

# SPATIOTEMPORAL GRAPH CONVOLUTIONAL NEURAL NETWORKS FOR METRO FLOW PREDICTION

Shiyuan Jin<sup>1</sup>, Changfeng Jing<sup>1\*</sup>, Yi Wang<sup>1</sup>, Xinxin Lv<sup>1</sup>,

<sup>1</sup> School of Geomatics and Urban Spatial Informatics, Beijing University of Civil Engineering and Architecture, Beijing 100044, China – (S.J.) 2108160320003@stu.bucea.edu.cn, (C.J.) jingcf@bucea.edu.cn, (Y.W.) 2108160119006@stu.bucea.edu.cn, (X.L.) 2108570020068@stu.bucea.edu.cn

Commission IV, WG IV/8

**KEY WORDS:** Metro Flow Forecasting, Citywide Metro Network, Spatiotemporal Graph Neural Network, Spatial Dependence, Temporal Dependence, Deep Learning.

## ABSTRACT:

Forecasting urban metro flow accurately plays an important role for station management and passenger safety. Owing to the limitations of non-linearity and complexity of traffic flow data, traditional methods cannot satisfy the requirements of effectively capturing spatiotemporal dependencies at the metro network level, which makes it difficult to demonstrate high performance. In this paper, a novel deep learning method is proposed based on Graph Neural Networks (GNN), named STGCN-Metro (SpatioTemporal Graph Convolutional Network based on Metro network), to forecast the short-term inflow and outflow volumes of metro passengers. The proposed model is composed of two spatiotemporal convolutional blocks, which is integrated with the Dilated Convolutional Neural Network (DCNN) and Cluster-Graph Convolutional Network (Cluster-GCN). The DCNN is employed with different dilation rates to capture temporal dependence in larger receptive field. In addition, compare with GCN, the Cluster-GCN is applied the graph clustering algorithms to reduce computational resources considering spatial heterogeneity. A real-world dataset collected in Shanghai metro stations is conducted for validation, and the results demonstrate that the proposed model achieves higher performance, outperforming some well-known baseline models.

## 1. INTRODUCTION

Predicting urban metro flow is one of essential mission of the Intelligent Transportation Systems (ITS). In particular, realizing prediction of flow volume accurately can help passengers to make travel planning in advance, alleviate traffic pressure, reasonably allocate metro staff and coordinate operation schedules (Wang et al., 2021). Therefore, it is fundamental to achieve accurate prediction for metro network.

Urban metro flow forecasting is a typical spatiotemporal prediction problem. Passenger flow-volume refers to the number of metro card-swiping at each station during a fixed period of time. From the perspective of spatial, taking the fixed stations as nodes and complex metro lines as connection edges, forms a graph structure in the network of metro stations. Different stations have different degrees of impact on the same station. From the perspective of temporal, the flow-volume has varying records on certain station as time goes by. Besides, there are certain patterns in the inflow and outflow volumes aspect. The flow-volume data of different stations has varying influence at different times in the future. In a word, the correlations of flow-volume in metro network reflect powerful dynamics in both the spatial perspective and temporal perspective. It means that metro flow forecasting should consider two aspects: spatial and temporal. How to structure models to explore spatiotemporal patterns and extract spatiotemporal correlations in complex passenger flow data is an extremely challenging problem.

With the development of computer technology, more time series data with geospatial information has been obtained in the field of transportation, so that more and more scholars have done a lot of

researches on the prediction of these data. Early, classic statistical methods were adopted to forecast traffic flow, but the methods constrain the highly non-linear feature representability of traffic flow data (Yu et al., 2017). Later, the traditional machine learning methods have been further applied to complete traffic prediction tasks. These methods are based on well-trained samples to forecast non-linear traffic flow. However, the performance of methods heavily relies on manually extracting feature engineering (Guo et al., 2019). Hence, there are difficult to yield the best performance. Recently, a growing number of researchers are applying deep learning methods to handle high-dimensional spatiotemporal data, i.e., Recurrent Neural Network (RNN) (Connor et al., 1994) are employed to capture temporal dynamic, but ignore spatial dependence; Convolutional Neural Network (CNN) (Krizhevsky et al., 2012) is employed to extract spatial features of grid-based data. However, there are topological relationships and flow correlations between adjacent metro stations. Therefore, the convolutional operations based on regular grid cannot accurately capture spatiotemporal dependencies at the metro network level (Han et al., 2019), which makes it difficult to obtain higher prediction accuracy.

In order to overcome these challenges, we present a novel deep learning method based on Graph Neural Network (GNN), named STGCN-Metro (SpatioTemporal Graph Convolutional Network based on Metro network), which is used for passenger flow forecasting based on metro network. The main contributions are as follows:

(1) The STGCN-Metro model designs a novel spatiotemporal block, which is integrated with dilated convolutional neural

\* Corresponding author, Changfeng Jing jingcf@bucea.edu.cn

network and cluster-graph convolutional network. The former is adopted to capture the temporal dynamic of traffic flow to model temporal dependence. The latter is adopted to capture the spatial topological structure of metro network to model spatial dependence.

(2) The prediction results of the STGCN-Metro model show excellent performance under inflow and outflow volumes of passengers, which indicates the short-term forecasting ability of model in complex and different traffic flow.

