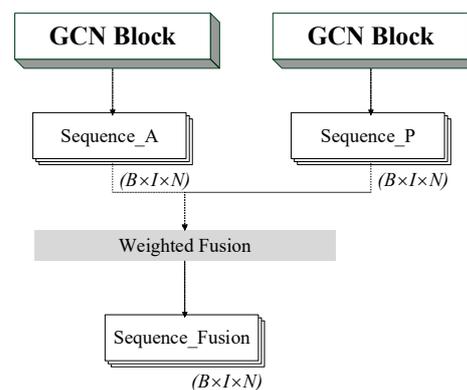


Figure 2. Spatial feature module framework.

4.4. Transformer Module

The Transformer framework is shown in Figure 3, where the model consists of several encoding and decoding layers stacked on top of each other. Each encoding layer consists of a multi-head attention layer and a feed-forward neural network layer. In order to enhance the generalization ability of the model, additional residual connections and layer normalizations are also added between the multi-head attention layer and the feed-forward neural network layer. The decoding layer is structurally similar to the coding layer, and the decoding layer consists of two multi-head attention layers and one feed-forward neural network layer, with layer normalization added between the multi-head attention layer and the feed-forward neural network layer. Unlike the encoding layer, the first multi-head attention mechanism of the decoding layer employs a masking mechanism to prevent future information in the sequence from being used for prediction.

1. Positional encoding

When performing the task of predicting sequence data, the sequence order usually carries important information. When the input data is imported into the Transformer module, the position of each batch is first encoded at the positional encoding layer to provide positional information for each position of the input sequence, and the positional

encoding is performed using a combination of sine and cosine functions. The calculation formula is shown in Equations (9) and (10).

$$PE_{(POS,2i)} = \sin(pos/10000^{2i/d_{model}}) \quad (9)$$

$$PE_{(POS,2i+1)} = \cos(pos/10000^{2i/d_{model}}) \quad (10)$$

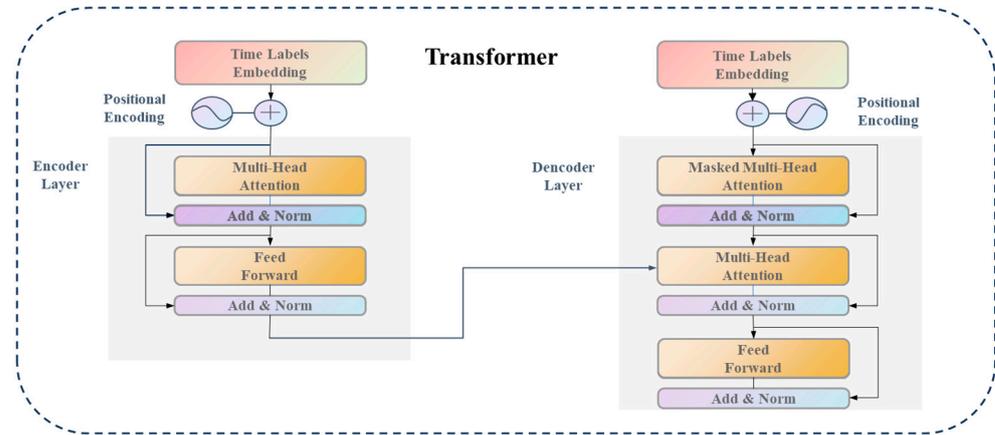


Figure 3. Transformer model framework.

2. Multi-attention mechanism

The multiple attention mechanism will transform the input sequence linearly to obtain the representation of Query, Key, and Value; these linear transformations will be realized by different weight matrices. After the transformation is performed, attention computation is performed for each Key using each Query. After the computation is completed, the Softmax function is used for normalization to obtain the attention weight matrix, and finally the attention weight matrix and the value matrix are multiplied and summed to obtain the output of the multi-head attention mechanism. The calculation process is as in Equation (11):

$$\text{Attention}(Q, K, V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V \quad (11)$$

where Q, K, V denotes the vector matrix of queries, keys, and values; and d_k is the feature dimension used for normalization.

5. Case Study

5.1. Data Source and Processing

The URT network studied in this paper is a localized real subway network in Hangzhou, which consists of three URT lines with a total of 81 stations, and the network structure is shown in Figure 4.

In this paper, the AFC data from the morning peak hour (6:30–9:30) of each station in the study area for 25 consecutive days in January 2019 are selected, and they are corrected for missing values and outliers. The time series of the inbound passenger flow of each station during the morning peak hour were collated by taking 5min and 15min as the time granularity, respectively. The attributes of the two datasets are detailed in Table 5. Therefore, 80% of them were randomly selected as the training set and the rest as the test set.

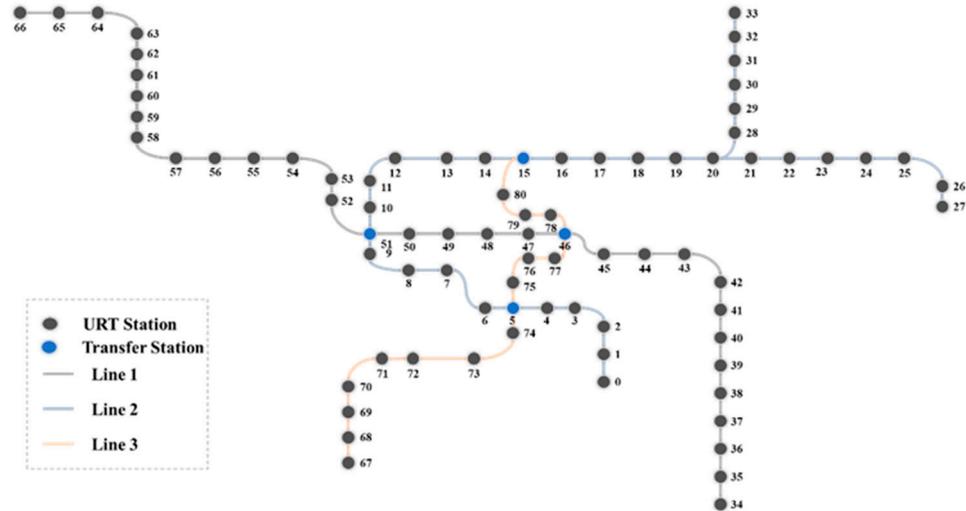


Figure 4. Structure of the case road network.

