Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

# Deep learning in the COVID-19 epidemic: A deep model for urban traffic revitalization index☆

Zhiqiang Lv [a], Jianbo Li [a,*], Chuanhao Dong [b], Haoran Li [b], Zhihao Xu [b]

[a] *College of Computer Science & Technology, Qingdao University, Qingdao 266071, China*
[b] *Institute of Ubiquitous Networks and Urban Computing, Qingdao 266070, China*

## ARTICLE INFO

## ABSTRACT

The research of traffic revitalization index can provide support for the formulation and adjustment of policies related to urban management, epidemic prevention and resumption of work and production. This paper proposes a deep model for the prediction of urban Traffic Revitalization Index (DeepTRI). The DeepTRI builds model for the data of COVID-19 epidemic and traffic revitalization index for major cities in China. The location information of 29 cities forms the topological structure of graph. The Spatial Convolution Layer proposed in this paper captures the spatial correlation features of the graph structure. The special Graph Data Fusion module distributes and fuses the two kinds of data according to different proportions to increase the trend of spatial correlation of the data. In order to reduce the complexity of the computational process, the Temporal Convolution Layer replaces the gated recursive mechanism of the traditional recurrent neural network with a multi-level residual structure. It uses the dilated convolution whose dilation factor changes according to convex function to control the dynamic change of the receptive field and uses causal convolution to fully mine the historical information of the data to optimize the ability of long-term prediction. The comparative experiments among DeepTRI and three baselines (traditional recurrent neural network, ordinary spatial–temporal model and graph spatial–temporal model) show the advantages of DeepTRI in the evaluation index and resolving two under-fitting problems (under-fitting of edge values and under-fitting of local peaks).

## 1. Introduction

### 1.1. Urban traffic revitalization index

Traffic situation is a key feature that reflects the health and order of urban operations, production and life. The road situation and travel data [1] can well reflect the recovery of urban production and consumption activities. The data of urban traffic revitalization index (referred to in this paper simply as TRI) is calculated by Didi platform [2]. The researchers of Didi platform carry on the fitting, cross-validating and weighting to the data of urban traffic trajectory and road congestion. The TRI is an important indicator that reflects the urban traffic travel situation. Based on smart transportation technology and data analysis capabilities, the TRI allows users to visually see the recovery of each urban traffic. It provides more information for the orderly promotion of the recovery of

**Table 1**
The results of evaluation indexes. The relationship between GDP and TRI in some cities. The unit of GDP ordinate is 100 million (yuan) and the TRI is the probability, whose value range is [0, 1].

| Urban name | GDP (first quarter) | GDP (second quarter) | Growth rate (GDP) | Mean TRI (first quarter) | Mean TRI (second quarter) | Growth rate (TRI) |
|---|---|---|---|---|---|---|
| Chengdu | 3845.57 | 4453.06 | 15.80% | 0.3791 | 0.7903 | 108% |
| Dongguan | 1923.7 | 2437.58 | 26.71% | 0.3759 | 0.8263 | 119.81% |
| Foshan | 2153.98 | 2472.52 | 14.79% | 0.3890 | 0.7988 | 105.37% |
| Fuzhou | 1838.94 | 2521.67 | 37.13% | 0.4357 | 0.8524 | 95.65% |
| Guangzhou | 5228.8 | 5739.49 | 9.77% | 0.4538 | 0.8143 | 79.45% |
| Guiyang | 838.76 | 994.9 | 18.62% | 0.4168 | 0.8769 | 110.37% |
| Haikou | 350.03 | 371.6 | 6.16% | 0.2672 | 0.8467 | 216.85% |
| Hangzhou | 3379 | 4009 | 18.64% | 0.3491 | 0.8081 | 131.45% |
| Huizhou | 879.14 | 1003.66 | 14.16% | 0.3721 | 0.7962 | 113.97% |
| Jinhua | 922.3 | 1178.99 | 27.83% | 0.3332 | 0.7852 | 135.66% |
| Kunming | 1513.27 | 1762.47 | 16.47% | 0.4239 | 0.7921 | 86.85% |
| Nanchang | 1204.62 | 1424.26 | 18.23% | 0.2679 | 0.6779 | 152.98% |
| Nanjing | 3247 | 3365.35 | 36.45% | 0.3323 | 0.7877 | 137.06% |
| Nanning | 1010.18 | 1244.32 | 23.18% | 0.3699 | 0.8158 | 120.54% |
| Ningbo | 2463.8 | 3023.4 | 22.71% | 0.3603 | 0.7701 | 113.73 |
| Qingdao | 2504.87 | 3009.86 | 20.16% | 0.4027 | 0.7789 | 93.40% |
| Quanzhou | 1997.64 | 2670.49 | 33.68% | 0.3523 | 0.7520 | 113.44% |
| Xiamen | **1318.02** | **989.34** | **−24.94%** | **0.8539** | **0.4936** | **−0.4219%** |
| Shanghai | 7856.62 | 9500.18 | 20.92% | 0.4210 | 0.8542 | 102.93% |
| Shaoxing | 1225 | 1436.15 | 17.24% | 0.3371 | 0.7646 | 126.82% |
| Shenzhen | 5785.6 | 6848.7 | 18.37% | 0.4455 | 0.8913 | 100.07% |
| Suzhou | 3743.93 | 5306.27 | 41.73% | 0.3659 | 0.7773 | 112.43% |
| Tianjin | 2874 | 3435.28 | 19.53% | 0.2826 | 0.8016 | 183.65% |
| Wenzhou | 1352.7 | 1723.2 | 27.39% | 0.3239 | 0.7964 | 145.87% |
| Wuxi | 2465.24 | 3051.08 | 23.76% | 0.3513 | 0.7584 | 115.90% |
| Changsha | 2704.9 | 2916.31 | 78.16% | 0.4151 | 0.8316 | 100.34% |
| Zhengzhou | 2469.8 | 2989.8 | 21.05% | 0.2354 | 0.7696 | 226.96% |
| Zhongshan | 597.22 | 792.1 | 32.63% | 0.3544 | 0.8052 | 127.19% |
| Chongqing | 4987.66 | 6222.17 | 24.75% | 0.2938 | 0.8499 | 182.29% |

production and life. It scientifically and objectively reflects the activity of urban traffic. In general, the amount of traffic activity raises with the increase of the TRI. The TRI is an indicator specially established for the impact of the COVID-19 on urban transportation. In the concept of TRI, the TRI of each city is close to 1 during the non-epidemic period at the end of 2019. With the outbreak of the COVID-19 at the end of 2019, the transportation of people has been greatly affected, which directly led to a sharp drop in the TRI of each city. In the face of the COVID-19, the government and health departments have formulated and implemented various epidemic prevention measures. Production and life in various cities have gradually recovered. The vitality of urban traffic will gradually return to normal levels.

The transportation industry is the basic and leading industry of the national economy and it is an important part of the national economy. The development of the transportation industry promotes rapid economic growth. The economic growth in turn drives the development of the transportation industry. Therefore, the TRI has dynamic synchronization with economic development. The Table 1 shows the relationship between the GDP of urban and TRI. The GDP of China is calculated on a quarterly basis, so we calculate the GDP and mean TRI of multiple Chinese cities in the first two quarters of 2020. From the statistical results, we can learn that almost all cities have experienced substantial growth in GDP in the second quarter. At the same time, TRI has also increased accordingly. Because of the second wave of the outbreak in Xiamen, its GDP in the second quarter declined. Its TRI for the second quarter decreased by 0.4219%. Based on the above results, GDP and TRI have a dynamic synchronization regardless of whether GDP is increasing or decreasing.

### 1.2. The economic impact of the COVID-19

The COVID-19 has had a profound and significant impact on the transportation industry [3–6] and it has had a serious impact on the world economy [7–9]. With the spread of the epidemic, the decline in industrial demand and the rising risk of virus infection have led to a sharp drop in Chinese transportation services. At the end of 2019, the TRI of Chinese city is close to 1 under normal conditions. However, due to the spread of the epidemic, the TRI of all cities in China have declined to varying degrees in early 2020. Fig. 1 shows the relationship between the industrial added value increased year on year in early 2020 and the number of suspected cases in China. We can know that the number of suspected cases reached its peak in February. Due to the impact of epidemic, the value of industrial growth declined severely in March, and even turned into a negative growth. After March, with the help of strong prevention and control measures, the epidemic has gradually subsided in China and the number of suspected cases has rapidly decreased. Urban production and life have gradually recovered and traffic vitality will gradually return to normal level [10].
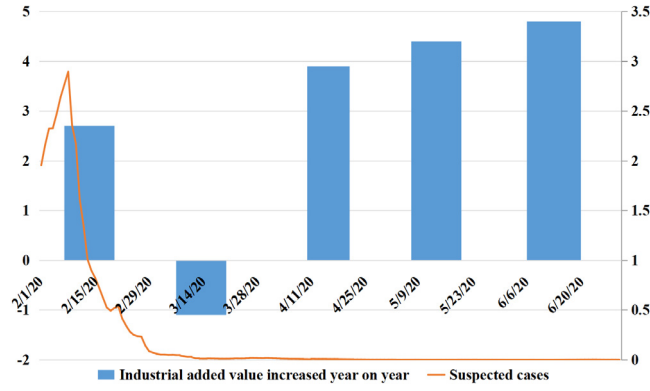
**Fig. 1.** The relationship between the industrial added value increased year on year and the number of suspected cases. The abscissa represents the date. The industrial added value increased year on year is proportional, which corresponds to the main axis on the left. The unit of the number of suspected cases is 10,000 (person), which corresponds to the secondary axis on the right.
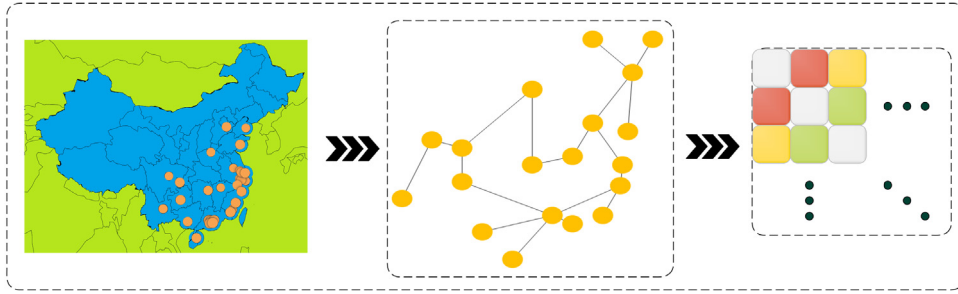


**Fig. 2.** The process of transforming urban spatial correlation into adjacency matrix. The points on the left map represent the locations of some Cities in China. The middle graph represents the topological relations of city locations. The right graph represents the adjacency matrix corresponding to the topological relations of urban locations.

