

Limi-TFP: Citywide Traffic Flow Prediction With Limited Road Status Information

Qiuyang Huang^{ID}, Hongfei Jia^{ID}, Yuanbo Xu^{ID}, Yongjian Yang^{ID}, and Gaoxi Xiao^{ID}, *Senior Member, IEEE*

Abstract—Citywide traffic flow prediction is a basic but challenging task in the field of Intelligent Transportation Systems (ITS). Most existing traffic flow prediction methods require historical traffic data of all the target roads in both training and application stages. However, it is not easy, and usually costly, to collect real-time traffic data across the entire road network. Therefore, many existing methods have to narrow down the scope of their prediction on the parts of road networks. In this paper, we propose a deep-learning model, named Limi-TFP, which has the ability to identify a limited number of monitored roads, and achieves citywide traffic flow prediction by using the historical traffic data of these selected roads. Specifically, an embedding module is proposed to capture the spatial context (road topology structure) and the attributes (e.g., road type, length, lanes, etc.) of each road by embedding all road segments into vectors. A road ranking method is then developed for selecting a limited number of roads to be monitored. A group of multi-head attention mechanisms is exploited for capturing the dynamic correlations with each monitored road. Meanwhile, the fusion with external factors, including the point of interests (POIs) and weather data, is also taken into account to further improve the prediction accuracy. Extensive experiments on the real-world datasets demonstrate that the proposed method achieves the superior performance and is robust with the presence of noise compared to the state-of-the-art baselines, with only 5% roads being monitored.

Index Terms—Deep learning, intelligent transportation systems, multi-head attention, road ranking, traffic flow prediction.

Manuscript received 3 May 2022; revised 19 August 2022; accepted 22 October 2022. Date of publication 27 October 2022; date of current version 14 March 2023. This work was supported in part by the Open Fund of Key Laboratory of Urban Land Resources Monitoring and Simulation, Ministry of Natural Resources under Grant KF-2021-06-006, in part by the Project of National Natural Science Foundation of China under Grants 52072143, 62072209, and 62002123, in part by the Science and Technology Project of Education Department of Jilin Province under Grant JKJH20221010KJ, in part by the Jilin Provincial Development and Reform Commission Project under Grant 2020C017-2, in part by the Ministry of Education, Singapore, under Contract RG19/20, and in part by the Future Resilient Systems Programme (FRS-II) at the Singapore-ETH Centre (SEC), which was supported by NRF Singapore, through its CREATE programme. The review of this article was coordinated by Prof. Luca D’Acierno. (*Corresponding author:* Yuanbo Xu.)

Qiuyang Huang is with the College of Transportation, Jilin University, Changchun 130012, China, and also with the Key Laboratory of Urban Land Resources Monitoring and Simulation, Ministry of Natural Resources, Shenzhen 518034, China (e-mail: huangqy@jlu.edu.cn).

Hongfei Jia is with the College of Transportation, Jilin University, Changchun 130012, China (e-mail: jiahf@jlu.edu.cn).

Yuanbo Xu and Yongjian Yang are with the College of Computer Science and Technology, Jilin University, Changchun 130012, China (e-mail: yuanbox@jlu.edu.cn; yyj@jlu.edu.cn).

Gaoxi Xiao is with the School of Electrical and Electronic Engineering, Nanyang Technological University, Singapore 639798 (e-mail: egxxiao@ntu.edu.sg).

Digital Object Identifier 10.1109/TVT.2022.3217521

I. INTRODUCTION

R EAL-TIME road traffic prediction at a citywide scale is of great importance for the intelligent transportation systems (ITS) [1], [2], [3]. It not only facilitates providing a smarter road route for travelers by the routing systems [4], [5], [6], but also helps identify future congestion in the urban road network [7], [8], [9].

Specific to the problem of traffic prediction, many methods have been proposed. Early approaches typically are based on time series analysis [10], [11], [12]. Such methods cannot capture the complex external features, such as the road’s surrounding points of interest (POIs) and the weather conditions, large errors could thus exist in the prediction results. The long short-term memory (LSTM) networks [13], [14], [15], by effectively utilizing the characteristics of long-term and short-term historical data, manages to perform well in traffic prediction. It however depends on a long receptive field, and cannot capture the spatial correlation of the road network. Recently, many innovative deep-learning based traffic prediction methods have been developed [16], [17], [18], [19], [20], [21], [22], [23], [24], which significantly improve the prediction accuracy by capturing the dynamic spatial-temporal correlations.

With no exception, however, all the above methods require the historical traffic information for all target roads in both the training and the application stages, which means that, to make traffic flow prediction for the entire road network, real-time traffic data must be frequently collected on all roads. Collecting real-time traffic data of the entire road network however is a costly and challenging task: 1) massive computing and storage resources are required for long-period traffic monitoring for the whole road network; 2) for vehicle trajectory data based traffic forecasting systems, a time-consuming map-matching [25] process is needed to extract traffic conditions of each road, making them difficult to meet the real-time requirements in practical applications, especially for short-term traffic prediction; 3) as to those road sensor data based traffic prediction applications, typically only certain parts of roads are installed with sensors, making such methods not be able to predict the traffic flow for the entire road network.

In this paper, we aim to build a model which, upon the completion of training and testing, only requests the historical monitoring traffic data of a limited number of roads to predict the traffic flow of the entire road network. As demonstrated in an example in Fig. 1(a), only three colored roads are selected as monitored roads to infer the traffic flow for all roads. This will

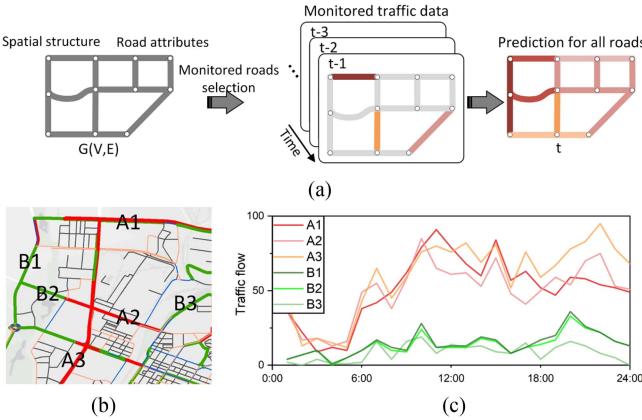


Fig. 1. An example of traffic prediction and traffic flow analysis. (a) An example that three colored roads are selected to be monitored, which allows making traffic prediction for all the roads in the next time slice t . (b) A part of the urban road network. Different colors indicate different road types, such as trunk roads (red color), secondary roads (green color), etc. (c) Traffic flow of different roads labeled in (b). It denotes that the traffic patterns between roads with similar attributes and closer spatial distance are more similar, making it feasible to select a part of roads to represent the global traffic conditions.

be of great significance for improving the efficiency of traffic forecasting at a lower cost in practice. As shown in Fig. 1(b) and (c), the traffic patterns are more similar between the roads which have similar attributes and closer spatial distance. Such strong correlations make it possible to represent the global traffic conditions through a limited number of roads.