(3) The proposed approach is evaluated on publicly-available real-world dataset collected in Shanghai metro stations. Experimental results demonstrate our model outperforms other well-known baselines.

## 2. RELATED WORK

Passenger flow forecasting is an important but challenging task, for which a number of scholars have conducted extensive researches. The researches methods can be divided into three categories: the statistical methods, the machine learning methods and the deep learning methods (Han et al., 2021).

### 2.1 Statistical Methods

The statistical methods belong to mathematical analysis methods that predict future values based on historical time series, which are used to predict traffic flow, such as History Average (HA) model (Brian and Michael, 1997), Kalman filtering (Liang et al., 2019), AutoRegressive Integrated Moving Average (ARIMA) model (Ahmed and Cook, 1979) and its variations the Seasonal AutoRegressive Integrated Moving Average (SARIMA) model (Ni et al., 2017), etc. The HA model obtains prediction results by calculating average values of traffic flow in history periods, which is commonly used in baseline model because of simple calculation. In addition, the ARIMA model has been frequently used to deal with time series, especially for passenger flow forecasting. Study proved the ARIMA model has fine results though simulation experiment both in accuracy and effect (Liu et al., 2021). Moreover, the emergence of hybrid models leads to a certain improvement in prediction. Li et al. combined symbolic regression and ARIMA to model different forms of relationships under the passenger flow dataset, the evaluation results indicate the hybrid model outperforms the ARIMA model (Li et al., 2018). In view of the dynamic volatility and nonlinearity of the metro passenger flow, Chen et al. integrated the Generalized AutoRegressive Conditional Heteroscedasticity (GARCH) and ARIMA to model mean and volatility of passenger flow to achieve the short-term prediction (Chen et al., 2020b). These methods have simple algorithm and capture linear feature successfully. However, they rely on the stationary assumption and fail to consider spatiotemporal correlations. Therefore, these methods tend to constrain the highly non-linear feature representability of traffic flow data, while the emergence of machine learning brings new opportunities.

### 2.2 Machine Learning Methods

The machine learning methods are able to reflect complex non-linear relations of traffic flow data, which can make up the flaws of statistical methods. The representative works are K-Nearest Neighbor (KNN) (Pinlong et al., 2016), Support Vector Regression (SVR) (Smola and Schlkopf, 2004), neural networks (Kranti et al., 2013), etc., which are commonly compared to time series models. For example, Tang et al. employed ARIMA, linear

regression and SVR to compare prediction effect of short-term passenger flow (Tang et al., 2019). Wang et al. proposed a novel hybrid model based on Support Vector Machine (SVM) to capture periodic and non-linear features of passenger flow, compared with SARIMA and SVM, the model has the best performance on unstable weekend and holiday (Wang et al., 2018). Liu et al. proposed based on the modified Least-square SVM model for forecasting passenger flow on holiday, results demonstrate the model outperforms the ARIMA (Liu and Yao, 2017). To consider the trend factor and time interval factor of passenger flow, Bai et al. adopted enhanced KNN method for passenger flow forecasting, which gains better performance than original KNN method (Bai et al., 2019). These methods can extract non-linear feature for complicated traffic flow. However, there is still difficult in simultaneously considering spatiotemporal correlations of high-dimensional traffic flow. Besides, the prediction performance of approaches heavily relies on feature engineering that requires abundant expert experiences in the corresponding domain.

### 2.3 Deep Learning Methods

Compared with the machine learning algorithms, there are deep learning methods enable to automatically model more complicated dependencies, which has made the focus on modeling complicated spatiotemporal data (Aqib et al., 2019). The RNN is suitable for handling complicated time series to model dynamic temporal dependence. Its successors such as Long Short-Term Memory (LSTM) (Hochreiter and Schmidhuber, 1997), Gated Recurrent Unit (GRU) (Chung et al., 2014) are applied for prediction tasks. Considering the irregular fluctuation of passenger flow caused by various factors, the single LSTM method cannot make great predictions. As a result, scholars proposed LSTM in combination with other methods, such as Seasonal-Trend decomposition based on Loess (Chen et al., 2020a), stacked auto-encoders (Jia et al., 2019), wavelet (Yang et al., 2021), evaluation results indicate the hybrid model has superior performance than single LSTM. However, these methods mentioned only the temporal dependence, while neglect spatial dependence. The CNN was originally designed for regular grids data (i.e., images). Recently, the CNN models have been used to capture the spatial dependence in Euclidean space. Zhang et al. proposed CNN-based named DeepST model to forecast flow-volume (Zhang et al., 2016). Since then, CNN-based models have been frequently applied in passenger flow prediction. In order to capture the spatiotemporal dependencies, Zhang et al. aggregated the output of the three residual neural networks to construct ST-ResNet model, and achieved prediction of regional traffic flow (Zhang et al., 2017). Chen et al. proposed Conv-LSTM hybrid model to extract spatiotemporal features for solving the short-term prediction problem of metro congestion delay (Chen et al., 2021b). However, these models need to be applied for regular grid data, while the traffic flow is essentially a graphical data (Wang and Jing, 2022). There are topological relationships and flow correlations between adjacent metro stations. Therefore, the operation cannot accurately capture the spatiotemporal dependencies in metro network. With the development of GNN, some models have appeared, such as Graph Convolutional Networks (GCN) (Kipf and Welling, 2016), Chebyshev networks (ChebNet) (Defferrard et al., 2016), diffusion convolutional neural networks (Atwood and Towsley, 2016), etc. They are more suitable for capturing spatial dependence due to that preserve the realistic topological information. Meanwhile, the GCN achieves greatly preserve the metro network globality by convolving the whole structured graphs, which is theoretically superior to CNN that can only capture adjacent spatial pattern because of the limited kernel