### 1.3. Graph neural network

With the development of deep learning, significant technological breakthroughs have been made in the fields of image, speech and natural language processing [11–13]. Deep learning is good at processing structured data, such as voice, images and text. However, there are a lot of unstructured data in the real world, such as social network topology and knowledge map. Unstructured data does not have the same spatial locality as images, whose data range is arbitrary and topology is complex. Compared with the most basic structure (Multi-Layer Perception, MLP) of neural network, graph neural network [14] adds an adjacency matrix to participate in the process of matrix calculation, as shown in formula (1). $\sigma$ represents nonlinear transformation. $A$ represents the adjacency matrix, $X$ represents the feature matrix and $W$ represents the weight matrix.

$$H = \sigma(AXW) \tag{1}$$

The graph spatial–temporal network [15] captures the spatial correlation and temporal dependence of data. The spatial–temporal graph has a global graph structure and the value of each node changes with the passage of time. For example, many works have established graph spatial–temporal networks for PeMS data [16]. The basic realization process of these networks uses data stations of PeMS as nodes of the graph. The input of the nodes are time series, such as traffic flow, vehicle speed and road occupancy rate. The edges of the graph spatial–temporal networks represent the distances of latitude and longitude among data stations. The task of the graph spatial–temporal networks is to predict the future value of node or the type of node. The spatial correlation among cities conforms to the topological relationship. This topological relationship can be converted into an adjacency matrix [17], which can be used in the computational process of graph neural network, as shown in Fig. 2.

### 1.4. Contribution

Based on the above research theories, this paper converts the spatial location relationship of Chinese cities into graph structure and combines the data of COVID-19 epidemic and TRI to accomplish the task of urban traffic recovery prediction. The main contributions of this paper are as follows:

- This paper proposes a new graph spatial–temporal network (Deep Model for Traffic Revitalization Index, DeepTRI). This is the first spatial–temporal network used to predict the TRI and the first work to combine the TRI with the data of COVID-19 epidemic.
- According to the characteristics of epidemic data, a special Graph Data Fusion (GDF) module is proposed in this paper. GDF assigns different weights to four types of epidemic data. The features of spatial correlation of the graph structure are fully mined through the Spatial Convolution Layer composed of Graph Convolutional Network (GCN) and GDF.
- This paper designed a Temporal Convolution Layer which is used to calculate time sequence features of temporal dependencies. In order to overcome the weakness that traditional recursive neural network does not support parallel computing and has a slow training speed., the Temporal Convolution Layer replaces the gating mechanism of the recurrent neural network with a multilayer residual structure. We use the dilation factor that changes according to a convex function with an exponent of 2 to control the receptive field of the deep network and use the mechanism of causal convolution to maintain the influence of historical data on the temporal features.
- We use DeepTRI to train the data of TRI and four types of epidemic data and set up three types of baselines (traditional recurrent neural network, graph convolution and graph spatial–temporal network) to verify the performance of DeepTRI. The experimental results show that DeepTRI has obvious advantages in evaluating index and resolving two under-fitting problems (underfitting of edge values and underfitting of local peaks).

## 2. Related work

The basic principle of this work is to use the location information among cities and the temporal series changes of TRI to build a model of the graph structure. The basic features among cities are mapped to the edge value information of the graph. The temporal series characteristics of TRI of cities are mapped to the sequence data of the graph nodes. Therefore, the knowledge involved in this work is the establishment of graph structure with adjacency matrix and the data prediction of unstructured graph.

### 2.1. Basics

The basic design process of the graph spatial–temporal network is that the time changes of traffic data (vehicle flow, vehicle speed, road occupancy rate and etc.) are designed as the input values of the graph nodes and the position relationship among the nodes is designed as the weight of the graph edges. Considering that traditional methods cannot meet the requirements of long-term traffic prediction and the traffic data has complex nonlinear problems, the Spatial-Temporal Graph Convolutional Networks (STGCN) [18] is used to predict the time series of traffic data based on graph neural networks. In order to improve the training speed of the model, STGCN uses a pure convolution structure to build spatial–temporal convolution blocks. The spatial–temporal convolution blocks are composed of two Temporal Gated-Conv blocks and one spatial Graph-Conv block. STGCN shows the advantages of accuracy and training speed in comparison with statistical methods (ARIMA), traditional machine learning methods and spatial–temporal models. The Attention Based Spatial-Temporal Graph Convolutional Networks (ASTGCN) [19] is proposed for the task of traffic flow prediction in traffic networks. ASTGCN separately builds models for three temporal characteristics of traffic flow (recent, daily-periodic and weekly-periodic data). The computational processes of the three temporal characteristics are independent of each other. Each computational process first uses the temporal attention mechanism to calculate the dynamic dependence among different times and uses the spatial attention mechanism to calculate the dynamic correlation between different positions. Finally, the spatial–temporal convolution module is used to calculate the spatial features of the graph and the dependency of adjacent nodes in the traffic network. ASTGCN is compared with traditional methods, recurrent neural networks and other graph spatial–temporal networks. Even the accuracy of ASTGCN without the attention mechanism is higher than other models. After adding the attention mechanism, ASTGCN decreased by an average of 6.2% in the RMSE indicator and 4.5% in the MAE indicator. The accuracy of ASTGCN without the attention mechanism is higher than other models. After adding the attention mechanism, the RMSE of ASTGCN decreased by 6.2% and the RMSE of ASTGCN decreased by 4.5%. The Temporal Graph Convolutional Network (T-GCN) [20] combines Graph Convolution Network (GCN) and Gated Recursive Unit (GRU) to finish the task of vehicle speed prediction. GCN calculates spatial correlation by building the topological relationship (connectivity) of the transportation network. GRU calculates the temporal dependence by calculating the time series changes of vehicle speed.

### 2.2. Adjacency matrix

The adjacency matrices of STGCN and ASTGCN represent the relationship of distance among nodes. However, the adjacency matrix of T-GCN represents the connectivity among nodes. Therefore, the adjacency matrix of T-GCN is composed of 0 and 1. In the comparison experiment, although T-GCN performs well in the long-term prediction, the prediction results of edge value and local peak value appear the problem of underfitting. Different from the above research, the Diffusion Convolutional Recurrent Neural Network (DCRNN) [21] uses digraph to calculate spatial–temporal features of traffic data. The DCRNN uses two-way random walk mechanism to calculate spatial correlation and uses encoder–decoder mechanism to calculate the temporal dependence. The diffusion of DCRNN mainly reflects in that the entire network generates more predictions by maximizing the time range of the target value. Compared with the baselines of experiment, the accuracy of DCRNN is improved by 12%–15%. GCNN-DDGF [22] designs four adjacency matrices based on the data of shared bicycle system, including spatial distance matrix, demand matrix, average trip duration matrix and demand correlation matrix. It calculates the interrelationships among different types of stations to predict

the short-term demand in a large-scale shared bicycle network. The Graph Convolutional Neural Network with Data-driven Graph Filter (GCNN-DDGF) [23] focuses on the similarity relationship of dynamic flow among nodes. It proposes three modes for building adjacency matrix, including distance graph and interaction graph and related graph. The distance graph uses the reciprocal of the node distance to represent the weight among nodes. The interaction graph uses the number of driving records between two nodes to represent the weight among nodes. The related graph uses the Pearson Correlation Coefficient to calculate the correlation between the inflow or outflow of two nodes in a fixed time interval as the weight among nodes. GCNN-DDGF does not calculate the above three adjacency matrices separately, but it fuses the three adjacency matrices into one adjacency matrix and performs the process of graph convolution.

### 2.3. Special graph

In the field of taxi demand research, in order to capture the relative relationship of space movement of passengers among different areas, the Grid-Embedding based Multi-task Learning (GEML) [24] divides urban areas into grids. Graph nodes represent geographic areas (grid form) and the edges among nodes represent passenger demand. GEML uses a multi-task learning module to calculate the dependency features of time series data (the inflow and outflow of the grid). The Spatial-Temporal Synchronous Graph Convolutional Networks (STSGCN) [25] proposed a special graph structure (local spatial–temporal graph). The local spatial–temporal graph focuses on three kinds of influences of a node on its neighbor nodes: the correlation influence based on spatial relationship, the time dependence influence based on the same node and the influence of spatial–temporal correlation of neighbor node in the transfer process of a previous time step to a subsequent time step. In the local spatial–temporal graph, each node has not only the value of its neighbor nodes at the same time step, but also the value of two time steps before and after itself. The value of its neighbor nodes at two time steps before and after are regarded as its second-order neighbors. The local spatial–temporal convolution uses graph convolution to allow each node to consider its own relationship with its multi-order neighbors. The mutual influence between each node and its first-order neighbors includes spatial dependence and temporal correlation. The influence among second-order neighbors includes the correlation between time and space. The work [26] proposes two adaptive modules which are suitable for graph convolution, including: node adaptive parameter learning (NAPL) and data adaptive graph generation (DAGG). NAPL does not share the parameters among all nodes, but it maintains a unique parameter of space for each node to learn the pattern of a specific node. DAGG overcomes the disadvantage of traditional graph convolutional network that network requires a pre-defined adjacency matrix and it can automatically calculate the implicit interdependence of nodes from the data.