In this report, we propose a novel deep-learning based model, named Limi-TFP, which can make real-time citywide traffic flow prediction by using historical data of a limited number of monitored roads. Specifically, an embedding module converts the spatial context and the attributes of each road into a vector. A road ranking method is then applied to select a limited number of monitored roads to represent the global traffic conditions. Moreover, a multi-head attention mechanism is adopted, which, by comprehensively considering the influences of spatial, temporal, and external factors, can capture the dynamic correlations between the target roads and the monitored roads.

The contributions of this work are summarized as follows:

- We present a novel model, called Limi-TFP, which achieves citywide traffic flow prediction by using the historical traffic data of a limited number of monitored roads;
- We propose a road ranking algorithm to select the monitored roads, which are used to represent global traffic conditions. Besides, an effective multi-head attention mechanism is adopted to capture the dynamic correlations between target roads and each monitored road;
- The proposed model comprehensively considers many complex factors, including the spatial structure of the whole road network, the attributes of each road, the POI distribution surrounding each road, the real-time weather conditions and the dynamic traffic parameters. Therefore, the model is with high robustness for different traffic conditions;
- The proposed model only needs the traffic parameters data of a limited number of monitored roads, so that the time cost of data processing is greatly reduced. Therefore, it is more

suitable for the real-time citywide traffic flow prediction, especially for the short-term traffic flow prediction task;

- We evaluate our model on two different traffic modes, taxis and buses, in Changchun city of China. Extensive experiments verify that our model, with only 5% roads being selected as monitored roads, achieves satisfactory prediction performances and has a stronger tolerance of data loss than the state-of-the-art methods.

The rest of this paper is organized as follows. Section II reviews the related works on traffic prediction. Section III formalizes the problem of Limi-TFP. Section IV elaborates on the proposed model. Experiments with real-word datasets are presented in Section V. Conclusion and future work are discussed in Section VI.

II. RELATED WORK

A. Time Series and Machine Learning Methods

Traffic prediction is an important component of ITS that has been intensively studied in the last few decades. Traditional time series based methods like ARIMA and its variations are widely used in forecasting tasks in many fields [10], [26]. Traffic conditions however are complicated and could be affected by many factors (e.g., weather conditions, traffic congestion, traffic accident, etc.), making it difficult for time series models to capture such complex external features. The performances of time-series models in traffic prediction thus have been unsatisfactory. Machine learning based methods, such as SVM [27] and XGBoost [28], can model more complex factors and perform well in traffic prediction fields. Such basic machine learning methods however require extensive initial work on feature selection, to which domain knowledge is essential.

B. Deep Learning for Traffic Flow Prediction

Deep Learning approaches have been successfully applied in many domains as they could automatically calculate the impacts of input variables on the targets. LSTM as one of the most famous deep learning techniques, can effectively capture the long-term and short-term patterns of traffic data and has shown superior performance in traffic prediction tasks [13], [14], [15]. Another well-known deep learning technique, convolutional neural network (CNN), which is good at modeling spatial correlations, has also been adopted for traffic prediction [29], [30], [31], [32]. These approaches transform traffic data into pixel values and then use CNN to achieve citywide traffic prediction. Some complex and innovative deep learning models have also been developed for traffic prediction [23], [24], achieving further improved performance and accuracy due to their enhanced abilities in capturing complex external features. More recently, approaches motivated by the graph convolution have been proposed, such as GCN [33] and STGCN [34]. Urban road network is of a typical graph structure, it is hence not a surprise that these graph convolution based methods achieve promising results in traffic prediction. The state-of-the-art methods, such as those reported in [18], [35], by combining the attention mechanism and encoder-decoder architecture with GCN, achieve excellent

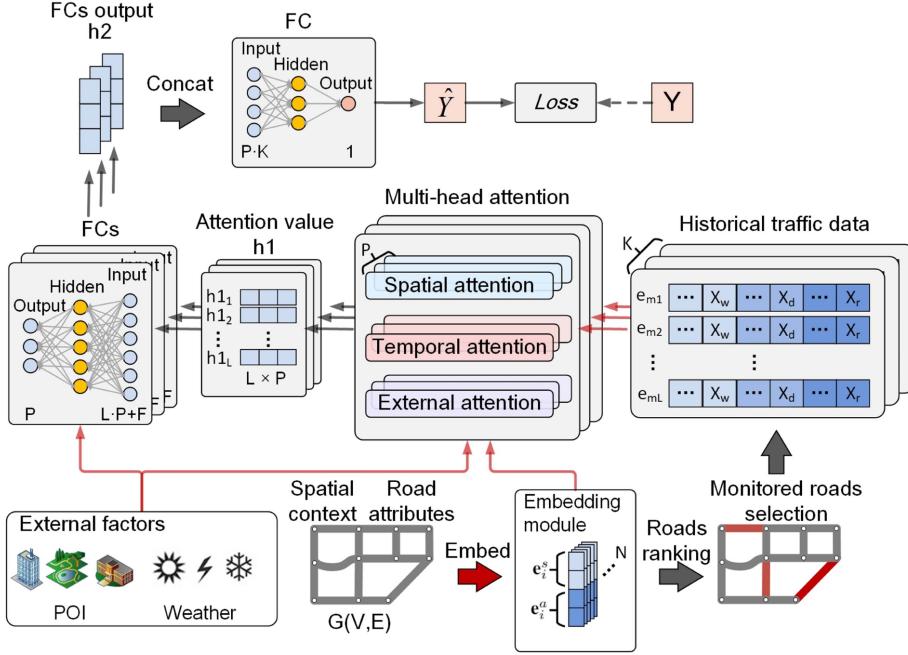


Fig. 2. The framework of Limi-TFP. The red arrow represents the input data flow of the model.

performances in traffic prediction, thanks to their strong capabilities in modeling dynamic spatio-temporal correlations.