window size (Chen et al., 2021a). Hence, the GNN-based models can be applied to effectively capture the irregular spatiotemporal dependencies at metro network rather than CNN-based models.

Based on aforementioned background, we proposed a novel deep learning method based on GNN that can accurately capture spatiotemporal dependencies at the metro network level.

### 3. METHODOLOGY

#### 3.1 Metro Flow Prediction on Graphs

In this work, the metro network is defined on graph. The metro stations are used as nodes and the connections between stations as edges of the graph. At the time step  $t$ , in graph  $G = (V_t, E, W)$ ,  $V_t$  is a set of vertices, each of which is denoted observation vector of  $n$  stations at time step  $t$ ;  $E$  is a set of edges connected by stations;  $W \in R^{N \times N}$  denotes the weighted adjacency matrix. The graph  $G$  processes a feature matrix  $X_t \in R^{N \times F}$  at each time step  $t$ . Our problem is to forecast the passenger flow in next  $T$  time intervals given the graph  $G$  and historical  $s$  step values. We regard the  $n$  stations passenger flow information on metro network as input data, expressed as  $X$ :

$$X = \begin{bmatrix} X_{t-s}, X_{t-s+1}, \dots, X_t \\ x_{t-s}^1 & x_{t-s+1}^1 & \dots & x_t^1 \\ x_{t-s}^2 & x_{t-s+1}^2 & \dots & x_t^2 \\ \dots & \dots & \dots & \dots \\ x_{t-s}^n & x_{t-s+1}^n & \dots & x_t^n \end{bmatrix} \quad (1)$$

Where  $X_{t-i}$  represents the flow vector counted at the  $i_{th}$  time intervals before  $t$  timestamp, and  $x^j$  denotes the passenger flow of  $j_{th}$  station. Besides, the function  $f$  maps relation between input and output, which is represented as follows:

$$[X_{t+1}, X_{t+2}, \dots, X_{t+T}] = f([X_{t-s}, X_{t-s+1}, \dots, X_t], G) \quad (2)$$

#### 3.2 Network Architecture

This paper defines graph structure based on the metro network and focus on constructing model about spatiotemporal graph convolutional network of passenger flow. As shown in Figure 1, the framework STGCN-Metro is a combination of two spatiotemporal convolutional blocks (ST-Conv Blocks) and fully-connected output. Each of ST-Conv Blocks contains two temporal convolutional layers and one spatial convolutional layer. The details of method are described as follows.

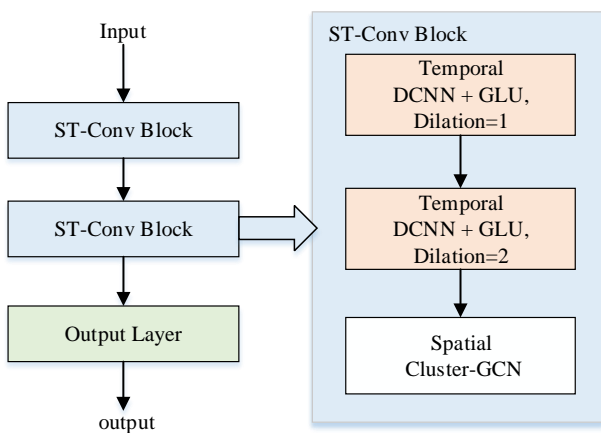


Figure 1. The framework of STGCN-Metro.

#### 3.3 Temporal Dependence Modeling

We adopt the Dilated Convolutional Neural Network (DCNN) (Yu and Koltun, 2016) as temporal convolution layer to capture temporal dependence. Although RNN-based methods are commonly used in time series analysis, the method suffers from complicated gating mechanisms and time-consuming iterations (Yu et al., 2017). Conversely, convolutional neural network is simple in structure as well as efficient in training. The DCNN adds dilation to standard convolution map, which allows to increase large receptive field for wider range of information and reduce calculation complexity (Sun et al., 2020). In this work, stacking two convolution layers with different dilation rates can capture multi-scale contextual information, which not only reduces computational resources, but also increases the sensitivity of the temporal component.