## 3. Model design

### 3.1. Overview

The main structure of DeepTRI is shown in Fig. 3. It establishes a graph structured adjacency matrix ($W$) based on the spatial position relationship among cities. The urban TRI data ($\{V_t\}$) is processed in the form of adaptive graph structure and then participates in the calculation process of the first Temporal Convolution Layer. The first Temporal Convolution Layer performs scale compression and feature transformation for TRI data instead of simple linearization and splicing operations. The $W$ is combined with the output of the first Temporal Convolution Layer ($\{TC_t^1\}$) to calculate the spatial correlation of TRI data. The GDF fuses and calculates the multi-scale features of epidemic data of the various cities ($\{U\}$) and whole country ($\{C\}$). The second Temporal Convolution Layer calculates the temporal dependence of the fused data ($\{SC_t\}$). From Fig. 3, we can know that Temporal Convolution Layer is composed of multi-level residual blocks. Each residual block contains causal convolution and dilated convolution. Causal convolution enables the TRI feature of the current node of the hidden layer to be associated with all historical information, which strengthens the time dependence of the sequence. Dilated convolution enables the convolution process of the hidden layer to obtain a larger and dynamic receptive field without changing the number of hidden layers. The DeepTRI is completely composed of the convolutional network, so the BatchNorm2D is used to prevent the gradient disappearance and gradient explosion of the convolutional network. Finally, the Fully Connected Layer maps the data features to the sample space to get the final output ($\{v_t'\}$).

### 3.2. Definition

The prediction problem of TRI can be defined as model predicts the data ($v_t$) of the next time period ($t$) from historical data ($\{v_{t-s}, \ldots, v_{t-2}, v_{t-1}\}$, $s$ represents the number of historical data), as the formula (2) shows. $F$ represents the method of building model.

$$v_t' = F\left(\left[v_{t-s}, \ldots, v_{t-2}, v_{t-1}\right]\right) \tag{2}$$

The topological relationship among cities is described as a graph structure $G = (W, V)$. The $W$ represents the adjacency matrix among city nodes. The $V$ represents the TRI data of each city node of per unit time step.

The input data of model has four dimensions, namely [*batch size, node count, time step, channel*]. The *batch size* represents the number of samples which are selected for one epoch. The *node count* represents the number of city nodes. The *time step* represents the number of days. The *channel* represents the number of data characteristics.
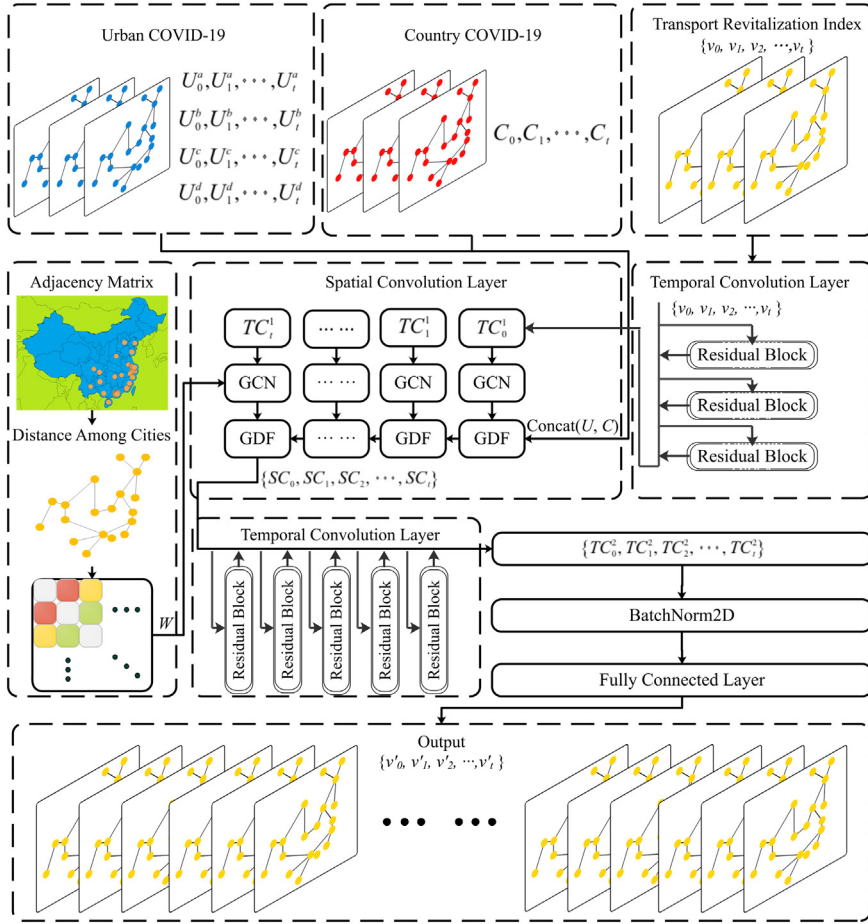
**Fig. 3.** The structure of DeepTRI.

The first Temporal Convolution Layer realizes the linear change of the data and the underlying feature calculation, whose shape is namely [*batch size, node count, time step − k + 1, Output1*]. *k* is the size of the convolution kernel of the first Temporal Convolution Layer. *Output1* is the output dimension of the first Temporal Convolution Layer.

The Spatial Convolution Layer contains two processes (GCN and GDF). The GCN itself only changes in data features and its final shape does not change. The GDF involves the number of people cured by the country's epidemic, so its shape should be increased by 1, namely [*batch size, node count, time step − k + 1, Output1 + 1*].

The second Temporal Convolution Layer calculates the time dependence of data features to form high-level features. Its shape is [*batch size, node count, time step − 2(k + 1), Output2*]. *Output2* is the output dimension of the second Temporal Convolution Layer.

The output value of the model is produced by Fully Connected Layer, namely [*batch size, node count, time step*]. The value of *time step* indicates the range of TRI that the model needs to predict. For example, the model predicts the TRI value of each city in the next three days when the value of *time step* is set to 3.

### 3.3. Spatial convolution layer

The main function of the Spatial Convolution Layer is to calculate the spatial correlation between the TRI in each city and the epidemic data. The graph composed of the distribution of cities is established depending on the distance among cities. The GDF fuses the output of GCN with epidemic data by the method of multiple scales. More importantly, The GDF assigns different weights to different types of epidemic data.

**Graph convolutional network**

The basic process of GCN is that each node of graph is affected by adjacent and farther nodes at any time and changes its own status until the final balance is reached. The closer the adjacent node, the greater the effect on the original node. The Laplacian matrix can make a positive comparison between the transfer intensity of data and the difference of status. In order to add the effect of the original node on itself into computational process, an improved version of Laplace matrix is used, as shown in formula (3).

$$L = D^{-\frac{1}{2}} A D^{-\frac{1}{2}} \tag{3}$$

$A$ represents the adjacency matrix ($W$) after $W$ adds self-joins. $D$ represents the degree distribution of nodes of $A$. It adds its own degree matrix and solves the problem of self-transfer. The normalization operation of the adjacency matrix is realized by the two sides of the adjacency matrix are multiplied by the value which is the degree of node being taken square root and inverted. The original spectral convolution realizes the filter of product of each node and Fourier transform. However, because the eigenvectors are higher-order variables and the decomposition of Laplace matrix is inefficient in the large graph structure, we use the $K$-order Chebyshev polynomials to approximate the computational process of Laplacian matrix, as shown in formula (4).

$$g(\theta) \cdot y \approx \sum_{i=1}^{K} \theta_i \cdot T_i\left(\tilde{L}\right) \cdot y \tag{4}$$

The $y$ is the output of the first Temporal Convolution Layer. The computational process of $T_i(\tilde{L})$ as shown in formula (5), which represents the recursive definition of Chebyshev polynomials. This method is called $K$-localized convolution algorithm the $K$-localized convolution algorithm which ensures that the current node only considers the influence of nodes whose value in the range of $K$. The above process greatly reduces the time complexity of graph convolutional network.

$$
\begin{aligned}
T_i\left(\tilde{L}\right) &= 2\tilde{L} \cdot T_{i-1}\left(\tilde{L}\right) - T_{i-2}\left(\tilde{L}\right) \\
T_0\left(\tilde{L}\right) &= 1, T_1\left(\tilde{L}\right) = \tilde{L} \\
\tilde{L} &= \frac{2}{\lambda_{max}} L - I_n
\end{aligned}
\tag{5}
$$

**Graph data fusion**

The TRI data passes through the first layer of Temporal Convolution Layer to form the underlying features of TRI data. After GCN processing, it has the preliminary spatial correlation features of urban location. The above process is the data processing process of matrix (a) in Fig. 4. The GDF is indispensable to DeepTRI, because the spatial features of matrix (a) that represents the relationship of TRI among cities are not significant. The experimental results of DeepTRI without GDF lead to poor performance, we have verified this theory in our experiments. The GDF fuses the epidemic data of each city with the TRI data of each city and the national epidemic data by the method of multiple scales to increase the spatial features of each city's traffic revitalization. We achieve multi-scale fusion by performing operations such as concat and weighted summation of different dimensions on the matrix (a) representing the underlying features of TRI and the original epidemic data. The multi-scale fusion process fuses different levels of semantic features to improve the accuracy of prediction.

For the computational process of (b) in Fig. 4, GDF considers firstly the proportion of different types of epidemic data in the modeling process. There are four types of epidemic data used by DeepTRI, including the number of confirmed cases, the number of suspected cases, the number of cured cases and the number of dead cases. Although four types of epidemic data are all about the COVID-19 epidemic, they have different distribution characteristics, as shown in Fig. 5. Fig. 5 shows the changes in Chengdu's TRI and four types of epidemic data within three months. The change of cured cases has obvious dynamic correlation with the TRI and the other three data have obvious differences with the TRI in the process of change. Therefore, GDF proportionally allocates the dynamic correlation of the TRI according to the four types of epidemic data.

Such as the computational process in Fig. 4, the four types of epidemic data are mapped into 4 channels in the matrix (b) and the channel representing cured cases is linearized into 7 dimensions. The proportion of the four types of data is 1:1:1:4. After the proportional allocation process, the feature matrix has the same dimensions as the matrix (a) which is processed by Con2d with a convolution whose kernel is 3×3. Because the matrices (a) and (b) have the same matrix dimensions at this time, we apply the weighted summation method to these two matrices. The matrix (c) represents the cured cases of the whole country. It has preliminary data features after passing the Conv2d with a kernel of 3×1. Finally, the matrix (c) is concatenated to the matrix which is the result of add for the features of (a) and (b) to complete the process of multi-scale data fusion of GDF.