Table 5. Table of dataset attributes.

Dataset	5 min Dataset	15 min Dataset
Statistics interval	5 min	15 min
Dataset size	900 * 81	300 * 81

5.2. BGCSTFFN Model Evaluation Parameter Setting

The model input data include passenger flow sequence features, the URT station adjacency matrix, POI similarity features, and the weather and time label features. The input feature is the original passenger flow sequence of each station, and the data structure is $900 * 81$ or $300 * 81$, where 900 and 300 denote the number of bars in the sample and 81 is the number of stations. The input sequence size of the model passenger flow is $B * I * N$, where B denotes the batch_size, I denotes the input step size, and N is the total number of stations. The adjacency matrix is used to describe whether the stations are adjacent to each other, which is a quantitative description of the network topology, and the data structure of both the POI information similarity matrix and the station adjacency matrix is $81 * 81$.

The proposed deep learning model is implemented with the Pytorch (v2.5.0) framework in the python programming language. Three classical error assessment metrics are chosen to evaluate the performance of the model experimental results:

Mean Absolute Error (MAE):

$$MAE = \frac{1}{N} \sum_{i=1}^N |y_i - \hat{y}_i| \quad (12)$$

Root Mean Square Error (RMSE):

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (y_i - \hat{y}_i)^2} \quad (13)$$

Goodness of Fit (R-squared, R^2):

$$R^2 = 1 - \frac{\sum_{i=1}^N (y_i - \hat{y}_i)^2}{\sum_{i=1}^N (y_i - \bar{y})^2} \quad (14)$$

In which, y_i , \hat{y}_i , and \bar{y} are the actual values, predicted values, and mean values of passenger flow for URT stations, and N is the total number of URT stations.

5.3. Multi-Step Prediction Results Analysis

Different input time steps and prediction steps were used by performing cross-over experiments on the two datasets. The time interval for the first dataset was set to 5 min, while the second dataset used a time interval of 15 min. The results of the experiments are shown in Table 6.

Table 6. Results for different combinations of model inputs and prediction steps.

Prediction Step	No. 1 Step	No. 2 Step	No. 3 Step	No. 4 Step	No. 5 Step	No. 8 Step	No. 12 Step
5 min dataset input step							
4 steps	86.24%	86.28%	85.98%	86.11%	85.34%	84.79%	83.21%
6 steps	87.69%	85.57%	83.64%	84.01%	84.01%	84.07%	83.57%
8 steps	89.63%	88.87%	86.66%	85.24%	84.69%	85.78%	83.27%
12 steps	90.38%	89.35%	87.32%	85.66%	85.03%	84.21%	84.85%
24 steps	89.77%	89.21%	85.98%	85.04%	85.25%	84.21%	84.34%
15 min dataset input step							
4 steps	92.02%	91.98%	91.35%	91.18%	90.85%	90.01%	89.44%
6 steps	92.43%	92.09%	91.85%	91.62%	91.31%	90.31%	90.33%
8 steps	93.17%	92.91%	92.61%	92.07%	91.15%	90.65%	90.55%
12 steps	94.10%	93.91%	93.14%	92.83%	92.24%	91.30%	90.76%
24 steps	93.79%	92.98%	92.98%	92.84%	92.22%	91.24%	90.76%

Note: better than 90%; between 85% and 90%; between 80% and 85%.

Having been experimentally validated, this model demonstrates excellent prediction ability, achieving more than 80% prediction accuracy in a given situation, despite an input time step of only four or six steps. The predictive performance of the model shows a gradual increase as the input step length grows. For training sequences of different lengths, there are significant differences between the predicted and actual values. However, constrained by the structure of the model, the prediction accuracy almost reaches its peak when the input step length reaches 12 steps.

In addition, we find that regardless of the value of the input time step scenario, when the prediction step is smaller, the prediction accuracy is higher. This is because the model can capture the current trend more accurately, and also, smaller prediction intervals can help us make more accurate predictions.

By testing the dataset at different time intervals, i.e., over different time periods, the experimental results show that different time intervals of data will have a significant impact on the performance of the model. On the 15 min dataset, the model showed a superior prediction accuracy than on the 5 min dataset. This difference may be due to the fact that a wider range of time periods is provided, and therefore, simpler temporal information can be obtained. In addition, the model shows excellent prediction on both sets of samples.

To summarize the experimental results, we conclude that the model proposed in this paper is an efficient spatio-temporal feature extraction model. It shows a good performance and strong generalization ability in time series prediction tasks. In practical applications, the combination of inputs and prediction steps can be reasonably selected to significantly improve the prediction performance based on the task requirements and data characteristics.

5.4. Comparison with Baseline Models

In order to verify the performance of the proposed model in short-term forecasting, we compare the proposed model with the following baseline model.

ARIMA: A representative traditional statistical model. It combines autoregressive (AR), difference (I), and moving average (MA) components for modeling and forecasting of time series data with certain trends and seasonality. It has only passenger flow sequences as input features.

CNN: A deep learning model with two convolutional layers and two linear layers is built. The first convolutional layer contains 32 convolutional kernels, and the second convolutional layer has 64 kernels, each of which is 3×3 in size.

GRU [16]: Gated Recurrent Unit (GRU) is a neural network model for processing sequential data and belongs to a variant of Recurrent Neural Networks (RNNs). A deep learning model with two GRU layers and one linear layer is built. Each GRU layer consists of 512 hidden neural units.

LSTM: A deep learning model for URT passenger flow prediction with two LSTM layers and a linear layer is developed. Each LSTM layer consists of 512 hidden neural units.

GCN-GRU: A deep learning model for URT passenger flow prediction consisting of one GCN layer and two GRU layers is constructed. Each GRU layer consists of 81 hidden neural units.

GCN-LSTM [26]: A deep learning model for URT passenger flow prediction consisting of one GCN layer and two LSTM layers is constructed. Each LSTM layer consists of 81 hidden neural units.