While most of the work reported above request monitoring the whole road network. Our method, on the other hand, comprehensively considers the spatial structure of the whole road network and the attributes of each road, consequently the prediction task can be done with fewer monitored roads.

III. PROBLEM FORMULATION

Notations: The road network is represented as a directed graph $G = (V, E)$, where V stands for the set of intersections and E is the set of road segments. $N = |E|$ is the number of road segments in G . The monitored road set is represented as $E_m \subseteq E$, and the number of E_m is $L = |E_m|$, $L \ll N$. The elements $e_i \in E$ and $e_{mi} \in E_m$ represent the i -th road segment in E and E_m respectively. The monitored traffic data at time slice t is represented as $\mathbf{X}_t = (\mathbf{x}_t^{e_{m1}}, \mathbf{x}_t^{e_{m2}}, \dots, \mathbf{x}_t^{e_{mL}})$, where $\mathbf{x}_t^{e_{mi}} \in \mathbb{R}^K$ represents K types of basic traffic parameters (e.g., traffic flow, traffic speed, etc.) obtained on e_{mi} at time slice t . The input traffic data of the model (\mathbf{X}) is consisted of three time series segments: recent-period (\mathbf{X}_r), daily-period (\mathbf{X}_d) and weekly-period (\mathbf{X}_w), details are as follows:

$$\begin{aligned}\mathbf{X}_r &= (\mathbf{X}_{t-T_p}, \mathbf{X}_{t-(T_p-1)}, \dots, \mathbf{X}_{t-1}), \\ \mathbf{X}_d &= (\mathbf{X}_{t-T_d*T_p}, \mathbf{X}_{t-T_d*(T_p-1)}, \dots, \mathbf{X}_{t-T_d}), \\ \mathbf{X}_w &= (\mathbf{X}_{t-T_d*T_p*7}, \mathbf{X}_{t-T_d*(T_p-1)*7}, \dots, \mathbf{X}_{t-T_d*7}), \\ \mathbf{X} &= (\mathbf{X}_r, \mathbf{X}_d, \mathbf{X}_w) \in \mathbb{R}^{L \times T \times K},\end{aligned}\quad (1)$$

where T_p is the length of each time series segment, T_d is the number of time slices in one day, and T represents the total length of the input time slices.

TABLE I
KEY VARIABLES DESCRIPTION

Variable	Description
E_m	The selected Monitored road set.
N	The number of all roads.
L	The number of the monitored roads.
T	The number of total time slices.
K	The number of traffic parameter types.
P	The number of attention heads.
\mathbf{e}_i^s	Spatial embedding vectors of road e_i .
\mathbf{e}_i^a	Attributes embedding vectors of road e_i .
\mathbf{e}_i	Final embedding vector of road e_i .
s_{c_i}	Ranking score of a cluster c_i .
s_{e_j}	Ranking score of road e_j in the cluster c_i .
$score_{e_j}$	Final ranking score of road e_j .
$\mathbf{us}_{etarget,e_{mi}}^{j,k}$	The spatial correlations between target road e_{target} and monitored road e_{mi} at time slice j of the k -th type traffic parameter.
$\alpha_{etarget,e_{mi}}^{j,k,p}$	The normalized spatial attention score of the p -th dimension of $\mathbf{us}_{etarget,e_{mi}}^{j,k}$.
$\mathbf{ut}_{t_{target},t_j}^{i,k}$	The temporal correlation between the target time slice and the j -th time slice of the monitored road e_{mi} of the k -th type traffic parameter.
$\beta_{t_{target},t_j}^{i,k,p}$	The normalized temporal attention score in the p -th dimension of $\mathbf{ut}_{t_{target},t_j}^{i,k}$.
$\mathbf{ue}_{etarget,e_{mi}}^{j,k}$	The external correlations between target road e_{target} and monitored road e_{mi} at time slice j of the k -th type traffic parameter.
$\gamma_{etarget,e_{mi}}^{j,k,p}$	The normalized external attention score in the p -th dimension of $\mathbf{ue}_{etarget,e_{mi}}^{j,k}$.
$\mathbf{h1}$	The output of Multi-head attention module, $\mathbf{h1}_{i,k} \in \mathbb{R}^P$ represents the attention value of road e_{mi} in the k -th attention layer.
$\mathbf{h2}$	The output of FCs (the first fully-connected neural network module), $\mathbf{h2}_k \in \mathbb{R}^P$ is output of k -th layer.

Problem Statement: There are two tasks in our model. The first one is to select the monitored road set E_m , and the second one is to predict the traffic flow of the target time slice t for

all roads E . The input of the model is the historical monitored traffic data \mathbf{X} , road network G , and the external factors \mathcal{F} (including POIs, weather conditions, etc.). The output is $\hat{\mathbf{Y}} = (\hat{\mathbf{y}}_t^{e_1}, \hat{\mathbf{y}}_t^{e_2}, \dots, \hat{\mathbf{y}}_t^{e_N})$, where $\hat{\mathbf{y}}_t^{e_i}$ is the predicted value of road e_i at time slice t .

For readability, The key variables used in this paper are summarized in [Table I](#).

IV. MODEL DESCRIPTION

[Fig. 2](#) presents the overall framework of Limi-TFP. Specifically, an embedding module is adopted to convert the spatial context and attributes of road segments into vectors, so that we can learn the hidden spatial correlations between each road. According to the embedding vector of each road, we establish a road ranking method and select the top L roads as the monitored roads. We integrate K types of recent, daily and weekly historical traffic data input into K separate attention layers. Each attention layer is composed of spatial, temporal and external factors attention mechanisms, which can dynamically assign different weights to the input data, and each road segment will generate an output $\mathbf{h1}_{i,k} \in \mathbb{R}^P$, which represents the attention value of road e_{mi} in the k -th attention layer. Next, we feed the attention values of each attention layer and the external factors into fully-connected layers (FCs); the output of the k -th fully-connected layer is $\mathbf{h2}_k \in \mathbb{R}^P$. Finally, another fully-connected layer (FC) is used to connect the output values of FCs. The final output $\hat{\mathbf{Y}}$ is the predicted traffic flow of each road. We detail each step in the following subsections.