In addition, the Gated Linear Units (GLU) are proved powerful to control information flow (Dauphin et al., 2017). Therefore, the GLU are used for convolutional network in temporal layer. They can retain the non-linear characteristics of complex dataset (Xu et al., 2021). Mathematically, given the input  $X \in R^{N \times F \times T}$ , representing  $f$  feature matrix of  $n$  points in  $t$  time series, the gating mechanism is represented as:

$$H = g(X * k_1 + b) \odot \sigma(X * k_2 + c) \quad (3)$$

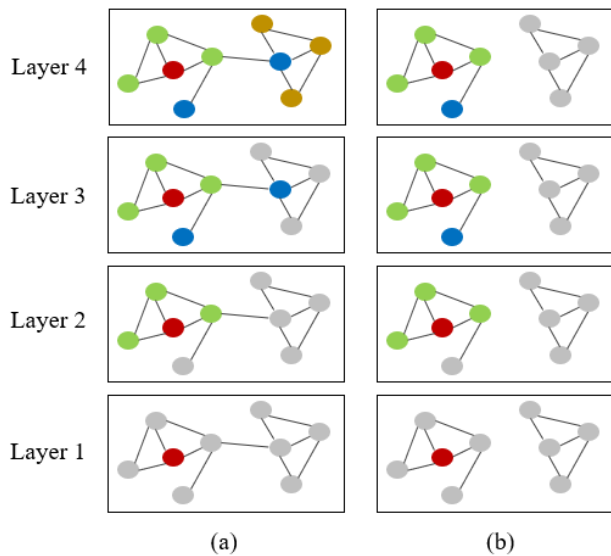
Where  $k_1, k_2$  are convolutional kernels,  $b, c$  are the model learnable parameters,  $g(\cdot)$  is the activation function,  $\sigma(\cdot)$  is the sigmoid function that controls the pass of information to next layer, and  $\odot$  is the element-wise Hadamard product. We take the DCNN as convolutional approach, and then through gating  $A \odot \sigma(B)$  gains output layer in this paper, which is able to capture complicated temporal dependence.

#### 3.4 Spatial Dependence Modeling

We adopt Cluster-Graph Convolutional Network (Cluster-GCN) as spatial convolution layer to acquire spatial dependence. The traditional methods, such as CNN and CNN-based variants, cannot reflect the complex spatial topological structure well for metro network. In contrast, the GCN is often applied by scholars due to its ability to process graph structured data (Zhao et al., 2020). After that, (Chiang et al., 2019) proposed the expansion of GCN, named Cluster-GCN. The graph nodes clustering algorithm is employed to divide the nodes of the graph into multiple clusters, the nodes of these clusters and the corresponding edges are formed into subgraphs, and then using subgraphs to train separately. (Chiang et al., 2019) shows the difference between traditional graph convolution and Cluster-GCN approach. In Figure 2, it defines a red node as starting node, which is used for neighbourhood nodes expansion. As the number of layers increases, eventually a forward propagation of the red node needs to use all nodes of the graph, while Cluster-GCN cuts the original large graph into two small subgraphs, reducing the memory pressure. In comparison with GCN, that is shown to be significant to improve the utilization by sampling the subgraph and limiting neighbourhood search within the subgraph (Chiang et al., 2019). Besides, the spatial heterogeneity is considered, and the phenomenon of over-smoothing is less frequent. Therefore, by approaching Cluster-GCN to capture spatial dependence with less layers reduces computational resources and guarantees extraction accuracy. A  $l$ -layer Cluster-GCN can be expressed as:

$$X^{l+1} = \sigma \left( (\tilde{A} + \lambda \text{diag}(\tilde{A})) X^l W^l \right) \quad (4)$$

Where  $X^l$  and  $X^{l+1}$  denotes feature input and output matrix respectively,  $\hat{A} = (D + I)^{-1}(A + I)$  represents to add identity to  $A$  and normalization.  $A$ ,  $D$  denotes adjacency matrix and degree matrix respectively.  $W^l$  is the weighted matrix in  $l$ -layer,  $\lambda$  is the learnable parameter, and  $\sigma(\cdot)$  is the sigmoid function.



**Figure 2.** The neighbourhood expansion contrast of (a) traditional graph convolution and (b) Cluster-GCN model.

In addition, geographically speaking, we take the distance among stations as an important factor in the Cluster-GCN module. The adjacency matrix of metro graph is calculated by normalizing and exponential function the distances in the metro network. The adjacency matrix with weighted  $W$  is formed as:

$$W_{ij} = \begin{cases} \exp\left(\frac{d_{ij}-\mu}{\sigma}\right), & i \neq j \\ 0, & \text{otherwise} \end{cases} \quad (5)$$

Where  $d_{ij}$  denotes the distance along the metro network between station  $i$  and station  $j$ , which decides the weight of edge  $W_{ij}$ .  $\mu$  and  $\sigma$  is the mean and standard deviation of distances respectively.