Transformer: A deep learning model with one Transformer layer and one linear layer is constructed. The Transformer layer consists of sixteen attention headers, and there are six layers for both the encoder and the decoder.

From the experimental results in Tables 7 and 8, we can draw the following conclusions.

Table 7. The 5 min dataset comparison of model results.

Model	One-Step Prediction			Two-Step Prediction			Three-Step Prediction		
	MAE	RMSE	R ²	MAE	RMSE	R ²	MAE	RMSE	R ²
ARIMA	35.21	60.47	54.74%	35.21	60.47	54.74%	35.21	60.47	54.74%
CNN	34.23	58.66	63.40%	34.17	39.14	61.96%	25.66	42.01	60.32%
GRU	21.24	35.66	74.76%	22.47	36.76	71.47%	23.15	38.66	70.76%
LSTM	20.77	33.47	77.54%	21.11	37.75	75.55%	23.67	38.01	74.57%
GCN-GRU	19.64	32.97	80.40%	21.24	35.66	79.76%	21.24	35.66	78.34%
GCN-LSTM	18.54	27.45	82.26%	19.97	31.68	81.43%	21.12	35.79	80.45%
Transformer	13.68	21.63	86.16%	19.77	33.24	85.97%	21.14	35.94	84.74%
BGCSTFFN	10.99	21.74	90.38%	11.24	21.67	89.35%	13.38	23.14	87.32%

Table 8. The 15 min dataset comparison of model results.

Model	One-Step Prediction			Two-Step Prediction			Three-Step Prediction		
	MAE	RMSE	R ²	MAE	RMSE	R ²	MAE	RMSE	R ²
ARIMA	40.21	72.98	56.45%	40.21	72.98	56.45%	40.21	72.98	56.45%
CNN	37.10	54.66	65.48%	37.25	55.34	64.96%	38.65	56.32	63.32%
GRU	35.10	50.66	75.40%	36.25	52.34	73.78%	37.39	52.32	73.31%
LSTM	32.77	46.67	80.55%	33.11	49.75	79.87%	34.67	50.01	78.01%
GCN-GRU	29.64	44.97	82.40%	30.24	45.66	81.76%	30.42	47.02	80.02%
GCN-LSTM	25.54	37.45	86.32%	26.18	39.02	85.89%	27.08	41.23	85.56%
Transformer	23.76	35.66	92.93%	23.77	38.42	91.97%	24.46	39.94	90.74%
BGCSTFFN	21.56	33.46	94.10%	23.29	37.24	93.14	25.84	39.87	93.08%

The prediction performance of the traditional statistical time series model ARIMA is consistent across tasks with different prediction steps. In contrast, deep learning and ma-

chine learning models show greater adaptability across different prediction steps, allowing them to better capture dynamic changes in the time series.

With the exception of traditional statistical models, the prediction error of all deep learning models increases as the prediction step size increases. This trend occurs because longer prediction steps introduce more uncertainties and disturbances that complicate the model's predictions. In addition, the prediction error of the combined model is smaller than that of the individual models because the combined model better combines the strengths of the individual models and thus improves the prediction performance. Overall, the deep learning models perform well in terms of prediction accuracy. Among these models, the GCN-LSTM and GCN-GRU models show a relatively superior prediction performance, emphasizing the superior ability of the GCN layer to extract spatio-temporal features. In contrast, the CNN models perform poorly, as they are unable to efficiently capture the spatio-temporal features present in the URT passenger flow data through convolution. Excellent prediction performance is shown in all tasks, especially in single-step prediction, which is very similar to the prediction capability of the BGCSTFFN proposed in this paper. However, the performance difference gradually expands as the prediction step size increases, which highlights the importance of considering both temporal and spatial correlations in challenging prediction tasks such as URT passenger flow prediction.

It is important to note that datasets with different time intervals have a significant impact on the performance of the model. In the 15 min dataset, the model demonstrates excellent predictive ability, consistently outperforming the 5 min dataset in terms of accuracy. This is due to the high volatility of the data in the 5 min dataset, the significant effect of noise, and the model's difficulty in capturing stable long-term trends. The 15 min dataset is smoother and more cyclical, allowing the model to better identify trends and make stable predictions.

The BGCSTFFN model performs best in terms of prediction accuracy in all prediction tasks. In addition, it has the least variation in performance at various prediction steps, revealing its superior performance in extracting spatio-temporal features and its stability in realizing multi-step prediction tasks.

5.5. Feature Ablation Experiment

The BGCSTFFN model performs best in terms of prediction accuracy in all prediction tasks. In addition, it has the least variation in performance at various prediction steps, revealing its superior performance in extracting spatio-temporal features and its stability in realizing multi-step prediction tasks.

Experimental Group I: BGCSTFFN-No *Adj_A*: Remove the GCN layer that extracts the adjacency feature matrix from the model while keeping all other configurations unchanged.

Experimental Group II: BGCSTFFN-No *R_POI*: Remove the GCN layer that extracts the POI information similarity matrix from the model while keeping all other configurations intact.

Experimental Group III: BGCSTFFN-No *Adj_A* and *R_POI*: Remove from the model the GCN layer that extracts the adjacency feature matrix and the POI information similarity matrix, while keeping all other configurations unchanged.

Experimental Group IV: BGCSTFFN-No *TW*: Remove time label and weather data from the model while keeping the other configurations unchanged.

Experimental Group V: BGCSTFFN-No *Adj_A*, *R_POI* and *TW*: Remove from the model the time labeling with the weather data and the GCN layer responsible for extracting the adjacency feature matrix and the POI information similarity matrix, while keeping all other configurations unchanged. The results of the experiment are shown in Table 9.

Table 9. Results of ablation experiments.