A. Embedding Module

The traffic flow transfer is closely related to the spatial structure of the road network, so it is critical to capture the spatial correlations between each road for traffic prediction. To this end, in the embedding module, we leverage the node2vec [36] approach to convert spatial structure information of each road into a vector. Specifically, we use random walk to obtain the context information (directly adjacent roads in a random walk trajectory) of each road. Considering the driving pattern of vehicles, it is unlikely that they will choose an already traveled road, and they tend to drive further away from the roads they have passed through during a period of time. So we adopt a depth-first search (DFS) strategy [36] in random walk procedure. Afterward, we use the Skip-Gram model [37] to learn the road representations, which adopts a two-layer fully-connected neural network and takes the one-hot encoding of the specific road and the context roads as input and output respectively. Finally, we obtain the spatial embedding vectors by multiplying the one-hot encoding of the input road and the learned weight matrix, represented as $\mathbf{e}_i^s \in \mathbb{R}^D$.

Besides the spatial structure, the road attributes (e.g., road type, road length, number of lanes, etc.) are also important factors affecting traffic prediction. We thus adopt an autoencoder [38], which is composed of a two-layer neural network, and takes the attributes encoding as input and output simultaneously. The attributes encoding consists of one-hot encoding of

road type, the normalized road length and the number of lanes. The attributes embedding vectors are represented as $\mathbf{e}_i^a \in \mathbb{R}^D$.

Finally, to preserve both the spatial structure and road attributes information, the final embedding vector of road e_i is represented as $\mathbf{e}_i = \mathbf{e}_i^s || \mathbf{e}_i^a$, where $||$ represents the concatenation operation.

B. Monitored Roads Selection

The selection of the representative roads to capture the global traffic conditions is a key issue in our model. It can be observed from [Fig. 1\(b\)](#) and [\(c\)](#) that, the traffic flow correlations is closely related to the spatial structure and attributes of each road. Accordingly, our monitored roads selection strategy is to use the k -means method to divide the roads into different clusters according to the hidden spatial and attributes information contained in the embedding vectors, where all the cluster centers are selected as initial monitored road set. More monitored roads will be iteratively selected based on a road ranking method.

Intuitively, the similarity between each element is higher in a smaller and denser cluster. To avoid over-selecting monitored roads in small clusters, we set higher ranking scores for the roads in large and sparse clusters. For a specific cluster, we assign higher ranking scores to the roads farther away from the cluster center and those selected monitored roads.

Specifically, we define the ranking score of a cluster c_i as follows:

$$s_{c_i} = e^{-\frac{1}{2} \left(\left(1 - \frac{|c_i|}{N} \right) + \left(1 - \frac{\bar{d}_{c_i}}{\max(\bar{d}_C)} \right) \right)}, \quad (2)$$

where $|c_i|$ is the number of road segments in cluster c_i , \bar{d}_{c_i} represents the average euclidean distance between all embedding vectors and the cluster center in c_i , and $\max(\bar{d}_C)$ represents the maximum average euclidean distance value of all the clusters.

The score of each road segment e_j in the cluster c_i is defined as follows:

$$s_{e_j} = e^{-\frac{1}{2} \left(\left(1 - \frac{d(e_j, e_{c_i})}{\max(d_{c_i})} \right) + \left(1 - \frac{\min(e_j, E_m)}{\max(d_m)} \right) \right)}, \quad (3)$$

where $d(e_j, e_{c_i})$ denotes the euclidean distance of the embedding vectors between road e_j and the cluster center e_{c_i} , $\max(d_{c_i})$ is the maximum euclidean distance between each road and e_{c_i} in cluster c_i , $\min(e_j, E_m)$ represents the minimum euclidean distance between e_j and the selected monitored road set E_m , and $\max(d_m)$ represents the maximum value of the minimal euclidean distance between E_m and the other roads.

Finally, the ranking score for a specific road e_j is defined as $score_{e_j} = s_{c_i} \cdot s_{e_j}$. We select one road with the highest ranking score value and add it to E_m , and then update the scores of all roads. Repeat this process until a total of L road segments are selected as monitored roads.

C. Attention Layer

[Fig. 3](#) illustrates an example of the p -th layer of the proposed multi-head attention mechanism, which contains three components, namely spatial attention, temporal attention and external attention, respectively. To capture the long-term and short-term temporal features, we integrate the recent, daily and weekly

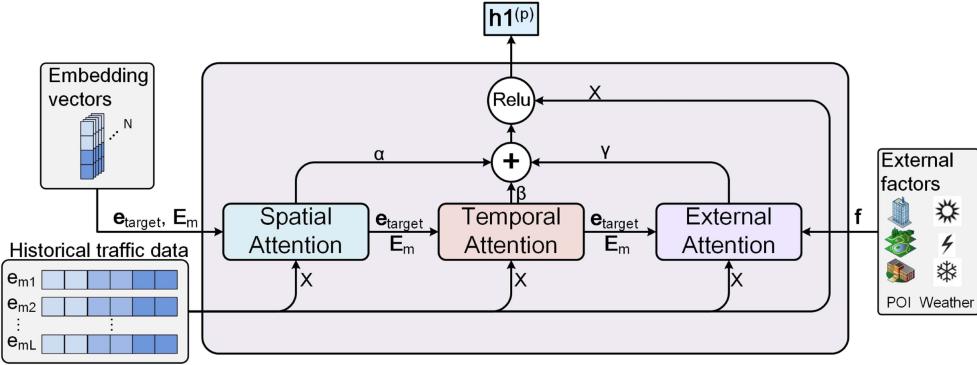


Fig. 3. The p -th layer of multi-head attention mechanism.

historical data of E_m as input. For convenience, we let $x_{i,j,k}$ denote the k -th dimension traffic data of road e_{mi} at the j -th time slice. We describe each attention component in detail in the following subsections.

1) *Spatial Attention*: The traffic flow of each target road has different correlations with each monitored road due to the different hidden distance between their embedding vectors. Besides, the correlations are highly dynamic under different traffic conditions. For instance, while the traffic conditions of two adjacent roads are almost the same or have a relatively stable proportion over a period of time, a congestion on a road however may significantly affect the traffic conditions and cause cascading effects on other roads. Motivated by such observations, we design a spatial attention mechanism to capture such dynamically changing correlations.