## 3.4. Temporal convolution layer

The traffic vitality represents the long-term traffic change process of the city. For a city, even if the short-term changes in traffic vitality are uncertain, we can obtain its temporal dependent features from long-term changes. Therefore, we need to design a neural network layer to calculate the epidemic and TRI data, so that the model can obtain multi-scale context information to improve the accuracy of prediction. According to the four-dimensional data structure designed in this paper, we design a novel Temporal Convolution Layer to calculate the features of temporal dependence, as shown in Fig. 6. The role of Temporal Convolution Layer is mainly reflected in the following three aspects:

- It retains all historical information and uses causal convolution to calculate long-term historical information.
- It uses dilated convolution to expand the receptive field of the convolution process. The dilation factor changes according to the convex function so that the model does not have too large receptive fields and do not cause local information loss during deep calculation.
- It is completely composed of a convolutional network and uses a multi-layer residual structure [27] to replace the gating structure of the traditional recurrent neural network. It overcomes the phenomenon that the traditional recurrent neural network does not support parallel computing and the training speed is slow [28].
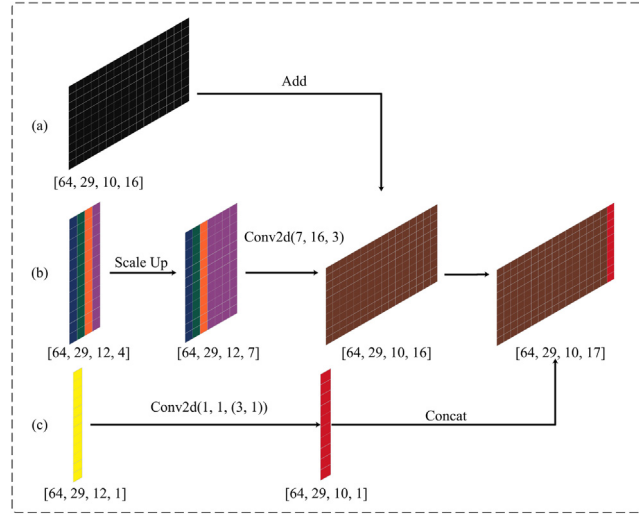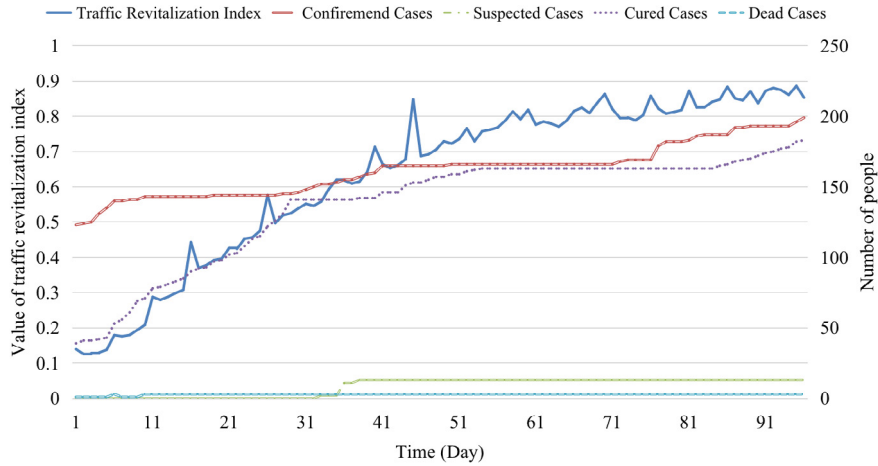
**Fig. 4.** GDF structure.



**Fig. 5.** The time curve of Chengdu's TRI and four types of epidemic data. The abscissa represents time (day). The ordinate on the left represents the value of TRI. The ordinate on the right represents the number of people.

**Causal convolution**

The theory of causal convolution is applied to trace more distant historical information [29], the formula of causal convolution is shown in (6). $\{x_1, x_2, \ldots, x_t\}$ is the input sequence, $\{y_1, y_2, \ldots, y_t\}$ is the output sequence of hidden layer and $\{f_1, f_2, \ldots, f_k\}$ is the filter sequence. Causal convolution only pays attention to historical information and ignores future information. The result of $y_t$ only is derived from the data before $x_t$. The larger the $K$, the more historical information that can be traced back. If the original input sequence of the current layer is $[0, i]$, the original input sequence of the next layer will become $[0, i+1]$.

$$y_t = \sum_{i=1}^{K} f_i \cdot x_{t-K+i} \qquad (6)$$

**Dilated convolution**

In the deep network, in order to increase the receptive field and reduce the amount of calculation, the model always needs to down-sampling. Although the receptive field can be increased, the spatial resolution is reduced. In order to not lose the resolution and still enlarge the receptive field, we use the dilated convolution [30]. Dilated convolution has a parameter which can control the dilation rate ($d$). The convolution kernel is filled with $d$-1 number of 0. Therefore, the receptive field will be different when different dilation rates are set. The above process can capture multi-scale information. The formula of dilated convolution is shown in (7).

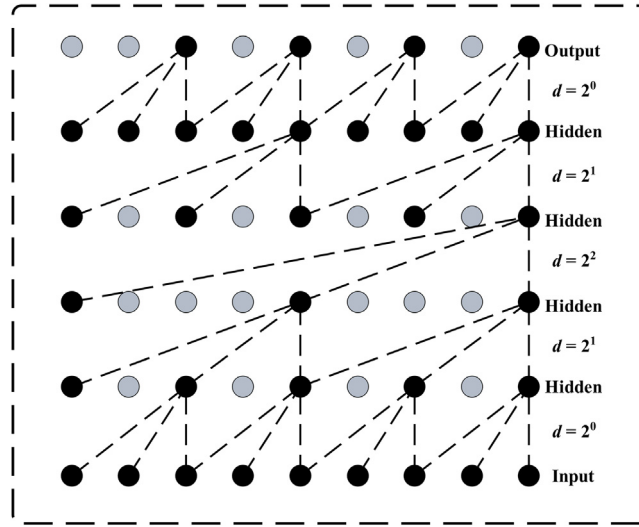$$y_t = \sum_{i=0}^{K-1} f_i \cdot x_{t-i \cdot d} \qquad (7)$$

8

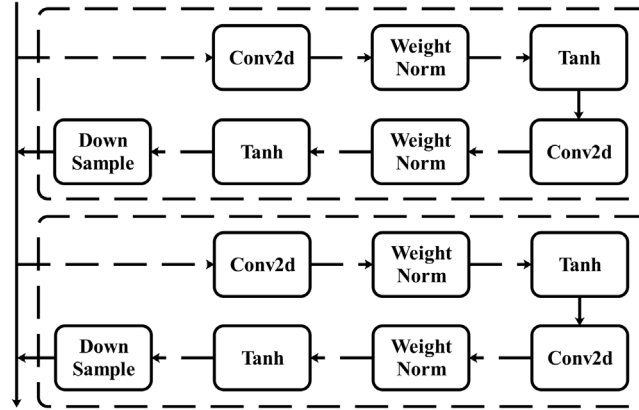**Fig. 6.** The structure of Temporal Convolution Layer.



**Fig. 7.** The structure of multi-level residual.

$d$ represents the dilation factor (dilation rate), which changes according to the convex function. The value range of $d$ is within the exponential range of 2. Increasing $d$ or $K$ can increase the range of the receptive fields. However, the receptive fields make the information obtained by long-distance convolution have no correlation with the deepening of the network layer, which means that the local information of the model is lost resulting in the loss of local information. Therefore, we change the dilation factor according to the convex function, as shown in the change process of $d$ in Fig. 6. This design mode ensures that the range of the receptive field is restricted to a certain extent in the deep network and the loss of local information is reduced. The information transmission of dense neuron makes the model more accurate in the computational process of deep features.

**Multi-level residual structure**

In order to reduce the complexity of the training process, the residual structure is used to replace the gating structure in the traditional recurrent neural network, as shown in Fig. 7. The residual structure mainly includes a two-layer convolutional network and a nonlinear mapping process. The Weight Norm [31] is to use the method (reparameterization weight normalization) to achieve data normalization. The Weight Norm is often used to accelerate model convergence. It can suppress the range of gradient when the gradient rotates by normalizing the weight and achieve self-stabilize of the gradient.

We use the Tanh as the activation function after the convolutional network for the following reasons:

The value range of the TRI data used in this experiment is [0,1]. The Z-Score standardization process is performed after the data is loaded. The formula of Z-Score is as follows.

$$Z = \frac{X - \overline{X}}{s} \tag{8}$$

$X$ represents the original data, $\overline{X}$ represents the mean and $s$ represents the standard deviation. From the computational process of Z-Score, it saves the difference between the original data and the average. So, the $Z$ of the average is equal to 0, the $Z$ of the data

which is greater than the average is positive and the $Z$ of the data which is smaller than the average is negative. After the TRI data is standardized, its minimum value is −2.399 and its maximum value is 1.415. The Tanh function has a mean value of 0, which is more conducive to improving training efficiency. Since the output range of Tanh is [−1,1], the gradient value of the parameter has an opposite sign during the training process so that the zigzag phenomenon is not easy to occur when updating weight and it is easy to reach the optimal value [32].

### 3.5. Post-processing

Based on the above theory of Weight Norm, we add Batch Norm [33] in the process of data normalization to deal with the problem of unstable performance caused by excessive data volume after activating the Temporal convolution Layer. The Batch Norm is as the formula (9) shown.

$$y = \frac{x - \mu}{\sqrt{v + eps}} \times gamma + beta \tag{9}$$

$\mu$ and $v$ are the mean and variance of the input data, respectively. The $eps$ is a stability factor added to increase the stability of the computational process and its value is equal to *1e-5* The *gamma* and *beta* are coefficient matrices that can be learned. The process of data normalization can make the direction of the gradient descent to the direction of the vertical contour in space. The above process can improve the convergence speed of the model and it can average the weight of each level of data in the feature calculation to improve the accuracy of the model. Finally, the fully connected layer maps the distributed features calculated by the network to the sample target space.

## 4. Experiment

### 4.1. Data preparation

The time range of the TRI data [2] used in this experiment is from February 10, 2020 to June 30, 2020. The time type of TRI data is continuous working days. It contains 29 major cities in China. The TRI data is obtained by fitting, cross-validating and weighting the massive traffic data provided by the Didi Travel Platform. It can scientifically and objectively reflect the urban traffic activity. The change trend of the index can reflect the trend of traffic activity, thereby indirectly reflecting the trend of urban recovery. The epidemic data includes four types of data (the number of confirmed cases, the number of suspected cases, the number of cured cases and the number of dead cases) for 29 cities and the number of cured cases for the entire China. The time range of epidemic data is the same as that of the TRI data. The epidemic data comes from the National Health Commission of the People's Republic of China [34].