Group	5 min Dataset			15 min Dataset		
	MAE	RMSE	R ²	MAE	RMSE	R ²
Control Group	14.74	24.47	90.38%	24.14	54.27	94.10%
Experiment Group I	12.73	19.37	87.89%	24.22	39.04	93.96%
Experiment Group II	12.90	20.04	87.25%	25.63	40.98	93.59%
Experiment Group III	13.42	21.41	87.18%	24.07	36.80	93.33%
Experiment Group IV	12.67	19.38	88.06%	22.93	35.53	93.98%
Experiment Group V	13.68	21.63	86.16%	23.76	35.66	92.93%

The models in Experimental Groups I, II, and III exhibit a lower prediction performance compared to the control group, indicating that features such as adjacency and POI information similarity have varying degrees of impact on model predictions when removed separately. When both feature types are removed at the same time, the model's performance is more severely affected. In Experimental Group IV, the predictive performance of the model decreases significantly, emphasizing the major role of timestamped data and weather data in the BGCSTFFN model. The model in Experimental Group V performs the worst, highlighting the importance of inputting multiple data sources in the URT passenger flow prediction task.

This comprehensive analysis shows that the key role of various input features in the BGCSTFFN model has been successfully verified by ablation experiments. The experimental results show that the contribution of the features is ranked as follows: the POI information similarity feature, the adjacency feature, and the time label and weather features. This provides an important guideline for the BGCSTFFN model to maintain the excellent predictive performance of URT station passenger flow prediction during peak hours.

5.6. Predictive Performance Analysis for Different Types of Stations

In order to validate the prediction performance of the model at the individual station level, we chose the optimal step combination task for the model, i.e., input step 12 for prediction step 1; selected station serial numbers 4, 15, 20, and 39, which were four different types of stations; and compared and analyzed the prediction results under two different datasets.

Station description: URT station number 4 is a non-interchange commuter station that primarily serves an area located near a medium to large residential neighborhood with relatively limited commercial and office amenities.

URT station number 15 is an important intercity and urban transportation hub where rail, urban rail, and surface public transportation converge to form a key node of the transportation network.

URT station number 20 is a comprehensive interchange station located between a large commercial area and a densely populated residential area, characterized by a high population density in the surrounding area.

URT station number 39 is a non-interchange, office-based subway station located in an area dominated by commercial and office spaces, which also contains some educational facilities and residential areas.

As can be seen from Figures 5–12, the prediction models proposed in this paper perform more accurately in capturing the process of passenger flow changes and the overall trend of different types of stations, and the overall prediction accuracy generally exceeds 85%. This result indicates that the model is able to better capture and predict the changes in passenger flow at different types of stations over different time periods, especially in most cases, and accurately reflects the fluctuations and trends in passenger flow. However,

despite this, there are significant differences in the prediction effects of the same type of stations on different datasets. Specifically, the prediction results obtained from the dataset based on 5 min intervals performed relatively poorly compared to the dataset based on 15 min intervals.

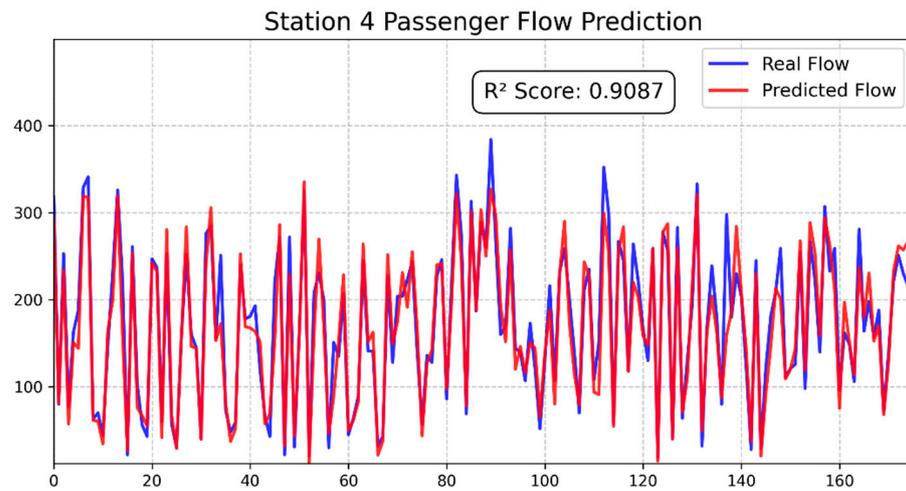


Figure 5. Prediction result of the 5 min dataset for station 4.

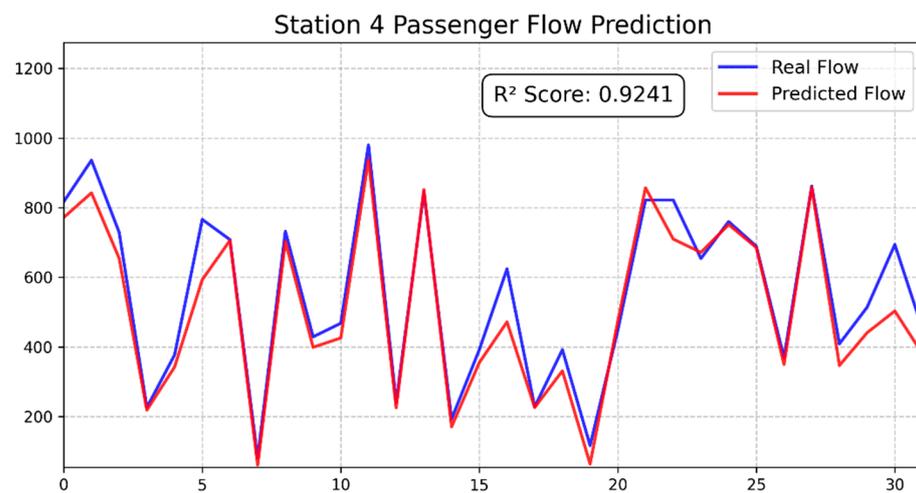


Figure 6. Prediction result of the 15 min dataset for station 4.