Specifically, we concatenate the hidden embedding vectors through a matrix, and adopt a non-linear transformation to compute the relevance between the target road e_{target} and the monitored road e_{mi} at a certain time slice as follows:

$$\mathbf{us}_{e_{target}, e_{mi}}^{j,k} = \mathbf{V}_s \tanh (\mathbf{w}_{s2} (\mathbf{e}_{target}^\top \mathbf{W}_{s1} \mathbf{e}_{mi}) + \mathbf{w}_{s3} x_{i,j,k} + \mathbf{b}_s), \quad (4)$$

where $\mathbf{us}_{e_{target}, e_{mi}}^{j,k} \in \mathbb{R}^P$ represents the dynamic correlations between the predicted target road and the i -th monitored road in E_m . The learnable parameters are $\mathbf{W}_{s1} \in \mathbb{R}^{2D \times 2D}$, $\mathbf{V}_s \in \mathbb{R}^{P \times P}$, and $\mathbf{w}_{s2}, \mathbf{w}_{s3}, \mathbf{b}_s \in \mathbb{R}^P$. Note that we use multi-head attention in each attention layer to jointly attend to the information from different representation sub-spaces at different positions. Then the p -th dimension of $\mathbf{us}_{e_{target}, e_{mi}}^{j,k}$ will be normalized via softmax function:

$$\alpha_{e_{target}, e_{mi}}^{j,k,p} = \frac{\exp(u_{e_{target}, e_{mi}}^{j,k,p})}{\sum_{e_m \in E_m} \exp(u_{e_{target}, e_{mi}}^{j,k,p})}. \quad (5)$$

2) *Temporal Attention*: The time factors also have significant influences on the traffic prediction. Traffic conditions at a specific road segment may have periodic changes at different times of the day, and also have different patterns on weekdays and weekends. Therefore, the target road and each monitored road have different correlations at different time. Generally, for the monitored road segments with small hidden distances from

the target road, the traffic conditions in the previous several time slices may have a higher correlation with the condition in the predicted time slice. However, for those monitored roads farther away, their impacts on the target road may only be felt after a relatively longer delay. Besides, traffic conditions may have a strong influence on the temporal correlations. For example, a congestion or a traffic accident occurring on a road segment may affect the traffic for a few hours. To model such properties, we design a temporal attention mechanism where the correlation between the target road e_{target} and a monitored road e_{mi} at the j -th time slice is defined as:

$$\begin{aligned} \mathbf{ut}_{t_{target}, t_j}^{i,k} &= \mathbf{V}_t \tanh (\mathbf{w}_{t2} (\mathbf{e}_{target}^\top \mathbf{W}_{t1} \mathbf{e}_{mi}) \\ &\quad + \mathbf{w}_{t4} (\boldsymbol{\tau}_{target}^\top \mathbf{W}_{t3} \boldsymbol{\tau}_j) \\ &\quad + \mathbf{w}_{t5} x_{i,j,k} + \mathbf{b}_t), \end{aligned} \quad (6)$$

$$\beta_{t_{target}, t_j}^{i,k,p} = \frac{\exp(u_{t_{target}, t_j}^{i,k,p})}{\sum_{t' \in T} \exp(u_{t_{target}, t'}^{i,k,p})}, \quad (7)$$

where $\mathbf{ut}_{t_{target}, t_j}^{i,k} \in \mathbb{R}^P$ represents the dynamic correlation between the target time slice and the j -th time slice of the monitored road e_{mi} , and $\beta_{t_{target}, t_j}^{i,k,p}$ is the temporal attention score in the p -th head of the k -th attention layer. Similar to that in the road attributes embedding, we transform the time features (e.g., day-of-week, time-of-day, weekday-or-weekend, etc.) through a two-layer autoencoder neural network. $\boldsymbol{\tau}_{target}, \boldsymbol{\tau}_j \in \mathbb{R}^D$ denote the time embedding vectors of the target time slice and j -th time slice respectively. The learnable parameters include $\mathbf{W}_{t1} \in \mathbb{R}^{2D \times 2D}$, $\mathbf{W}_{t3} \in \mathbb{R}^{D \times D}$, $\mathbf{V}_t \in \mathbb{R}^{P \times P}$, and $\mathbf{w}_{t2}, \mathbf{w}_{t4}, \mathbf{w}_{t5}, \mathbf{b}_t \in \mathbb{R}^P$.

3) *External Attention*: Besides spatiotemporal correlations, external factors (such as POI, weather, etc.) may also significantly affect the correlations between the target roads and the monitored roads [39]. For example, in extreme weather conditions, the traffic conditions of certain roads will be severely affected, and people may change their driving routes, thus the correlations between roads will change accordingly. For different urban functional regions, their traffic patterns are also different. For example, in working regions, traffic flow tends to be higher in the morning and evening peaks, while in entertainment regions, the pattern may be totally different. To capture



Fig. 4. Study area. (a) Road clusters and bus stations. There are five clusters represented by different colors, the blue dots represent bus stations. There are totally 6,404 roads and 4,785 bus stations in our study area. (b) Monitored roads selection. The red lines represent the roads ranked in the top 5%, the green lines represent the roads ranked between 5% and 10%.

these properties, we transform the external factors of a road at each time slice into a D -dimensional vector, and the external attention score is calculated as follows:

$$\begin{aligned} \mathbf{ue}_{e_{target}, e_{mi}}^{j,k} &= \mathbf{V}_e \tanh (\mathbf{w}_{e2}^\top (\mathbf{e}_{target}^\top \mathbf{W}_{e1} \mathbf{e}_{mi}) \\ &\quad + \mathbf{w}_{e4} (\mathbf{f}_{target}^\top \mathbf{W}_{e3} \mathbf{f}_{mi,j}) \\ &\quad + \mathbf{w}_{e5} x_{i,j,k} + \mathbf{b}_e), \end{aligned} \quad (8)$$

$$\gamma_{e_{target}, e_{mi}}^{j,k,p} = \frac{\exp (ue_{e_{target}, e_{mi}}^{j,k,p})}{\sum_{e_m \in E_m} \exp (ue_{e_{target}, e_m}^{j,k,p})}, \quad (9)$$

where \mathbf{f}_{target} and $\mathbf{f}_{mi,j}$ represent the external embedding vectors of the target road at target time slice and the monitored road e_{mi} at j -th time slice, respectively. Similarly, $\mathbf{ue}_{e_{target}, e_{mi}}^{j,k} \in \mathbb{R}^P$, and $\gamma_{e_{target}, e_{mi}}^{j,k,p}$ is the normalized external attention score in the p -th head of the k -th attention layer. The learnable parameters include $\mathbf{W}_{e1} \in \mathbb{R}^{2D \times 2D}$, $\mathbf{W}_{e3} \in \mathbb{R}^{D \times D}$, $\mathbf{V}_e \in \mathbb{R}^{P \times P}$, and $\mathbf{w}_{e2}, \mathbf{w}_{e4}, \mathbf{w}_{e5}, \mathbf{b}_e \in \mathbb{R}^P$.