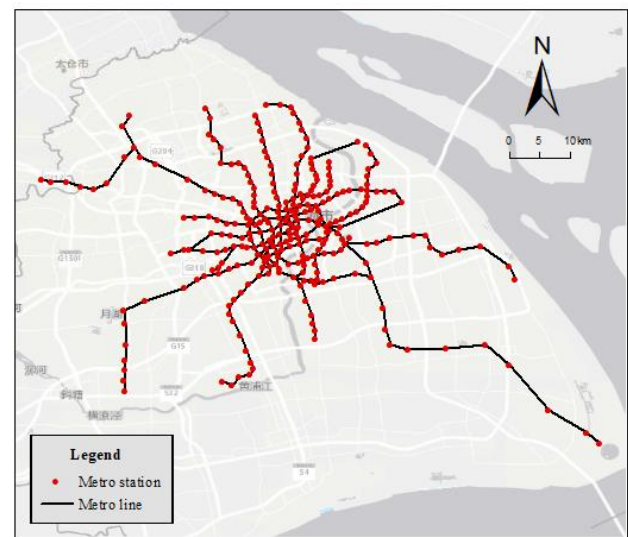
In summary, the STGCN-Metro model is feasible in handling spatiotemporal dependencies. On the one hand, the DCNN with gating mechanism is used to obtain temporal feature of passenger flow information for capturing the temporal dependence. On the other hand, the Cluster-GCN is used to obtain complex spatial topological structure of metro stations network for capturing the spatial dependence. And stacking these spatiotemporal convolutional layers builds two ST-Conv Blocks to achieve flow prediction task.

#### 4. STUDY AREA AND DATASET

In this paper, Shanghai, a developing rapidly city in China, is selected as our study area. The metro stations have become a network and surrounded by commercial areas, education areas, and residential areas, etc., so the passenger flow is highly concentrated. There were 14 metro running lines with 289 metro stations for metro system in 2015, as shown in Figure 3.

The metro flow datasets including inflow and outflow volumes are counted by card-swiping data of Shanghai, from Apr.1 to

Apr.30, 2015. The card-swiping data provides lots of information for metro operators including passenger ID, transaction date, transaction time, current station, transaction amount and transaction nature (e.g., preferential, non-preferential). It appears about nine million records per day. When the transaction amount is equal to 0, we define this card-swiping record as inbound status, otherwise, it is defined as outbound status. Taking the metro operating schedules and human activities into account, we select the time between 6:00 and 23:00 as study period. The passenger flow of stations is aggregated in 5-minutes intervals, so each station contains 204 time points per day.



**Figure 3.** Shanghai metro network.

## 5. EXPERIMENT

### 5.1 Experimental Settings

All experiments are implemented with Python3 programming language and the Pytorch deep learning packages. For the inflow and outflow volumes prediction, the settings of two ST-Conv Blocks in STGCN-Metro model are equivalent. The dilation rates of the DCNN model are 1, 2 respectively. And the Cluster-GCN model layer is set to 1. In this work, the Mean Square Error (MSE), reflecting the deviation of the ground truth and prediction value, is employed as loss function and minimized for 100 epochs. The batch size is set to 64 and the learning rate is set to 0.001. It is worth noting that the same parameters are also taken in the baseline models. Additionally, 80% of the flow-volume is utilized for training and the 20% is utilized for testing. It means that the passenger flow before Apr.24 is applied as training set, which contains eighteen weekdays and six weekends. And the remaining six days after Apr.24 is applied as testing set, which contains four weekdays and two weekends. Furthermore, we use the last 12 window values to predict the next window values, that is to say, 1-hour records are used to predict flow-volume in the next 5 minutes.

### 5.2 Evaluation Metrics

In order to evaluate the performances of the proposed model and baselines, the two widely metrics that Mean Absolute Error (MAE) and Root Mean Squared Error (RMSE) are chosen. The metrics calculations are expressed as:

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \quad (6)$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad (7)$$

Where  $n$  is the values number,  $y_i$  and  $\hat{y}_i$  are passenger flow, representing the ground truth and prediction value respectively. In generally, the smaller the results obtained by the above two formulas, the smaller the prediction errors of the model.

### 5.3 Baseline Methods

We compare STGCN-Metro model with the following six baseline models:

HA (Brian and Michael, 1997): Historical Average method. The last an hour average value of passenger flow-volume as the prediction value.

LSTM (Hochreiter and Schmidhuber, 1997): Long Short-Term Memory network, a special form of RNN.

GRU (Chung et al., 2014): Gated Recurrent Unit network, a special form of RNN.

GCN (Kipf and Welling, 2016): Graph Convolutional Network, a basis model of traffic prediction, which is composed of pre-defined graph convolutional layer components.

TGCN (Zhao et al., 2020): Temporal Graph Convolutional Network, which processes temporal features with GRU on the basis of graph convolution.

STGCN (Yu et al., 2018): SpatioTemporal Graph Convolution Network. The graph convolution and 1D convolution networks with GLU function are adopted to the spatial and temporal feature extraction respectively.

## 6. RESULT AND DISCUSSION

In this paper, two evaluation metrics are selected as the basis for our experiment. We compare the proposed model and baselines in inflow and outflow volumes prediction, the results are shown in Table 1.

Model	Inflow		Outflow	
	MAE	RMSE	MAE	RMSE
HA	21.70	46.45	27.76	61.18
GRU	18.12	44.96	27.35	65.61
LSTM	15.57	42.87	23.83	61.49
GCN	12.34	22.48	19.64	36.99
TGCN	12.68	28.72	19.08	44.03
STGCN	11.92	21.95	18.70	35.69
STGCN-Metro	11.65	21.23	17.87	34.73

Table 1. Results comparison of different methods.