### 4.2. Index comparison

**Base lines**

We set three types of baselines to verify the performance of the model, including traditional recurrent neural networks, spatial–temporal neural networks and graph spatial–temporal neural networks. All models are trained and evaluated on the same data set. The experimental result is the mean value of multiple training and evaluation results. The model structure of various baselines in the experiment is as follows:

- LSTM [35]: It controls the transmission state of data through a gating mechanism. Compared with traditional RNN which only mechanically superimposes one kind of memory, the LSTM retains the long-term memory and forget the unimportant information.
- GRU [36]: Its input and output structure are similar to traditional RNN and its processing logic is similar to LSTM. Compared with LSTM, GRU has one less "gating" inside and has fewer parameters than LSTM, but it can achieve similar functions and accuracy to LSTM. Considering the computing power and time cost of hardware, GRU is the choice of more researchers engaged in deep learning.
- STDN [37]: It uses local CNN and LSTM to calculate spatial–temporal information. A gated local CNN uses dynamic similarity among regions to model spatial dependence. A mechanism of periodic shifting attention is used to learn long-term periodic dependence. Finally, it uses the attention mechanism to model long-term periodic information and time translation.
- GCN [38]: It uses a filter to perform a weighted summation of pixels in a certain spatial area to obtain a new feature representation. The weighting coefficient is the parameter of the convolution kernel.
- STSGCN [39]: It consists of multiple graph convolution operations and an aggregation operation. The output of each graph convolution operation will be added to the input of the aggregation layer in a similar way to the residual neural network. Aggregation is achieved using maximum pooling.
- GCNN-DDGF [21]: It uses a traditional convolutional network and a graph convolutional network based on data-driven filtering to capture spatial correlation. It uses a recursive block of LSTM structure to capture the temporal dependence of data features.
- DCRNN [21]: It uses a directed graph to calculate the spatial–temporal features of the data in the adjacency matrix design process. We use its two-way random walk mechanism to calculate the spatial correlation among cities and use the Encoder-Decoder mechanism to calculate the temporal dependence of the data.

**Table 2**
The results of evaluation indexes.

| Model | RMSE | | | MAE | | | MAPE | | |
|---|---|---|---|---|---|---|---|---|---|
| | 1d | 2d | 3d | 1d | 2d | 3d | 1d | 2d | 3d |
| GRU | 6.15 | 6.34 | 6.52 | 3.86 | 3.90 | 3.94 | 6.88% | 7.54% | 8.32% |
| LSTM | 6.06 | 6.27 | 6.62 | 3.67 | 3.75 | 3.89 | 6.55% | 7.29% | 8.28% |
| STDN | 6.11 | 6.30 | 6.50 | 3.80 | 3.85 | 3.94 | 6.72% | 7.40% | 8.25% |
| GCN | 6.09 | 6.46 | 6.49 | 3.77 | 4.00 | 3.96 | 6.68% | 7.61%. | 8.29% |
| GCNN-DDGF | 5.86 | 5.97 | 6.02 | 3.87 | 3.97 | 3.73 | 6.13% | 7.01% | 7.45% |
| DCRNN | 5.46 | 5.85 | 5.97 | 3.78 | 3.86 | 3.65 | 6.11% | 6.72% | 6.89% |
| GEML | 5.23 | 5.54 | 5.69 | 3.55 | 3.59 | 3.72 | 6.11% | 6.43% | 6.57% |
| STSGCN | 5.02 | 5.42 | 5.62 | 3.41 | 3.54 | 3.69 | 5.94% | 6.13% | 6.47% |
| DAGG | 4.32 | 4.45 | 4.98 | 3.23 | 3.48 | 3.62 | 5.13% | 5.36% | 5.97% |
| T-GCN | 3.98 | 4.10 | 4.07 | 3.06 | 3.16 | 3.13 | 4.49% | 4.72% | 4.78% |
| ASTGCN | 3.58 | 3.64 | 3.73 | 2.72 | 2.77 | 2.84 | 4.39% | 4.57% | 4.86% |
| STGCN | 2.87 | 2.88 | 2.89 | 2.19 | 2.21 | 2.22 | 3.47% | 3.40% | 3.31% |
| DeepTRI | **1.36** | **1.34** | **1.53** | **0.98** | **0.95** | **1.14** | **1.44%** | **1.38%** | **1.62%** |

- GEML [24]: It divides the urban area into grids, but this work is aimed at the TRI relationship among multiple cities, so the grid division of a single city does not apply to this work. We only use the model result of GEML that is the multi-task learning module.
- DAGG [26]: It does not need to predefine the adjacency matrix that represents the spatial relationship of the city. The interdependence among city nodes can be automatically established during the training process.
- T-GCN [19]: This model combines graph convolutional network (GCN) and gated recursive unit (GRU). GCN is used to learn complex topological structures to capture spatial correlation and GRU is used to learn dynamic changes of traffic data to capture temporal dependence.
- STGCN [17]: It builds a spatial–temporal convolution block (ST-Conv block) composed of a continuous three-layer convolution network and a layer of GCN. The data is processed by two layers of ST-Conv blocks.
- ASTGCN [18]: It uses a spatial–temporal block (ST-BLOCK) composed of a spatial attention layer and a temporal attention layer. The data is processed by two layers of ST-BLOCKs. Finally, it performs a fully-connected process with three temporal characteristics.

**Index introduction**

This experiment uses three indexes to evaluate the performance of the model, including root mean square error (RMSE), mean absolute error (MAE) and mean absolute percentage error (MAPE), as shown in the formula (10).

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^{n} \left(\widehat{y}_i - y_i\right)^2}$$

$$MAE = \frac{1}{n} \sum_{i=1}^{n} \left|\widehat{y}_i - y_i\right| \tag{10}$$

$$MAPE = \frac{1}{n} \sum_{i=1}^{n} \left|\frac{\widehat{y}_i - y_i}{y_i}\right|$$

MAE and MAPE can measure the accuracy of a model's prediction results, but they are both based on the computational process of absolute error. Although the absolute error can obtain an evaluation value, we cannot know the pros and cons of the model performance represented by this evaluation value. Only through the comparison between models can we know the best model. Therefore, RMSE was also used in the evaluation process of this experiment. RMSE is used to measure the deviation between the predicted value and the true value. It is very sensitive to very large or very small errors in a set of data. So, the RMSE can well reflect the precision of the prediction results [40].

The evaluation results of all models for the three indexes are shown in Table 2. GRU and LSTM are traditional recurrent neural networks, which have the worst performance. Because they cannot handle complex spatial relationship. The STDN and GEML add a convolutional network to the traditional recurrent neural network model to calculate the spatial relationship and the accuracy of the model is improved. However, it is not a model for graph structure and it has no significant effect in this experiment. The GCN is suitable for calculating the spatial relationship of the graph structure, but it lacks the ability to rely on the temporal features of the calculation, so its performance is not best optimal. The city information and TRI data features of the dataset used in this work are very obvious and rich. Compared with the automatic graph structure model of DAGG, the adjacency matrix directly established in this work is more effective. The basic process of graph spatial–temporal neural networks (GCNN-DDGF, DCRNN, STSGCN, T-GCN, ASTGCN and STGCN) is that they use GCN for spatial relationship calculation and use recurrent neural network for the calculation of temporal dependence. Therefore, they cannot obtain the deep multi-scale context features of the data during the recursive modeling process.
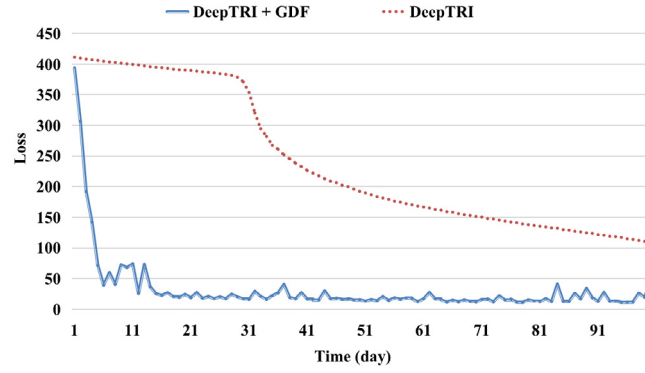
**Fig. 8.** The effect of GDF. The abscissa represents time (day). The ordinate represents the value of loss.
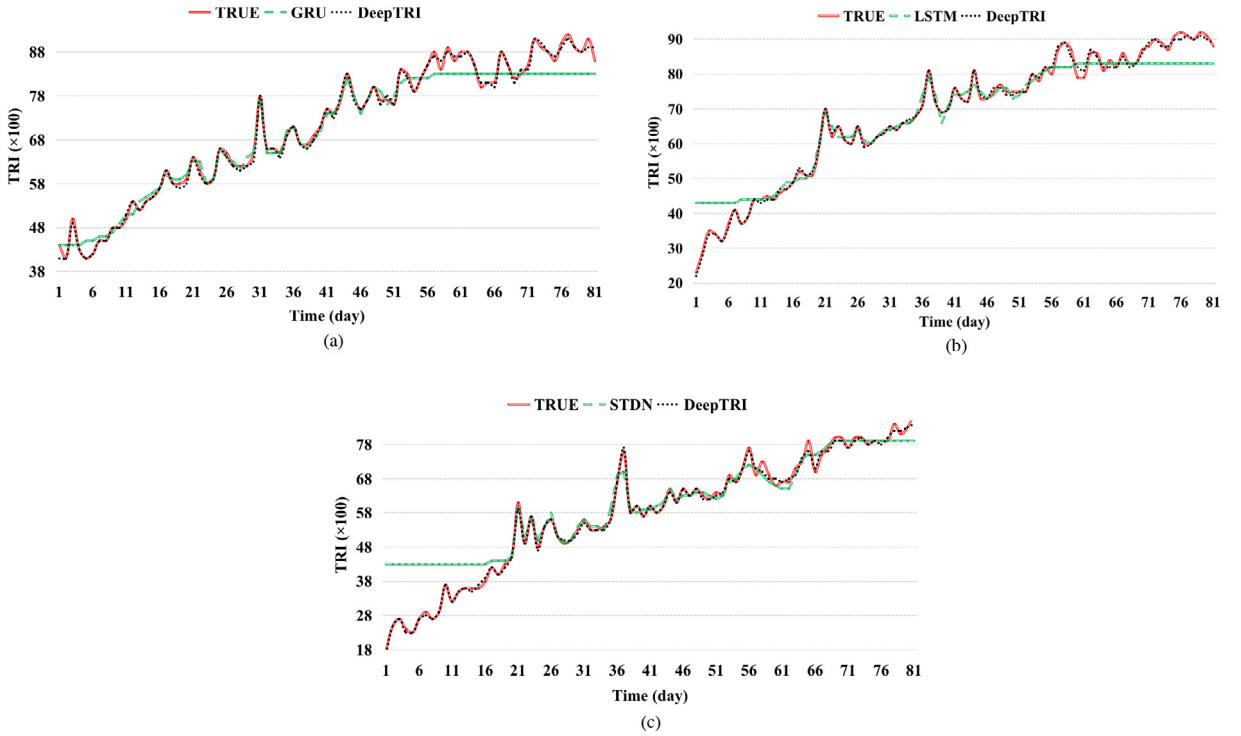


**Fig. 9.** The change of TRI of a single node over time. The abscissa represents time (day). The ordinate represents the value of TRI.