4) *Attention Components Fusion*: As shown in Fig. 3, we fuse the spatial, temporal and external factors attention components to generate the hidden state vector of each monitored road. $\mathbf{h1}_{i,k} \in \mathbb{R}^P$ denotes the final output of road e_{mi} at the k -th attention layer. The p -th dimension of $\mathbf{h1}_{i,k}$ is defined as follows:

$$\begin{aligned} h1_{i,k}^p &= \text{ReLU} \left(w_\alpha \sum_{j=1}^T \alpha_{e_{target}, e_{mi}}^{j,k,p} x_{i,j,k} \right. \\ &\quad + w_\beta \sum_{j=1}^T \beta_{t_{target}, t_j}^{i,k,p} x_{i,j,k} \\ &\quad \left. + w_\gamma \sum_{j=1}^T \gamma_{e_{target}, e_{mi}}^{j,k,p} x_{i,j,k} + b1 \right), \end{aligned} \quad (10)$$

where $w_\alpha, w_\beta, w_\gamma$ and $b1$ are learnable parameters. We use ReLU as the activation function.

D. Fully-Connected Layers

We build a fully-connected neural network for each attention layer to receive the expanded hidden state vectors and the external factors (e.g. POI, weather conditions, etc.). The input dimension is $\mathbb{R}^{L \cdot P + F}$ (F is the number of external factors), and the output dimension is \mathbb{R}^P . Afterwards, we use another fully-connected neural network to combine the outputs of all the previously fully-connected layers.

Finally, we train the proposed model Limi-TFP via back-propagation by minimizing the mean squared error (MSE) between the predicted values $\hat{\mathbf{Y}}$ and the ground truth \mathbf{Y} :

$$\mathcal{L}(\Theta) = \left\| \hat{\mathbf{Y}} - \mathbf{Y} \right\|_2^2, \quad (11)$$

where Θ denotes all learnable parameters in Limi-TFP.

V. EXPERIMENTS

A. Datasets

We validate our model on two different traffic modes, taxis and buses, over the same period in Changchun city, northeast in China. Specifically, we integrate the trajectory data of approximately 2,000 taxis and 1,500 buses, from May 1, 2017 to July 30, 2017. The taxi data is available for 24 hours each day, while the bus data is available for 15 hours (5 am to 8 pm) per day. For both of them, the GPS sampling interval is 30 seconds for each trajectory. Parts of the data have been released in simple formats for easy access, and could be found in our previous publication [40]. We select the main urban area of Changchun city as the research area, which covers 6,404 roads and 4,785 bus stations, as shown in Fig. 4.

B. Data Preprocessing

We first use an HMM based map matching method [25] to match all the GPS trajectories to the road network. In this way we can transfer the low sampling rate of GPS trajectories to a real driving path, including all the passing through road segments for

TABLE II
EXPERIMENTAL RESULTS COMPARISON OF DIFFERENT APPROACHES FOR TRAFFIC PREDICTION ON TAXI AND BUS DATASETS

Data	Methods	15 min			30 min			60 min		
		MAE	RMSE	MAPE	MAE	RMSE	MAPE	MAE	RMSE	MAPE
Taxi data	ARIMA	1.87	2.77	70.52%	2.80	3.94	67.63%	3.87	5.87	64.30%
	SVR	1.54	2.59	56.91%	2.73	3.75	51.34%	3.78	5.63	41.18%
	LSTM	1.47	2.48	42.56%	2.76	3.67	47.56%	3.70	5.61	40.43%
	LC-RNN	1.46	2.33	43.30%	2.69	3.59	46.94%	3.69	5.59	39.97%
	STGCN	1.38	2.25	39.86%	2.54	3.51	43.81%	3.57	5.51	38.29%
	GSTNet	1.41	2.13	38.75%	2.46	3.48	38.67%	3.44	5.40	36.13%
	GMAN	1.36	2.15	42.54%	2.41	3.37	36.32%	3.38	5.33	35.54%
	Limi-TFP-5%	1.43	2.18	41.70%	2.37	3.30	36.11%	3.31	5.27	34.69%
Bus data	Limi-TFP-TR-5%	1.39	2.15	39.12%	2.31	3.18	33.46%	2.89	5.13	30.86%
	ARIMA	1.54	2.65	43.26%	1.77	3.45	38.27%	2.45	4.68	34.30%
	SVR	1.42	2.43	37.33%	1.68	3.41	33.95%	2.42	4.56	31.41%
	LSTM	1.33	2.39	38.65%	1.59	3.38	31.63%	2.39	4.34	29.19%
	LC-RNN	1.35	2.32	35.37%	1.61	3.35	30.44%	2.34	4.29	28.44%
	STGCN	1.20	2.28	33.29%	1.42	3.24	29.72%	2.26	4.01	26.26%
	GSTNet	1.17	2.24	31.86%	1.39	3.22	27.03%	2.23	3.93	25.85%
	GMAN	1.18	2.19	29.54%	1.33	3.19	26.38%	2.15	3.88	24.98%
Bus data	Limi-TFP-5%	1.14	2.17	30.01%	1.31	3.17	26.01%	2.13	3.82	24.73%
	Limi-TFP-TR-5%	1.08	2.13	28.94%	1.22	3.06	25.58%	1.83	3.76	24.15%

each trajectory. Then we calculate the amount of the traffic flow and the average speed of traffic along each road in each time unit (5 minutes). The missing values are filled by the random normal distribution interpolation, which are calculated from the same time slice in historical data. We use July 1–16 as the target time period for training, July 17–23 for validation, and the last week in July 24–30 for testing. The length of time slice is set to 15 minutes, 30 minutes and 60 minutes respectively to test the performance of the proposed model on two datasets.

C. Experimental Settings

1) *Metrics*: To measure the prediction performance of different methods, we adopt three widely used metrics including Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), and Mean Absolute Percentage Error (MAPE), respectively. For each of these metrics, a smaller value indicates a better prediction performance.