The following can be thought: (1) The STGCN-Metro processes the best performance in evaluation metrics, which demonstrates the advantages of this model in extracting spatiotemporal dependencies. (2) The GRU and LSTM models achieve better prediction precision than statistical model, such as the HA model, since take the extraction temporal dependence into account. (3) Compared with the GRU and the LSTM models, the neural network-based models, including the GCN model, the TGCN model and the STGCN model, which emphasize topological

relationships and flow correlations between adjacent metro stations, generally have higher prediction results. (4) The MAEs of the GCN model are closed to the TGCN model whether inflow or outflow volume. However, the RMSEs of the GCN model are reduced by approximately 21.7% and 16.0% compared with the TGCN model in the inflow and outflow prediction respectively, which means that applying GRU to extract temporal dependence based on GCN cannot be combined well. (5) The MAEs and RMSEs of the STGCN model are the second smallest values both in the inflow and outflow volumes. It shows that the feasibility to consider temporal and spatial dependencies separately on graph neural networks, but the method employed can be further optimized.

From the test results of inflow and outflow prediction in Table 1, it can be seen the STGCN-Metro model, the STGCN model and the GCN model perform well. As a result, three models are chosen to explore more details. In this work, we take the crowded Xujiahui station in weekday as an example, Figure 4 and Figure 5 show the performance of three models for inflow and outflow volumes prediction. The overall fluctuation trend in predictions looks roughly similar. However, the model is superior to baselines in the morning and evening peaks, such as the evening peak of inflow volume around 18:00, the morning and evening peaks of outflow volume around 9:00 and 18:30. It shows that the proposed model is more sensitive than other baseline models in predicting peak passenger flow. At the same time, it also proves that the effectiveness of the model in spatiotemporal prediction. The property is meaningful for predicting station congestion and protecting metro operations.

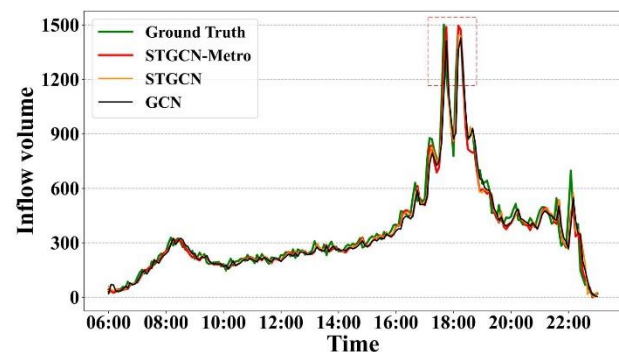


Figure 4. Inflow volume prediction in weekday.

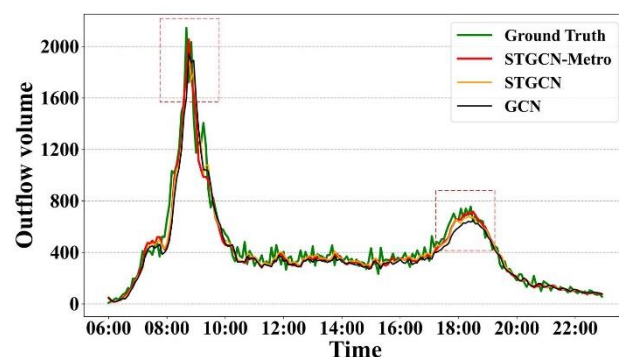


Figure 5. Outflow volume prediction in weekday.

## 7. CONCLUSION

In this paper, we propose a novel deep learning method based on graph neural networks, named STGCN-Metro, to predict the short-term passenger inflow and outflow volumes at the citywide



metro network. The model is a combination of two spatiotemporal convolutional blocks, each of which contains two temporal convolutional layers and one spatial convolutional layer. In temporal convolutional layer, the DCNN model is employed with different dilation rates to capture temporal dependence in larger receptive field. In spatial convolutional layer, compare with GCN, the Cluster-GCN is applied the graph clustering algorithms to reduce computational resources considering spatial heterogeneity. The effectiveness of the model is proved by real-world dataset collected in Shanghai metro stations. Experimental results demonstrate our model outperforms other well-known baselines (i.e., HA, GRU, LSTM, GCN, TGCN, STGCN). The model achieves end-to-end prediction that can accept the raw format of input data and automatically extract the feature from passenger flow. Besides, it has ability to capture spatiotemporal dependencies to gain better prediction accuracy. This work can not only assist passengers to make travel planning in advance, but also provide references for relevant metro departments to make effective plans. In the future, it is worth to consider other factors that affect passenger flow, such as urban structure, population density, weather, etc., to achieve more accurate prediction.

#### ACKNOWLEDGMENTS

This research was funded by the Beijing Natural Science Foundation (Grant #8222009), Training Programme for the Talents by Xicheng, Beijing (Grant #202137) and The Pyramid Talent Training Project of BUCEA (Grant #JDJQ20200306).