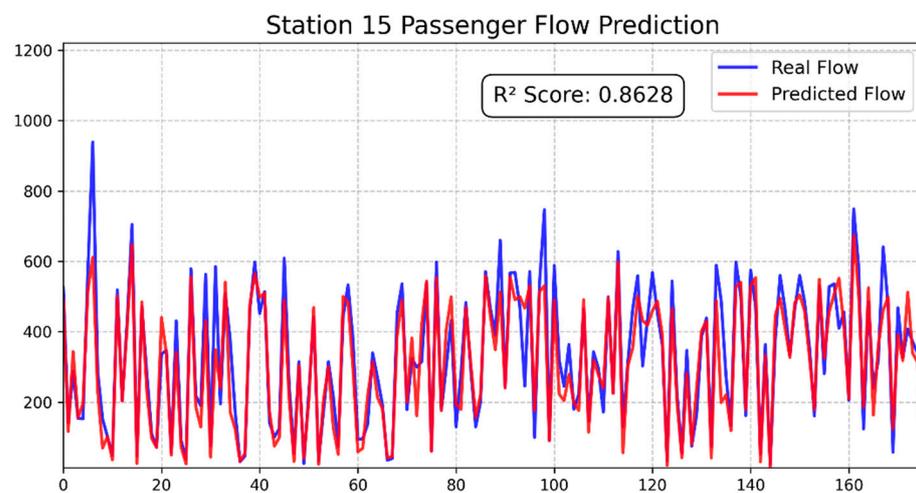


Figure 7. Prediction result of the 5 min dataset for station 15.

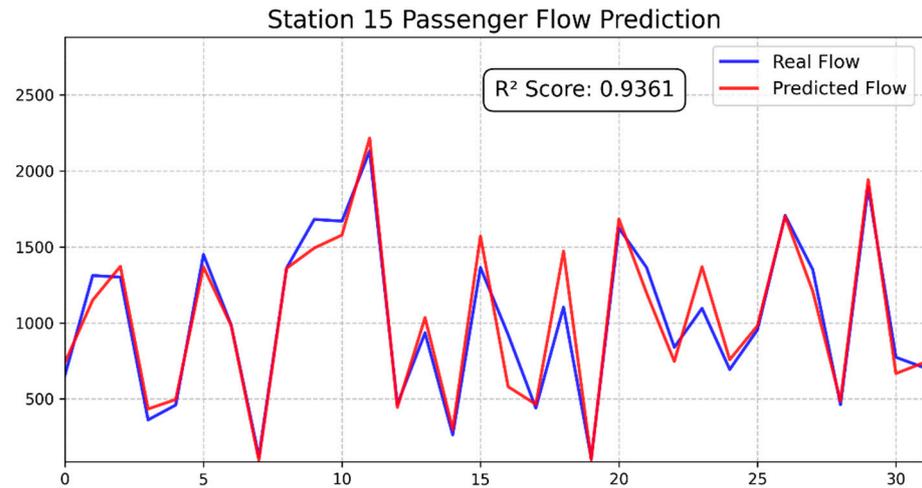


Figure 8. Prediction result of the 15 min dataset for station 15.

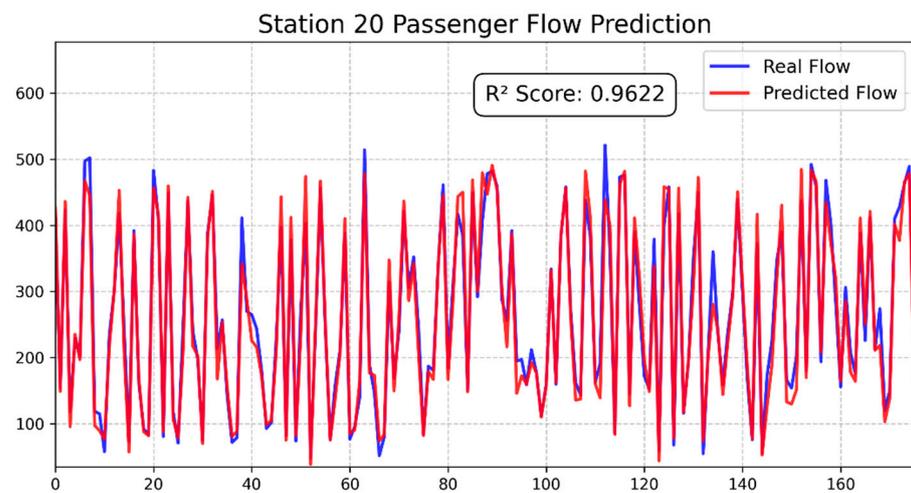


Figure 9. Prediction result of the 5 min dataset for station 20.

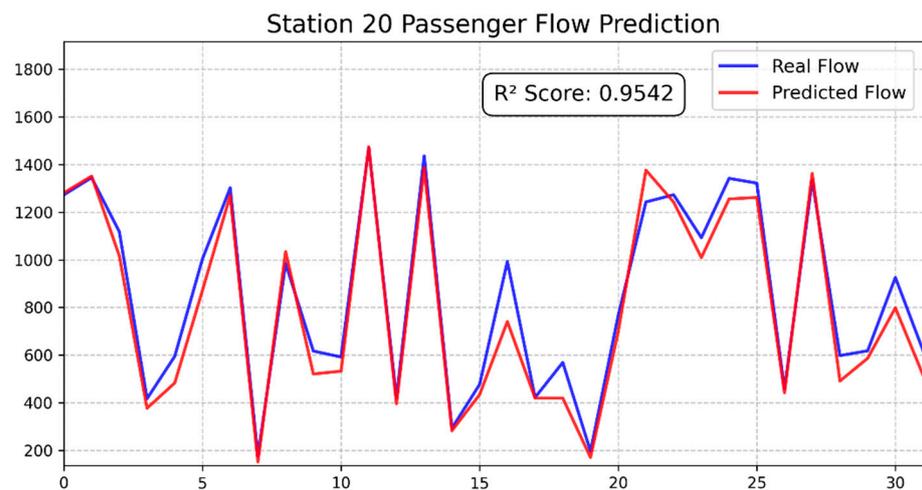


Figure 10. Prediction result of the 15 min dataset for station 20.

The reason for this phenomenon can be attributed to the effect of the time interval of the dataset on the fluctuation of passenger flow. In the dataset with 5 min intervals, the fluctuations in passenger flow changes are more frequent and drastic, especially during peak hours. This frequent fluctuation makes the model face more uncertainties and challenges in trend prediction. Since the data changes in each 5 min time interval are more subtle

and unstable, the model needs to capture the details of these short-term fluctuations more accurately, which is a difficult task for conventional prediction algorithms. As a result, the prediction effect of the dataset for 5 min time intervals is more complex and the prediction accuracy is somewhat compromised.