**Table 3**
The error reduction ratio.

| Time step (d) | RMSE | MAE | MAPE |
|---|---|---|---|
| 1 | 52.61% | 55.25% | 58.50% |
| 2 | **53.47%** | **57.01%** | **59.41%** |
| 3 | 47.06% | 48.65% | 51.06% |

Compared with the above graph neural network, DeepTRI has a significant effect with the smallest errors in the three dimensions of time. This is due to the fact that DeepTRI considers the impact of epidemic data on the change of TRI in both space and time. Compared with the indexes of the three dimensions time of STGCN, the error reduction ratio is shown in Table 3. When the time step of DeepTRI is 2d, the reduction ratio is the largest. This is because the causal convolution mechanism of DeepTRI's Temporal Convolution Layer can retain all historical information, which is more suitable for long-term information calculation.
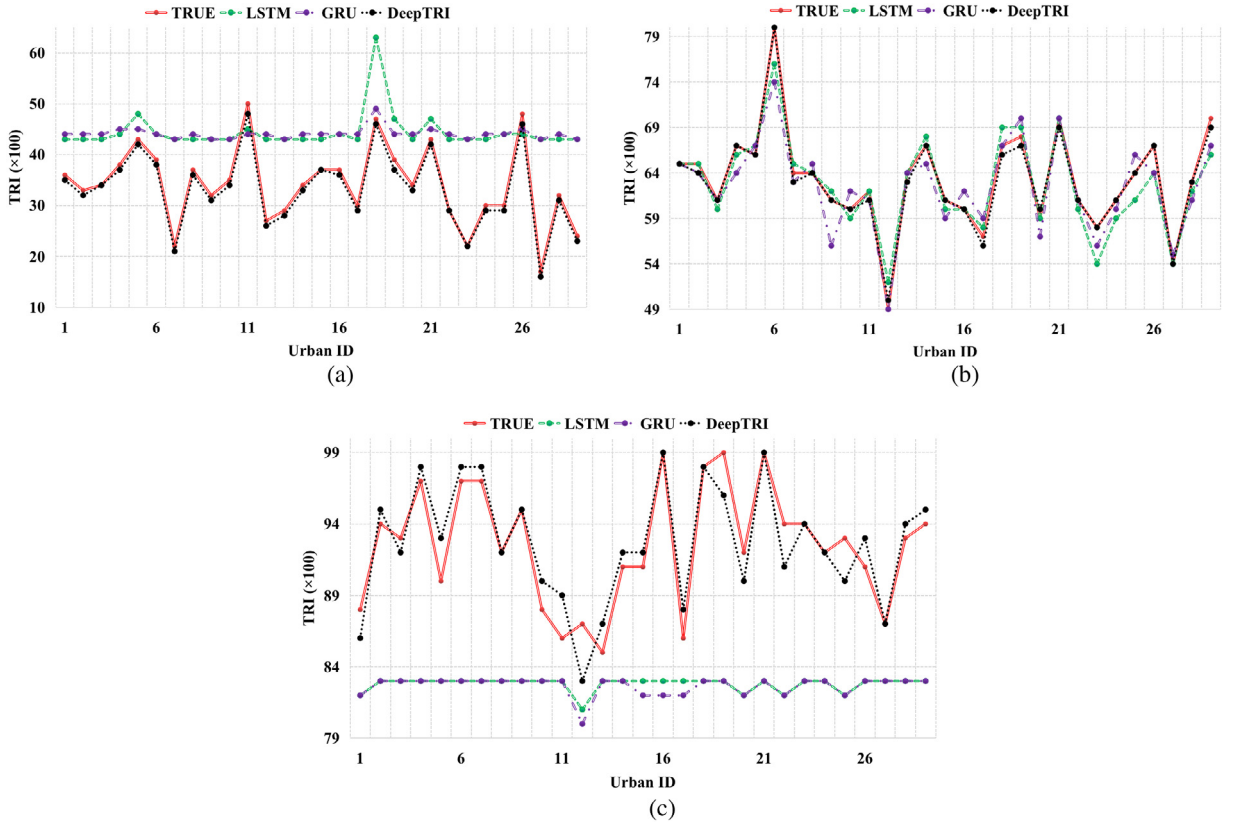
**Fig. 10.** The TRI value of all nodes in a single time period. The abscissa represents time (day). The ordinate represents the value of TRI.

### 4.3. The effect of GDF

The GDF plays an important role in the training process of model. The effect of GDF is shown in Fig. 8. Fig. 8 shows the changes in the loss value of the two models for 100 epochs and the experimental parameters are exactly the same. The DeepTRI without GDF converges very slowly in the early stage of the training process. Even if the loss has decreased significantly in the mid-term, the final effect is not ideal. With the help of GDF, the DeepTRI can converge quickly and the loss of training can be maintained in a small numerical range without overfitting. This is because the epidemic data has obvious dynamic correlation to the TRI and GDF amplifies their dynamic correlation. Finally, The GDF strengthens the model's weighting features of the modeling target.

### 4.4. Spatial-temporal results

This chapter mainly focuses on two under-fitting problems in the comparison of actual results of the model, including: under-fitting of edge values and under-fitting of local peaks [41].

**Under-fitting of edge values**

The problem of under-fitting of edge values means that the model has better prediction results for the median value of TRI, but it has poor prediction results for smaller or larger TRI. This kind of problem is shown in Fig. 9(a), (b) and (c). For example, in the (a), the predictive value of the TRI is basically consistent with the true value in the interval [48, 78], while there is an under-fitting problem in the predicted value in the interval [38, 48] and [78, 88]. The predictive result approximates a straight line. The under-fitting problem of edge value is reflected in all city matrices. Fig. 10(a) and (b) illustrate the spatial breadth of edge value. The reason for this problem is that the intermediate values overlap and the features are more obvious in the data set, while the edge data is sparse and the spatial correlation between the edge values is insufficient.

The above problem can be effectively avoided by the graph neural network, as shown in Fig. 11(a). The (a) is the modeling result of TRI with the GCN, which effectively avoids the problem of under-fitting of edge values. The GCN can capture the spatial relationship of the graph structure, which makes the predictive values have fitting ability in all intervals of the data. Fig. 11(b) and (c) show the results of complex graph spatial–temporal networks. From their comparison results, we can learn that the spatial–temporal network based on GCN can further suppress the under-fitting problem of edge values. In particular, the T-GCN as a graph spatial–temporal network exhibits a certain degree of under-fitting problem of edge values. This is because the T-GCN uses a gating mechanism to capture the temporal dependence of data. It relies too much on historical data and the weight of the edge value in
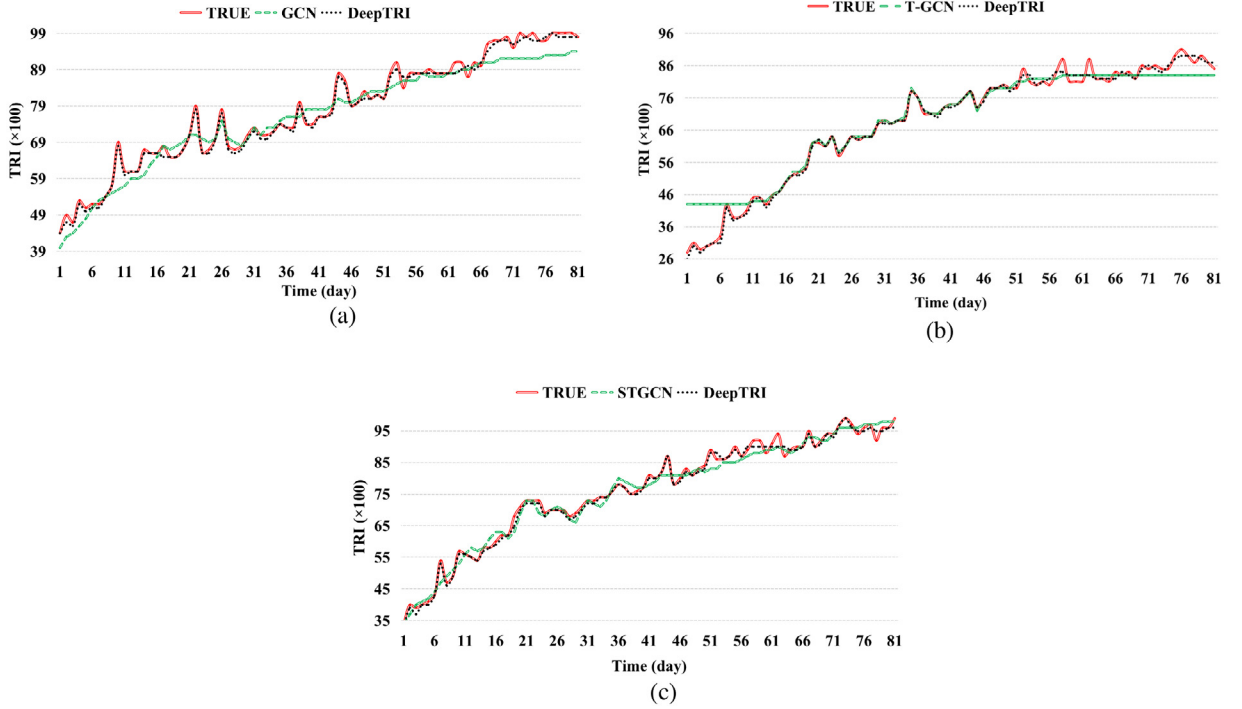
**Fig. 11.** The change of TRI of a single node over time. The abscissa represents time (day). The ordinate represents the value of TRI.
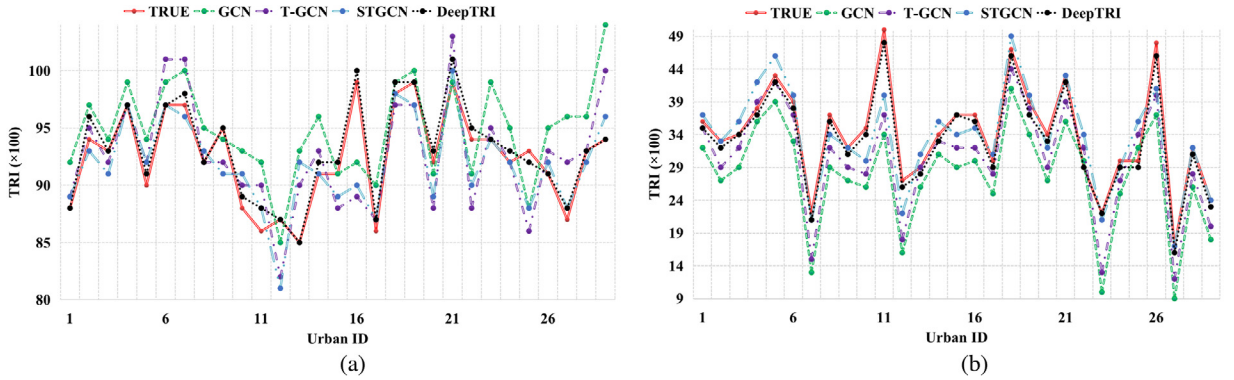