#### REFERENCES

- Ahmed, M. S., Cook, A. R., 1979: Analysis of Freeway Traffic Time-Series Data by Using Box-Jenkins Techniques, *Transportation Research Record*, 1-9.
- Aqib, M., Mehmood, R., Alzahrani, A., Katib, I., Albeshri, A., Altowaijri, S. M., 2019: Smarter Traffic Prediction Using Big Data, In-Memory Computing, Deep Learning and GPUs, *Sensors*, 19, 2206. doi.org/10.3390/s19092206.
- Atwood, J., Towsley, D., 2016: Diffusion-Convolutional Neural Networks, *Proceedings of the 30th International Conference on Neural Information Processing Systems*, 2001–2009.
- Bai, J., He, M., Shuai, C., 2019: Short-Term Passenger Flow Forecast in Urban Rail Transit Based on Enhanced K-Nearest Neighbor Approach, *19th COTA International Conference of Transportation Professionals*.
- Brian, L. S., Michael, J. D., 1997: Traffic Flow Forecasting: Comparison of Modeling Approaches, *Journal of Transportation Engineering*, 123, 261-266.
- Chen, D. W., Zhang, J. H., Jiang, S. X., 2020a: Forecasting the Short-Term Metro Ridership With Seasonal and Trend Decomposition Using Loess and LSTM Neural Networks, *IEEE Access*, 8, 91181-91187. doi.org/10.1109/access.2020.2995044.
- Chen, E. H., Ye, Z. R., Wang, C., Xu, M. T., 2020b: Subway Passenger Flow Prediction for Special Events Using Smart Card Data, *IEEE Transactions on Intelligent Transportation Systems*, 21, 1109-1120. doi.org/10.1109/tits.2019.2902405.
- Chen, P. F., Fu, X. D., Wang, X., 2021a: A Graph Convolutional Stacked Bidirectional Unidirectional-LSTM Neural Network for Metro Ridership Prediction, *IEEE Transactions on Intelligent Transportation Systems*. doi.org/10.1109/TITS.2021.3065404.
- Chen, W., Li, Z. P., Liu, C., Ai, Y., 2021b: A Deep Learning Model with Conv-LSTM Networks for Subway Passenger Congestion Delay Prediction, *Journal of Advanced Transportation*. doi.org/10.1155/2021/6645214.
- Chiang, W.-L., Liu, X., Si, S., Li, Y., Bengio, S., Hsieh, C.-J., 2019: Cluster-GCN: An Efficient Algorithm for Training Deep and Large Graph Convolutional Networks, *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*. doi.org/10.1145/3292500.3330925.
- Chung, J., Gulcehre, C., Cho, K., Bengio, Y., 2014: Empirical Evaluation of Gated Recurrent Neural Networks on Sequence Modeling, *NIPS 2014 Workshop on Deep Learning*.
- Connor, J. T., Martin, R. D., Atlas, L. E., 1994: Recurrent Neural Networks and Robust Time Series Prediction, *IEEE Transactions on Neural Networks*, 5, 240–254. doi.org/10.1109/72.279188.
- Dauphin, Y. N., Fan, A., Auli, M., Grangier, D., 2017: Language Modeling with Gated Convolutional Networks, *International Conference on Machine Learning*.
- Defferrard, M., Bresson, X., Vandergheynst, P., 2016: Convolutional Neural Networks on Graphs with Fast Localized Spectral Filtering, *Advances in Neural Information Processing Systems* 29.
- Guo, S., Lin, Y., Feng, N., Song, C., Wan, H., 2019: Attention Based Spatial-Temporal Graph Convolutional Networks for Traffic Flow Forecasting, *Proceedings of the AAAI Conference on Artificial Intelligence*. doi.org/10.1609/aaai.v33i01.3301922.
- Han, Y., Peng, T. X., Wang, C., Zhang, Z. H., Chen, G., 2021: A Hybrid GLM Model for Predicting Citywide Spatio-Temporal Metro Passenger Flow, *ISPRS International Journal of Geo-Information*. doi.org/10.3390/ijgi10040222.
- Han, Y., Wang, S., Ren, Y., Wang, C., Gao, P., Chen, G., 2019: Predicting Station-Level Short-Term Passenger Flow in a Citywide Metro Network Using Spatiotemporal Graph Convolutional Neural Networks, *ISPRS International Journal of Geo-Information*, 8. doi.org/10.3390/ijgi8060243.
- Hochreiter, S., Schmidhuber, J., 1997: Long Short-Term Memory, *Neural Computation*, 9, 1735-1780. doi.org/10.1162/neco.1997.9.8.1735.
- Jia, F. F., Li, H. Y., Jiang, X., Xu, X. Y., 2019: Deep learning-based hybrid model for short-term subway passenger flow prediction using automatic fare collection data, *IET Intelligent Transport Systems*, 13, 1708-1716. doi.org/10.1049/iet-its.2019.0112.
- Kipf, T. N., Welling, M., 2016: Semi-Supervised Classification with Graph Convolutional Networks, *5th International Conference on Learning Representations*, arXiv:1609.02907.
- Kranti, K., Parida, M., Katiyar, V. K., 2013: Short Term Traffic Flow Prediction for a Non Urban Highway Using Artificial Neural Network, *Procedia - Social and Behavioral Sciences*, 104, 755-764. doi.org/10.1016/j.sbspro.2013.11.170.