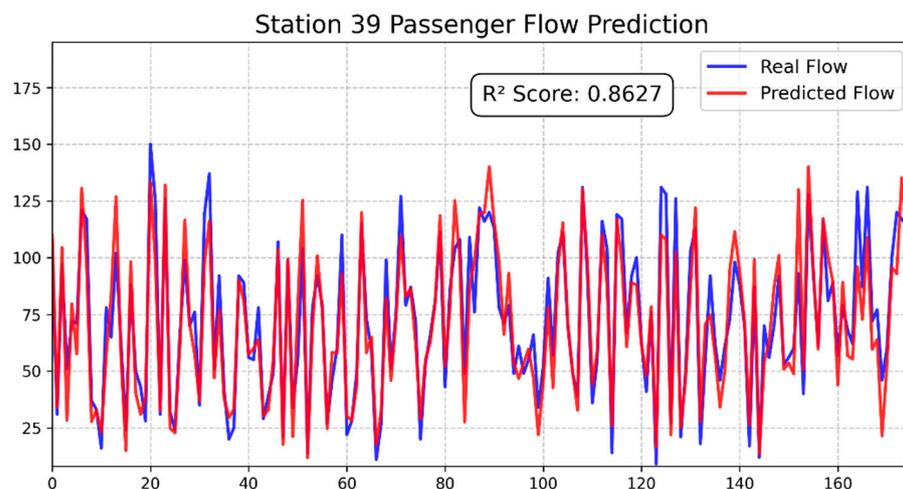


Figure 11. Prediction result of the 5 min dataset for station 39.

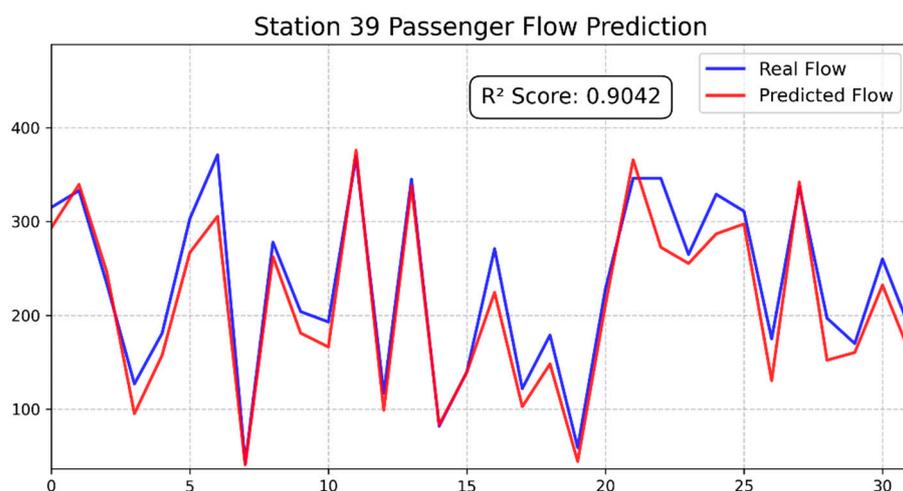


Figure 12. Prediction result of the 15 min dataset for station 39.

In contrast, the dataset with 15 min time intervals is able to present the overall trend of passenger flow more smoothly, reducing the impact of short-term fluctuations. This allows the model to better capture long-term trends and patterns, while ignoring sharp fluctuations over short periods of time, thus improving the accuracy of the predictions. Under longer time intervals, the passenger flow data tend to stabilize, and the model can more effectively identify the regular changes in it, and the prediction results are therefore more accurate.

In addition, the forecast performance accuracy of some stations is more exceptional, especially for Station 39. Due to the complexity and irregularity of passenger flow changes at this station, its prediction accuracy is the lowest among all four categories of stations. The passenger flow at station 39 is affected by a variety of factors, which may include changes in the surrounding environment, temporary activities, etc., leading to fluctuations in its passenger flow that are difficult to capture by conventional prediction models, which in turn affects the accuracy of the prediction results.

Across all datasets, the model demonstrated high accuracy in predicting sudden peaks in passenger flow. In most cases, the model was able to accurately predict the peaks of

mutations and was able to capture significant fluctuations in passenger flow. However, there are still a few peaks that were not predicted with sufficient accuracy, which may be due to the high volatility of the data themselves or the model's failure to adequately adapt to these sudden changes in patterns in specific cases. Nonetheless, in terms of overall performance, the model is quite satisfactory in dealing with sudden change peaks, demonstrating a strong prediction ability. Especially when facing different datasets and changing patterns, the model's performance remains stable with strong generalization ability, which can provide reliable prediction support for practical applications.

For the non-interchange commuter metro station with the URT station serial number 4, the frequency arrangement can be optimized to attract more commuter passengers by strengthening the connection between the surrounding residential areas and commercial facilities, as well as by gradually cultivating a stable passenger flow during morning and evening peak hours in conjunction with the development of residential areas.

For the transportation hub station with the URT station serial number 15, the interchange efficiency and connectivity of the hub station can be enhanced to strengthen cross-modal transportation guidance and divert passenger flows in different directions, so as to avoid the over-concentration of peak hour pressure. At the same time, increasing the number of commercial and retail facilities in the vicinity could attract increased off-peak hour passenger flow and promote all-weather passenger flow distribution.

For the comprehensive interchange station with the URT station serial number 20, it can take advantage of its location between commercial and residential areas to strengthen diversified commercial services and regional supporting facilities to attract the daily commuting high-density flow of people in the surrounding area, and at the same time control the over-concentration of passenger flow by guiding the diversion of short-distance and long-distance passenger flow.

For the office-type metro station with the URT station serial number 39, the frequency of trains during peak hours should be appropriately adjusted according to the mobility characteristics of weekday office crowds to avoid the over-concentration of passenger flows. At the same time, stable passenger flow during off-peak hours can be gradually cultivated with the help of the development of neighboring educational facilities and residential areas to alleviate the pressure of commuting peaks.