**Fig. 12.** The TRI value of all nodes in a single time period. The abscissa represents time (day). The ordinate represents the value of TRI.

the historical data is too small, which eventually leads to the above problem. Fig. 12(a) and (b) respectively show the degree of fit of all city nodes to larger and smaller edge values. We can know that the spatial breadth of edge values is suppressed with the help of graph convolutional networks and each city node shows different data changes.

**Under-fitting of local peaks**

The under-fitting problem of local peak means that the under-fitting problem of the predictive value occurs when the true data changes greatly. Although the overall evaluation index is good, the prediction results of local peaks are poor, as shown in Fig. 13. Fig. 13 is a part of the curve in Fig. 11(a). Although the addition of GCN solves the problem of underfitting of edge values, it brings about the problem of underfitting of local peaks. This problem occurs because GCN is only suitable for the calculation of spatial correlation of graph structure, but it lacks consideration of the recursive relationship of time series. In Fig. 9, the traditional recurrent neural network has significant advantages in the temporal-dependent calculation of the median value of TRI. The graph spatial–temporal network represented by STGCN shows the advantage of predicting edge values and local peaks. They use GCN to solve the problem of underfitting of edge values and use recursive networks to solve the problem of underfitting of local peaks.

**Overall improvement**

The predictive effect of DeepTRI is shown in Figs. 9, 10, 11 and 12. Figs. 9 and 10 show the superiority of DeepTRI in capturing time series features in a single node with continuous time period. Figs. 11 and 12 show the superiority of DeepTRI for capturing
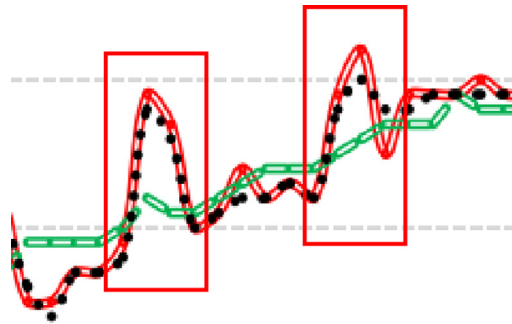
**Fig. 13.** Examples of local peaks.

the spatial features of all nodes at the same time. The accuracy of DeepTRI is better than traditional recurrent neural networks because the Spatial Convolution Layer can establish the spatial relative relationship among nodes and strengthen the dynamic synchronization features of urban nodes. The accuracy of DeepTRI is better than GCN because the Temporal Convolution Layer relies on the dilated convolution according to the convex function changing to establish a wide range of receptive fields and relies on the multi-level residual structure to overcome the long-term temporal dependence of the gating mechanism on historical data. The accuracy of DeepTRI is better than other graph spatial–temporal network because the special GDF structure fuses the epidemic data with the TRI data in the method of multiple scales according to different weights, it makes the trend of data feature more obvious.

## 5. Conclusion

The COVID-19 epidemic is a challenge that all mankind must cope with. Firstly, the emergence of COVID-19 has affected human activities. With the reduction of human travel activities, the vitality of the transportation industry gradually declines. Because the transportation industry is closely related to the development of the economy of city, the economy has been hit hard by the emergence of COVID-19. Studying the relationship between TRI and the COVID-19 epidemic is of great significance for government departments to grasp the traffic and economic development status of the city during the epidemic. We use the existing data of TRI and COVID-19 to realize the prediction research on the future urban traffic vitality. The special GDF module realizes multi-scale fusion of the data of TRI and COVID-19 and it plays a huge advantage in the modeling process. We rely on the location information of the city to establish a graph structure and fully explore the spatial relationship among city nodes in the Spatial Convolution Layer. The special Temporal Convolution Layer fully exploits the features of temporal dependence of TRI and COVID-19. DeepTRI shows an obvious advantage when facing with two under-fitting problems. However, the acquisition of TRI data has certain limitations. TRI data is based on the normal conditions of urban traffic as the boundary value (it is generally equal to 1). Its meaning refers to the process of a major accident in the city that affects the operation of the urban traffic and eventually returns to a normal state. Therefore, if a urban traffic has been operating normally, its TRI data will be difficult to obtain. The time range of the TRI data used in this work is from February 10, 2020 to June 30, 2020 in China. The emergence of the new crown epidemic was at the end of 2019. However, the initial stage of the epidemic did not receive the attention of people and relevant anti-epidemic policies were not designated to be formulated and implemented. Therefore, the traffic status of most cities in China was not affected at the beginning of the epidemic. With the development of the epidemic, China implemented strict epidemic prevention policies from February to June 2020. At the same time, the traffic in various cities in China has gradually returned to normal from the worst. The above process is the main source of TRI data in this work. In future work, we will expand the scale of data to analyze and predict the relationship between TRI and COVID-19 for the entire country and select more cities as nodes to join the graph structure. Furthermore, we can calculate the extent which represents that the country is affected by the epidemic when there are enough data of cities in a country. In our opinion, the significance of this research is that it provides a reference tool for the analysis and prediction of future public safety disasters.

## CRediT authorship contribution statement

**Zhiqiang Lv:** Methodology, Validation, Writing - original draft, Writing - review & editing. **Jianbo Li:** Project administration, Supervision. **Chuanhao Dong:** Formal analysis, Software. **Haoran Li:** Data Curation. **Zhihao Xu:** Investigation.

## Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

# References

[1] E. Jenelius, H.N. Koutsopoulos, Travel time estimation for urban road networks using low frequency probe vehicle data, Transp. Res. B 53 (2013) 64–81, http://dx.doi.org/10.1016/j.trb.2013.03.008.

[2] https://gaia.didichuxing.com.

[3] Tirachini Alejandro, Cats Oded, COVID-19 and public transportation: Current assessment, prospects, and research needs, J. Public Transp. 22 (1) (2020) 1–21, http://dx.doi.org/10.5038/2375-0901.22.1.1.

[4] Loske Dominic, The impact of COVID-19 on transport volume and freight capacity dynamics: An empirical analysis in german food retail logistics, Transp. Res. Interdiscip. Perspect. 6 (1) (2020) 1–6, http://dx.doi.org/10.1016/j.trip.2020.100165.

[5] Pengfei Wang, Kaiyu Chen, Shengqiang Zhu, Severe air pollution events not avoided by reduced anthropogenic activities during COVID-19 outbreak, Resour. Conserv. Recy. 158 (3) (2020) 1–9, http://dx.doi.org/10.1016/j.resconrec.2020.104814.

[6] Yen Muh Yen, Schwartz Jonathan, Chen Shey-Ying, Interrupting COVID-19 transmission by implementing enhanced traffic control bundling: Implications for global prevention and control efforts, J. Microbiol. Immunol. Infect. 53 (3) (2020) 377–380, http://dx.doi.org/10.1016/j.jmii.2020.03.011.

[7] Weidong Lai, Zhaoxi Deng, The influence and countermeasures of the "COVID-19" on the economic development of coastal economic zone of guangdong province, Int. J. Econ. Behav. Organ. 8 (2) (2020) 44–48, http://dx.doi.org/10.11648/j.ijebo.20200802.14.

[8] W Bartik Alexander, Bertrand Marianne, Cullen Zoe, The impact of COVID-19 on small business outcomes and expectations, Proc. Natl. Acad. Sci. 117 (30) (2020) 17656–17666, http://dx.doi.org/10.1073/pnas.2006991117.

[9] Andy Sumner, Chris Hoy, Eduardo Ortiz-Juarez, Estimates of the impact of COVID-19 on global poverty, UNU-WIDER 43 (9) (2020) 1–14, http://dx.doi.org/10.35188/UNU-WIDER/2020/800-9.

[10] Moritz U.G. Kraemer, Chia-Hung Yang, Bernardo Gutierrez, The effect of human mobility and control measures on the COVID-19 epidemic in China, Science 368 (6490) (2020) 493–497, http://dx.doi.org/10.1126/science.abb4218.

[11] Chunwei Tian, Lunke Fei, Wenxian Zheng, Deep learning on image denoising: An overview, Neural Netw. 131 (1) (2020) 251–275, http://dx.doi.org/10.1016/j.neunet.2020.07.025.