- Krizhevsky, A., Sutskever, I., Hinton, G. E., 2012: ImageNet Classification with Deep Convolutional Neural Networks, *Proceedings of the 25th International Conference on Neural Information Processing Systems*, 1097–1105.
- Li, L. C., Wang, Y. G., Zhong, G., Zhang, J., Ran, B., 2018: Short-to-medium Term Passenger Flow Forecasting for Metro Stations using a Hybrid Model, *KSCE Journal of Civil Engineering*, 22, 1937-1945. doi.org/10.1007/s12205-017-1016-9.
- Liang, S., Ma, M., He, S., Zhang, H., 2019: Short-Term Passenger Flow Prediction in Urban Public Transport: Kalman Filtering Combined K-Nearest Neighbor Approach, *IEEE Access*, 7, 120937-120949. doi.org/10.1109/ACCESS.2019.2937114.
- Liu, S. S., Yao, E. J., 2017: Holiday Passenger Flow Forecasting Based on the Modified Least-Square Support Vector Machine for the Metro System, *Journal of Transportation Engineering, Part A: Systems*, 143. doi.org/10.1061/JTEPBS.0000010.
- Liu, S. Y., Liu, S., Tian, Y., Sun, Q. L., Tang, Y. Y., 2021: Research on Forecast of Rail Traffic Flow Based on ARIMA Model, *Journal of Physics: Conference Series*, 1792, 012065. doi.org/10.1088/1742-6596/1792/1/012065.
- Ni, M., He, Q., Gao, J., 2017: Forecasting the Subway Passenger Flow Under Event Occurrences With Social Media, *IEEE Transactions on Intelligent Transportation Systems*, 18, 1623-1632. doi.org/10.1109/TITS.2016.2611644.
- Pinlong, C., Yunpeng, W., Guangquan, L., Peng, C., Chuan, D., Jianping, S., 2016: A spatiotemporal correlative k-nearest neighbor model for short-term traffic multistep forecasting, *Transportation Research Part C: Emerging Technologies*, 62, 21-34. doi.org/10.1016/j.trc.2015.11.002.
- Smola, A. J., Schlkopf, B., 2004: A Tutorial on Support Vector Regression, *Statistics and Computing*, 199-222.
- Sun, X., Li, J., Lv, Z., Dong, C., 2020: Traffic Flow Prediction Model Based on Spatio-Temporal Dilated Graph Convolution, *KSH Transactions on Internet and Information Systems*, 14, 3598-3614. doi.org/10.3837/tiis.2020.09.002.
- Tang, L. Y., Zhao, Y., Cabrera, J., Ma, J., Tsui, K. L., 2019: Forecasting Short-Term Passenger Flow: An Empirical Study on Shenzhen Metro, *IEEE Transactions on Intelligent Transportation Systems*, 20, 3613-3622. doi.org/10.1109/TITS.2018.2879497.
- Wang, J. C., Zhang, Y., Wei, Y., Hu, Y. L., Piao, X. L., Yin, B. C., 2021: Metro Passenger Flow Prediction via Dynamic Hypergraph Convolution Networks, *IEEE Transactions on Intelligent Transportation Systems*, 22, 7891-7903. doi.org/10.1109/TITS.2021.3072743.
- Wang, X. M., Zhang, N., Zhang, Y. L., Shi, Z. B., 2018: Forecasting of Short-Term Metro Ridership with Support Vector Machine Online Model, *Journal of Advanced Transportation*. doi.org/10.1155/2018/3189238.
- Wang, Y., Jing, C., 2022: Spatiotemporal Graph Convolutional Network for Multi-Scale Traffic Forecasting, *ISPRS International Journal of Geo-Information*, 11. doi.org/10.3390/ijgi11020102.
- Xu, C., Zhang, A., Xu, C., Chen, Y., 2021: Traffic speed prediction: spatiotemporal convolution network based on long-term, short-term and spatial features, *Applied Intelligence*, 52, 2224-2242. doi.org/10.1007/s10489-021-02461-9.
- Yang, X., Xue, Q. C., Yang, X. X., Yin, H. D., Qu, Y. C., Li, X., Wu, J. J., 2021: A novel prediction model for the inbound passenger flow of urban rail transit, *Information Sciences*, 566, 347-363. doi.org/10.1016/j.ins.2021.02.036.
- Yu, B., Yin, H., and Zhu, Z., 2018: Spatio-Temporal Graph Convolutional Networks: A Deep Learning Framework for Traffic Forecasting, *Proceedings of the 27th International Joint Conference on Artificial Intelligence*, 3634–3640.
- Yu, F. and Koltun, V., 2016: Multi-Scale Context Aggregation by Dilated Convolutions, *International Conference on Learning Representations*.
- Zhang, J., Zheng, Y., and Qi, D., 2017: Deep Spatio-Temporal Residual Networks for Citywide Crowd Flows Prediction, *Aaai'17*, 1655–1661.
- Zhang, J., Zheng, Y., Qi, D., Li, R., and Yi, X., 2016: DNN-Based Prediction Model for Spatio-Temporal Data, *Sigspacial '16*. doi.org/10.1145/2996913.2997016.
- Zhao, L., Song, Y., Zhang, C., Liu, Y., Wang, P., Lin, T., Deng, M., and Li, H., 2020: T-GCN: A Temporal Graph Convolutional Network for Traffic Prediction, *IEEE Transactions on Intelligent Transportation Systems*, 21, 3848-3858. doi.org/10.1109/TITS.2019.2935152.