2) *Hyperparameters*: The dimension of the input vector $\mathbf{X} = (\mathbf{X}_r, \mathbf{X}_d, \mathbf{X}_w) \in \mathbb{R}^{L \times T \times K}$ should be confirmed first, it determines how to integrate the input data. We set $K = 2$ because two types of traffic parameters data (traffic flow and average speed) are used in this paper; we set the length of \mathbf{X}_r , \mathbf{X}_d and \mathbf{X}_w is 8, so the total length of \mathbf{X} is $T = 24$; L is the number of the monitored roads. The other two dimensions also need to be confirmed, we set the number of attention heads is $P = 5$, the dimension of each embedding vector is $D = 64$. The number of clusters is a key point in monitored roads selection strategy, we find when it equals 5, the model achieves the best performance. The results of clustering roads into 5 categories are shown in Fig. 4(a). The initial learning rate is an important parameter in the training process, as the back propagation (BP) algorithm based on the gradient descent, if initial learning rate is too big, the model cannot converge; On the contrary, if it is too small, it will lead to a significant increase in training time, and the algorithm will get a local minimum with a certain probability. We set the initial learning rate is 0.01, the model performs excellent training effect.

D. Baseline Methods

We compare our model with the following baselines:

- **ARIMA** [10]: Auto-Regressive Integrated Moving Average method, a widely used time-series method.
- **SVR** [41]: Support Vector Regression, a regression method based on Support Vector Machine.
- **LSTM** [13]: Long Short Term Memory network, which solves the problem of gradient disappearance.
- **LC-RNN** [23]: Look-up Convolution Recurrent Neural Network, which adopts a road network embedded convolution method to learn more meaningful spatial features.
- **STGCN** [34]: Spatial-Temporal Graph Convolutional Network based on spatial method.
- **GSTNet** [42]: Global Spatial-Temporal Network for traffic flow prediction, which consists of several layers of spatial-temporal blocks.
- **GMAN** [35]: Graph Multi-Attention Network, which adapts an encoder-decoder architecture to model the impact of the spatio-temporal factors.

E. Results

1) *Performance Comparisons*: We compare the performance of the proposed model with some widely used existing methods and some latest graph-based methods. Note that Limi-TFP-5% means that the proposed model is with 5% roads being monitored. Specifically, Limi-TFP-TR-5% represents that, when predicting the traffic flow of each target road, the target road itself is included in the monitored road set. The performance of each model with different lengths of time slice are shown in Table II.

The main observations may be summarized as follows:

- Deep learning methods achieve significantly better performance than traditional time series method (ARIMA) and machine learning method (SVR). The graph-based models, e.g., STGCN, GSTNet, and GMAN, meanwhile outperform other deep learning methods, e.g., LSTM and

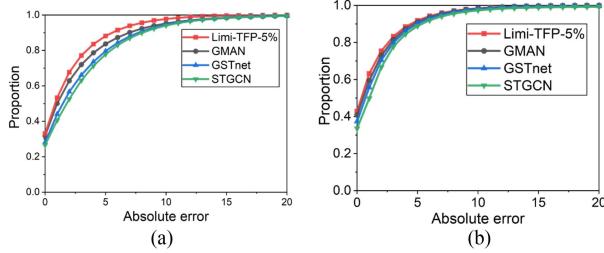


Fig. 5. Cumulative probability density of absolute error. (a) Taxi dataset. (b) Bus dataset.

LC-RNN. Such observations indicate that the spatiotemporal characters could help improve the accuracy for traffic flow prediction.

- Though the proposed model only requests the historical traffic data of 5% monitored roads as inputs, however, Limi-TFP-5% and Limi-TFP-TR-5% manage to achieve better prediction accuracy, especially for long-term forecasting (30 minutes or 1 h). It confirms that our model will be of great significance for improving the efficiency of traffic forecasting with lower data collection cost, and has obvious application advantages compared with the existing methods.
- Comparing Limi-TFP-5% and Limi-TFP-TR-5%, we could find that, by adding the target roads to the monitored road sets, the prediction accuracy could be further improved measured by all metrics for both datasets.
- Comparing the prediction accuracy of all methods on the two datasets in Table II, we could find that all methods perform better predictions on bus dataset in terms of all evaluation metrics. It is mainly because the route in the road network is fixed for each bus, and the departure time interval is relatively stable, which allows the spatiotemporal characteristics to be better captured in the prediction process.

For further analysis, we depict the cumulative distribution function (CDF) of the absolute prediction errors (time slice length is 1 h) of the proposed model and the state-of-the-art methods in Fig. 5. It shows that Limi-TFP-5% achieves smaller absolute errors in most proportions compared to other baselines, including GMAN, GSTnet, and STGCN.

2) *The Impact of Monitored Roads Selection:* To investigate the impact of monitored roads selection on the performances of our model, we compare the proposed method against two other selection methods: (1) random selection; (2) selecting roads according to historical average traffic flow from high to low. Then we test the performance of each method (time slice length is 1 h) under different percentages (from 1% to 10%) of monitored roads. Fig. 6 presents the MAE of the tests on the two datasets. It can be seen that our method exhibits significantly better prediction results, and when the proportion reaches 5%, our method achieves best performance. For method (2) and our method, at the beginning, the MAE gradually decreases as the proportion of the monitored roads increases. As the proportion continues to increase, instead, MAE increases slightly. It indicates that, if the number of monitored roads is too large, the

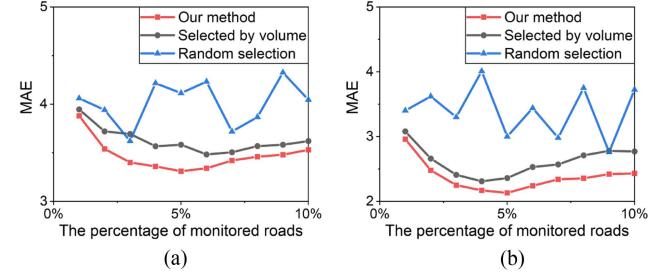


Fig. 6. The impact of monitored roads selection. (a) Taxi dataset. (b) Bus dataset.