[12] DeLiang Wang, Jitong Chen, Supervised speech separation based on deep learning: An overview, IEEE/ACM Trans. Audio 26 (10) (2018) 1702–1726, http://dx.doi.org/10.1109/TASLP.2018.2842159.

[13] Wei Emma Zhang, Quan Z. Sheng, Ahoud Alhazmi, Adversarial attacks on deep-learning models in natural language processing: A survey, ACM Trans. Intell. Syst. Technol. (TIST) 11 (3) (2020) 1–41, http://dx.doi.org/10.1145/3374217.

[14] Franco Scarselli, Marco Gori, Ah Chung Tsoi, The graph neural network model, IEEE Trans. Neural Netw. 20 (1) (2008) 61–80, http://dx.doi.org/10.1109/TNN.2008.2005605.

[15] Xiufang Sun, Jianbo Li, Zhiqiang Lv, Traffic flow prediction model based on spatio-temporal dilated graph convolution, KSII Trans. Internet and Inf. Syst. 14 (9) (2020) 3598–3614, http://dx.doi.org/10.3837/tiis.2020.09.002.

[16] Michael A. Hickner, Ghassemi Hossein, Seung Kim Yu, Alternative polymer systems for proton exchange membranes (PEMs), Chem. Rev. 104 (10) (2004) 4587–4612, http://dx.doi.org/10.1021/cr020711a.

[17] Frank Harary, The determinant of the adjacency matrix of a graph, Siam Rev. 4 (3) (1962) 202–210, http://dx.doi.org/10.1137/1004057.

[18] Bing Yu, Haoteng Yin, Zhanxing Zhu, Spatio-temporal graph convolutional networks: A deep learning framework for traffic forecasting, in: Proceedings of the 27th International Joint Conference on Artificial Intelligence, IJCAI, Stockholm, Sweden, 2018, pp. 3634–3640, http://dx.doi.org/10.24963/ijcai.2018/505.

[19] Shengnan Guo, Youfang Lin, Ning Feng, Attention based spatial–temporal graph convolutional networks for traffic flow forecasting, in: Proceedings of the AAAI Conference on Artificial Intelligence, AAAI, New York, USA, 2019, pp. 922–929, http://dx.doi.org/10.1609/aaai.v33i01.3301922.

[20] Ling Zhao, Yujiao Song, Chao Zhang, T-gcn: A temporal graph convolutional network for traffic prediction, IEEE Trans. Intell. Transp. Syst. 21 (9) (2019) 3848–3858, http://dx.doi.org/10.1109/TITS.2019.2935152.

[21] Yaguang Li, Rose Yu, Shahabi Cyrus, Diffusion convolutional recurrent neural network: Data-driven traffic forecasting, in: Proceedings of International Conference on Learning Representations, ICLR, Toulon, France, 2017, pp. 1–14, arXiv preprint arXiv:1707.01926.

[22] Lei Lin, Zhengbing He, Srinivas Peeta, Predicting station-level hourly demand in a large-scale bike-sharing network: A graph convolutional neural network approach, Transp. Res. C 97 (1) (2018) 258–276, http://dx.doi.org/10.1016/j.trc.2018.10.011.

[23] Di Chai, Leye Wang, Qiang Yang, Bike flow prediction with multi-graph convolutional networks, in: Proceedings of the 26th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems, ACM, Washington, USA, 2018, pp. 397–400, http://dx.doi.org/10.1145/3274895.3274896.

[24] Yuandong Wang, Hongzhi Yin, Hongxu Chen, Origin–destination matrix prediction via graph convolution: a new perspective of passenger demand modeling, in: Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, ACM, Alaska, USA, 2019, pp. 1227–1235, http://dx.doi.org/10.1145/3292500.3330877.

[25] Chao Song, Youfang Lin, Shengnan Guo, Spatial-temporal synchronous graph convolutional networks: A new framework for spatial-temporal network data forecasting, in: Proceedings of the AAAI Conference on Artificial Intelligence, AAAI, New York, USA, 2020, pp. 914–921, http://dx.doi.org/10.1609/aaai.v34i01.5438.

[26] Lei Bai, Lina Yao, Can Li, Adaptive Graph Convolutional Recurrent Network for Traffic Forecasting. arXiv preprint arXiv:2007.02842.

[27] Shaojie Bai, J. Zico Kolter, Vladlen Koltun, An empirical evaluation of generic convolutional and recurrent networks for sequence modeling, Arxiv 4 (1) (2018) 1–14, arXiv preprint arXiv:1803.01271.

[28] Zhipeng Cai, Xu Zheng, Jiguo Yu, A differential-private framework for urban traffic flows estimation via taxi companies, IEEE Trans. Ind. Inf. 15 (12) (2019) 6492–6499, http://dx.doi.org/10.1109/TII.2019.2911697.

[29] T.J. Brazil, Causal-convolution-a new method for the transient analysis of linear systems at microwave frequencies, IEEE Trans. Microw. Theory Tech. 43 (2) (1995) 315–323, http://dx.doi.org/10.1109/22.348090.

[30] Fisher Yu, Vladlen Koltun, Multi-scale context aggregation by dilated convolutions, ArXiv 23 (1) (2015) 1–13, arXiv preprint arXiv:1511.07122.

[31] Salimans Tim, P. Kingma Diederik, Weight normalization: A simple reparameterization to accelerate training of deep neural networks, in: Proceedings of Advances in Neural Information Processing Systems, Niagara, US, 2016, pp. 901–909, arXiv preprint arXiv:1602.07868.

[32] B.L. Kalman, S.C. Kwasny, Why tanh: choosing a sigmoidal function, in: Proceedings of the International Joint Conference on Neural Networks, IEEE, Baltimore, MD, USA, 1992, pp. 578–581, http://dx.doi.org/10.1109/IJCNN.1992.227257.

[33] Ioffe Sergey, Szegedy Christian, Batch normalization: Accelerating deep network training by reducing internal covariate shift, ArXiv 11 (1) (2015) 1–11, arXiv preprint arXiv:1502.03167.

[34] http://www.nhc.gov.cn/.

[35] Hochreiter Sepp, Schmidhuber Jürgen, Long short-term memory, Neural Comput. 9 (8) (1997) 1735–1780, http://dx.doi.org/10.1162/neco.1997.9.8.1735.

[36] Kyunghyun Cho, Bart van Merrienboer, Caglar Gulcehre, Learning phrase representations using RNN encoder–decoder for statistical machine translation, ArXiv 3 (3) (2014) 1–15, arXiv preprint arXiv:1406.1078.

[37] Huaxiu Yao, Xianfeng Tang, Hua Wei, Revisiting spatial–temporal similarity: A deep learning framework for traffic prediction, in: Proceedings of the AAAI Conference on Artificial Intelligence, AAAI, Hawaii, US, 2019, pp. 5668–5675, http://dx.doi.org/10.1609/aaai.v33i01.33015668.

[38] Thomas N. Kipf, Welling Max, Semi-supervised classification with graph convolutional networks, ArXiv 9 (4) (2016) 1–14, arXiv preprint arXiv:1609.02907.

[39] Chao Song, Youfang Lin, Shengnan Guo, Spatial-temporal synchronous graph convolutional networks: A new framework for spatial-temporal network data forecasting, in: Proceedings of the AAAI Conference on Artificial Intelligence, AAAI, New York, USA, 2020, pp. 914–921, http://dx.doi.org/10.1609/aaai.v34i01.5438.

[40] Tianfeng Chai, R.R. Draxler, Root mean square error (RMSE) or mean absolute error (MAE)? –Arguments against avoiding RMSE in the literature, Geosci. Model Dev. 7 (3) (2014) 1247–1250, http://dx.doi.org/10.5194/gmd-7-1247-2014.

[41] Gavrilov Andrei Dmitri, Jordache Alex, Vasdani Maya, Preventing model overfitting and underfitting in convolutional neural networks, Int. J. Softw. Sci. Comput. Intell. 10 (4) (2018) 19–28, http://dx.doi.org/10.4018/IJSSCI.2018100102.

**Zhiqiang Lv** was born in Weifang city, Shandong province, China in 1995. He received the bachelor degree in software engineering from Ludong University, China, in 2019. From 2016 to 2019, he studied in the software parallel group of the State Key Laboratory of Computer Architecture, Institute of Computer Technology Chinese Academy of Sciences. He is studying for the master degree in Qingdao University, China, majoring in computer technology. He has won 13 awards of the science and technology innovation in university, 1 invention patent and many university student scholarships. His main research directions are traffic demand research based on traffic trajectory and travel time, traffic data forecast research based on traffic flow, speed and traffic congestion, high performance parallel computing research based on deep learning and reinforcement learning.

**Jianbo Li** was born in Weifang city, Shandong province, China in 1980. He received the Ph.D. degree in computer science and technology department from the University of Science and Technology of China in 2009. From 2013 to 2014, he was a visiting scholar at Fordham University. He is currently the professor of the college of computer science & technology in Qingdao University and the director of the Institute of ubiquitous network and urban computing of the Qingdao University. He is the chairman of ACM Qingdao Branch, deputy secretary general of Qingdao Computer Society, senior member of China Computer Federation, and member of Internet of Things Professional Committee of China Computer Federation. His research interests include urban computing, mobile social networks and data offloading.

**Chuanhao Dong** was born in Jining city, Shandong province, China in 1993. He graduated from Qingdao University in 2018 with a master's degree in computer science and technology, and now works at the Institute of Ubiquitous Networks and Urban Computing, Qingdao. His main research directions are urban computing, intelligent transportation system and traffic flow prediction.

**Haoran Li** was born in Heze city, Shandong Province in 1992. He got a master degree in computer science and technology at Qingdao University in China. Now, He works at the Institute of Ubiquitous Networks and Urban Computing. His research focuses on urban computing, urban area classification and traffic big data mining.

**Zhihao Xv** was born in Binzhou city, Shandong province, China, in 1992. He received the bachelor degree in digital media technology from Qingdao University, China, in 2014. He got the master degree in Qingdao University, majoring in computer technology, in 2017. Now, He works at the Institute of Ubiquitous Networks and Urban Computing, Qingdao. His main research directions are deep learning, urban computing, Intelligent Transportation, crowd density research and crowd interest prediction.