Overall, the BGCSTFFN model proposed in this thesis consistently exhibits superior prediction performance in various situations, emphasizing its practicality and value for real-world applications. The BGCSTFF model exhibits excellent adaptability and reliability, and successfully responds to the challenge of predicting short-term, multi-station passenger flow during URT peak hours.

6. Conclusions

In this study, we propose a BGCSTFFN-based model that combines GCN and Transformer techniques to accurately predict short-term passenger flows during URT peak hours. The model integrates inputs from passenger flow time series, POI information similarity coefficients, station adjacency features, and weather and time labeling data, thus providing insights into the complex interactions between temporal and spatial features. Real passenger flow data from the URT in Hangzhou are used to evaluate the prediction performance of the model. The effectiveness of the model in short-term passenger flow prediction at URT stations during peak hours is experimentally verified.

The BGCSTFFN model consistently maintains superior and stable performance in prediction tasks with different combinations of inputs and prediction steps, showing strong robustness and demonstrating an excellent generalization ability. Both longer input steps and shorter prediction steps will help to improve the prediction accuracy. In practical appli-

cations, the appropriate combination of input and prediction steps should be selected based on the task requirements and data characteristics to optimize the prediction performance.

The BGCSTFFN model can effectively reveal the complex coupling relationship between multiple data sources while extracting the spatio-temporal features of passenger flow, and the effects of station adjacency features, POI information similarity features, and time label and weather features on the prediction accuracy are also effectively captured. In addition, the ablation experiments reveal the contribution of each feature to the model prediction accuracy, in which the POI information similarity feature has the greatest impact, followed by the adjacency feature and the time label and weather feature.

Although the current model has used spatial features based on inter-station adjacency and POI similarity, this spatial feature has not yet covered all possible dimensions, and subsequent studies can further introduce more spatial dimension features, such as inter-station distance, etc. In addition, the collection of event data sources can be explored to analyze the interference pattern of random events on URT station passenger flow in order to enhance the accuracy of passenger flow prediction under special events, thereby further improving the prediction performance of the model.

Author Contributions: Conceptualization, J.S.; data curation, X.Y. (Xiaofei Ye); formal analysis, J.S.; funding acquisition, X.Y. (Xiaofei Ye), X.Y. (Xingchen Yan), T.W., and J.C.; investigation, J.S.; writing—original draft, J.S.; writing—review and editing, X.Y. (Xiaofei Ye), X.Y. (Xingchen Yan), T.W., and J.C. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Natural Science Foundation of Zhejiang Province, China (No. MS25E080023), the Natural Science Foundation of Ningbo City, China (No.2024J130), the Fundamental Research Funds for the Provincial Universities of Zhejiang (No. SJLY2023009), the National “111” Center on Safety and Intelligent Operation of Sea Bridge (D21013), National Natural Science Foundation of China (Nos. 71971059, 52262047, 52302388, 52272334, and 61963011), the Natural Science Foundation of Jiangsu Province, China (No. BK20230853), the Specific Research Project of Guangxi for Research Bases and Talents (No. AD20159035), in part by the Guilin Key R&D Program [No. 20210214-1], and the Liuzhou Key R&D Program (No. 2022AAA0103).

Data Availability Statement: 2019 Hangzhou Metro AFC data for eighty-one stations on three lines are sourced from: <https://tianchi.aliyun.com/competition/entrance/231708/information> (accessed on 7 November 2023).

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Chen, Q.; Zhao, J. The use of Ls-SVM for short-term passenger flow prediction. *Transport* **2011**, *26*, 5–10. [[CrossRef](#)]
2. Cao, L.; Liu, S.; Zeng, X.; He, P.; Yuan, Y. Passenger Flow Prediction Based on Particle Filter Optimization. *Mechatron. Robot. Autom.* **2013**, *373–375*, 1256–1260. [[CrossRef](#)]
3. Yan, D.; Zhou, J.; Zhao, Y.; Wu, B. Short-term subway passenger flow prediction based on ARIMA. *Commun. Comput. Inf. Sci.* **2018**, *848*, 464–479.
4. Jiao, P.; Li, R.; Sun, T.; Hou, Z.; Ibrahim, A. Three revised Kalman filtering models for short-term rail transit passenger flow prediction. *Math. Probl. Eng.* **2016**, *2016 Pt 3*, 9717582. [[CrossRef](#)]
5. Li, Q.; Qin, Y.; Wang, Z.Y.; Zhao, Z.X.; Zhan, M.H.; Liu, Y.; Li, Z. The research of urban rail transit sectional passenger flow prediction method. *J. Intell. Learn. Syst. Appl.* **2013**, *5*, 227–231. [[CrossRef](#)]
6. Wei, Y.; Chen, M.C. Forecasting the short-term metro passenger flow with empirical mode decomposition and neural networks. *Transp. Res. Part C Emerg. Technol.* **2012**, *21*, 148–162. [[CrossRef](#)]
7. Sun, Y.; Leng, B.; Guan, W. A novel Wavelet-SVM short-time passenger flow prediction in Beijing subway system. *Neurocomputing* **2015**, *166*, 109–121. [[CrossRef](#)]
8. Tang, L.; Zhao, Y.; Cabrera, J.; Ma, J.; Tsui, K.L. Forecasting short-term passenger flow: An empirical study on Shenzhen metro. *IEEE Trans. Intell. Transp. Syst.* **2018**, *20*, 3613–3622. [[CrossRef](#)]
9. Roos, J.; Gavin, G.; Bonneval, S. A dynamic Bayesian network approach to forecast short-term urban rail passenger flows with incomplete data. *Transp. Res. Procedia* **2017**, *26*, 53–61. [[CrossRef](#)]

10. Singh, A.P.; Tripathi, A.; Dwivedi, R.K.; Garg, A.; Kumar, R. Prediction of passenger flow for north central railway region through ANN. *IOP Conf. Ser. Mater. Sci. Eng.* **2021**, *1136*, 012023. [[CrossRef](#)]
11. Gliovi, N.; Milenkovi, M.; Bojovi, N.; Vadlenka, L.; Avramovi, Z. A hybrid model for forecasting the volume of passenger flows on Serbian railways. *Oper. Res.* **2016**, *16*, 271–285. [[CrossRef](#)]
12. Ma, X.; Zhang, J.; Du, B.; Ding, C.; Sun, L. Parallel architecture of convolutional bi-directional LSTM neural networks for network-wide metro ridership prediction. *IEEE Trans. Intell. Transp. Syst.* **2018**, *20*, 2278–2288. [[CrossRef](#)]
13. Xiong, Z.; Zheng, J.; Song, D.; Zhong, S.; Huang, Q. Passenger flow prediction of urban rail transit based on deep learning methods. *Smart Cities* **2019**, *2*, 371–387. [[CrossRef](#)]
14. Peng, K.; Bai, W.W.; Liu, Y. Passenger flow forecast of railway station based on improved LSTM. In Proceedings of the 2nd International Conference on Advances in Computer Technology, Information Science and Communications 2020, CTISC 2020, Suzhou, China, 20–22 March 2020; pp. 166–1700.
15. Yang, D.; Chen, K.; Yang, M.; Zhao, X. Urban rail transit passenger flow forecast based on LSTM with enhanced long-term features. *Intell. Transp. Syst.* **2019**, *13*, 1475–1482. [[CrossRef](#)]
16. Xiu, C.; Sun, Y.; Peng, Q.; Chen, C.; Yu, X. Learn traffic as a signal: Using ensemble empirical mode decomposition to enhance short-term passenger flow prediction in metro systems. *J. Rail Transp. Plan. Manag.* **2022**, *22*, 100311. [[CrossRef](#)]
17. Li, Y.; Yin, M.; Zhu, K. Short term passenger flow forecast of metro based on inbound passenger plow and deep learning. In Proceedings of the 3rd IEEE International Conference on Communications, Information System and Computer Engineering, CISCE 2021, Beijing, China, 14–16 May 2021; pp. 777–780.
18. Hao, S.; Lee, D.H.; Zhao, D. Sequence to sequence learning with attention mechanism for short-term passenger flow prediction in large-scale metro system. *Transp. Res. Part C Emerg. Technol.* **2019**, *107*, 287–300. [[CrossRef](#)]
19. Guo, J.; Xie, Z.; Qin, Y.; Jia, L.; Wang, Y. Short-term abnormal passenger flow prediction based on the fusion of SVR and LSTM. *IEEE Access* **2019**, *7*, 42946–42955. [[CrossRef](#)]
20. Du, B.; Peng, H.; Wang, S.; Bhuiyan, M.Z.A.; Wang, L.; Gong, Q.; Liu, L.; Li, J. Deep irregular Convolutional Residual LSTM for urban traffic passenger flows prediction. *IEEE Trans. Intell. Transp. Syst.* **2019**, *21*, 972–985. [[CrossRef](#)]
21. Wang, B.; Ye, M.; Zhu, Z.; Li, Y.; Zhang, J. Short-term passenger flow prediction for urban rail stations using learning network based on optimal passenger flow information input algorithm. *IEEE Access* **2020**, *8*, 170742–170753. [[CrossRef](#)]
22. Chen, X.; Xie, X.; Teng, D. Short-term traffic flow prediction based on ConvLSTM Model. In Proceedings of the IEEE 5th Information Technology and Mechatronics Engineering Conference (ITOEC), Chongqing, China, 12–14 June 2020; pp. 846–850.
23. Li, S.; Liang, X.; Zheng, M.; Chen, J.; Chen, T.; Guo, X. How spatial features affect urban rail transit prediction accuracy: A deep learning based passenger flow prediction method. *J. Intell. Transp. Syst.* **2024**, *28*, 1032–1043. [[CrossRef](#)]
24. Wang, D.; Gao, R.; Su, W. Short-term inbound passenger flow prediction of urban rail transit network based on RS-Conv1dGRU. In Proceedings of the 2nd International Conference on Internet of Things and Smart City (IoTSC 2022), Xiamen, China, 18–20 February 2022.
25. Ye, J.; Zhao, J.; Ye, K.; Xu, C. Multi-STGCnet: A graph convolution based spatial-temporal framework for subway passenger flow forecasting. In Proceedings of the 2020 International Joint Conference on Neural Networks (IJCNN), Glasgow, UK, 19–24 July 2020.
26. Zhang, J.; Chen, F.; Cui, Z.; Guo, Y.; Zhu, Y. Deep Learning Architecture for Short-Term Passenger Flow Forecasting in Urban Rail Transit). *IEEE Trans. Intell. Transp. Syst.* **2021**, *22*, 7004–7014. [[CrossRef](#)]
27. Zhao, L.; Song, Y.; Zhang, C.; Liu, Y.; Wang, P.; Lin, T.; Deng, M.; Li, H. T-GCN: A Temporal Graph Convolutional Network for Traffic Prediction). *IEEE Trans. Intell. Transp. Syst.* **2020**, *21*, 3848–3858. [[CrossRef](#)]
28. Xia, D.; Shen, B.; Geng, J.; Hu, Y.; Li, Y.; Li, H. Attention-based spatial-temporal adaptive dual-graph convolutional network for traffic flow forecasting. *Neural Comput. Appl.* **2023**, *35*, 17217–17231. [[CrossRef](#)]
29. Zhang, C.H.; Yu, J.J.Q.; Liu, Y. Spatial-temporal graph attention networks: A deep learning approach for traffic forecasting. *IEEE Access* **2019**, *7*, 166246–166256. [[CrossRef](#)]
30. Zhang, J.L.; Chen, Y.J.; Panchamy, K.; Jin, G.; Wang, C.; Yang, L. Attention-based multi-step short-term passenger flow spatial-temporal integrated prediction model in URT systems. *J. Geo-Inf. Sci.* **2023**, *25*, 698–713. [[CrossRef](#)]
31. Liu, L.; Liu, Y.; Ye, X. Multi-sequence spatio-temporal feature fusion network for peak-hour passenger flow prediction in urban rail transit. *Transp. Lett.* **2024**, *17*, 86–102. [[CrossRef](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.