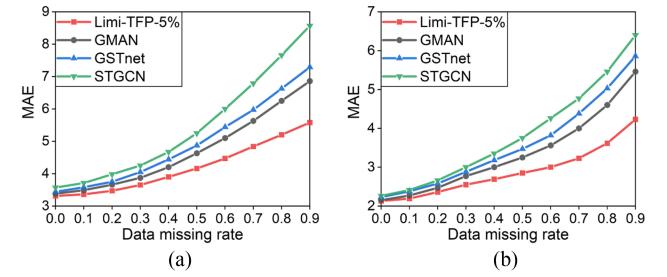


Fig. 7. Data missing tolerance comparison. (a) Taxi dataset. (b) Bus dataset.

TABLE III
THE COMPUTATION TIME COMPARISON ON TAXI DATASET (WITH 24 GB RAM, 3.20 GHZ INTEL(R) CORE(TM) i7-8700 CPU)

Methods	Training (hours/epoch)	Inference (seconds/slice)
STGCN	0.14	16
GSTNet	2.04	97
GMAN	0.89	8
Limi-TFP-5%	3.18	25

correlation between the target road and each monitored road will be disturbed, resulting in the decline of the model accuracy. For random selection method, the change of MAE is irregular and fluctuates greatly, which proves that the selection of monitored roads has important impacts on the accuracy of the model.

3) *Data Missing Tolerance Comparisons:* Data missing is a common phenomenon in traffic data collection, due to frequent equipment failures, poor signals, etc. A stronger tolerance of data missing helps make the traffic prediction model be more robust. To evaluate the tolerance for data missing, we randomly drop a fraction ranging from 10% to 90% of historical traffic data, then compare the performance of the proposed model versus the state-of-the-art methods (with time slice being 1 h). As shown in Fig. 7, Limi-TFP-5% shows obviously highest tolerance for data missing. This indicates that the proposed method can effectively use the limited traffic data to capture the global characteristics of the entire road network.

4) *Computation Time:* Table III presents the training time and inference time of STGCN, GSTNet, GMAN and Limi-TFP-5% on the taxi dataset for one hour ahead prediction. Similar conclusions can be extended to bus dataset. For a fair comparison, the training time is computed on one epoch, and the inference time is operated on one time slice prediction for all roads. Due to the time-consuming spatiotemporal correlation

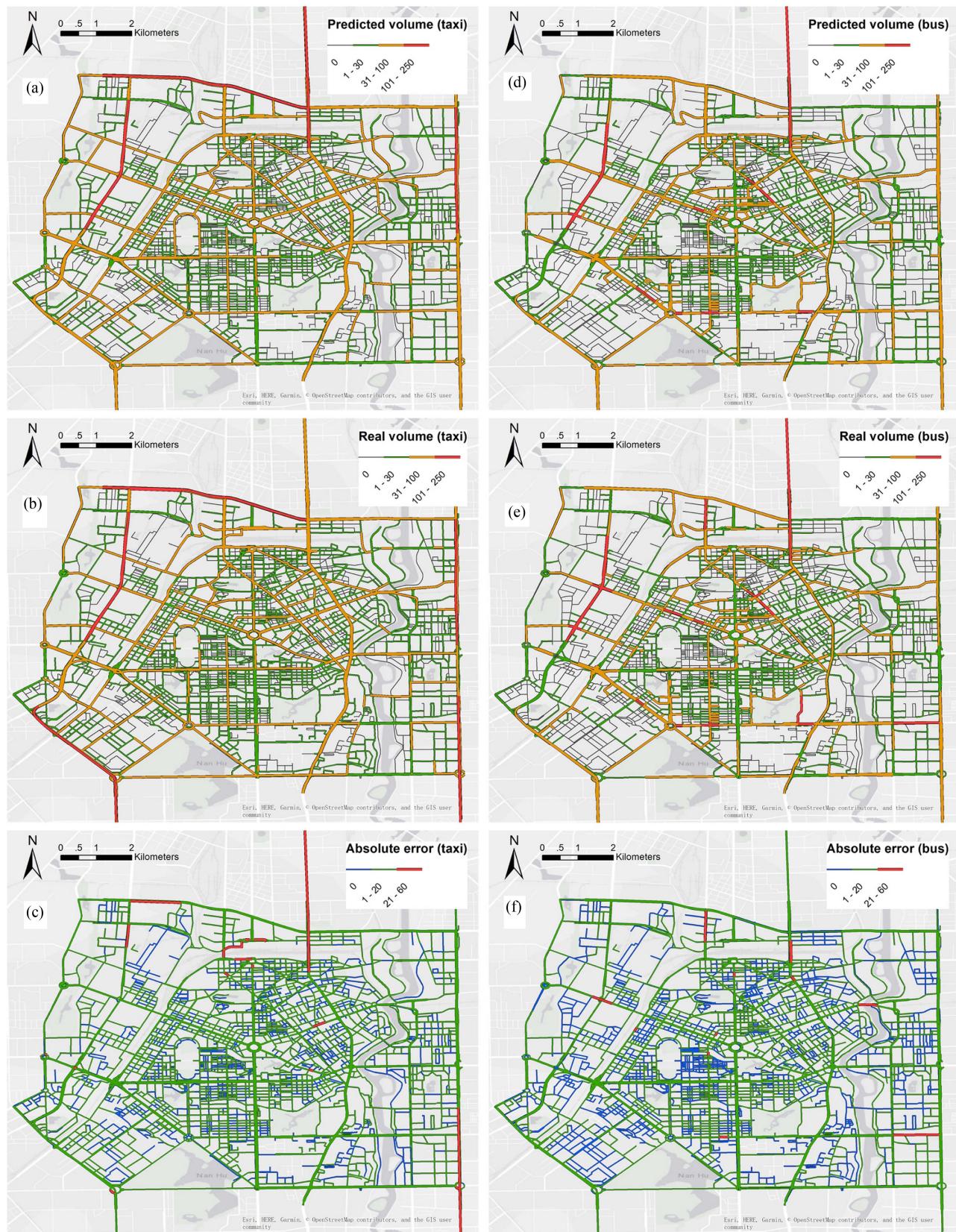


Fig. 8. Visualization examples. Visualizations of the predictions, the ground truth, and the prediction errors for a specific time period 7:00 am–8:00 am on July 24, 2017, on two datasets: (a) the predicted taxi flows; (b) the ground truth of taxi flows; (c) the absolute errors of taxi flows on each road; (d) the predicted bus flows; (e) the ground truth of taxi flows; and (f) the absolute errors of bus flows on each road.

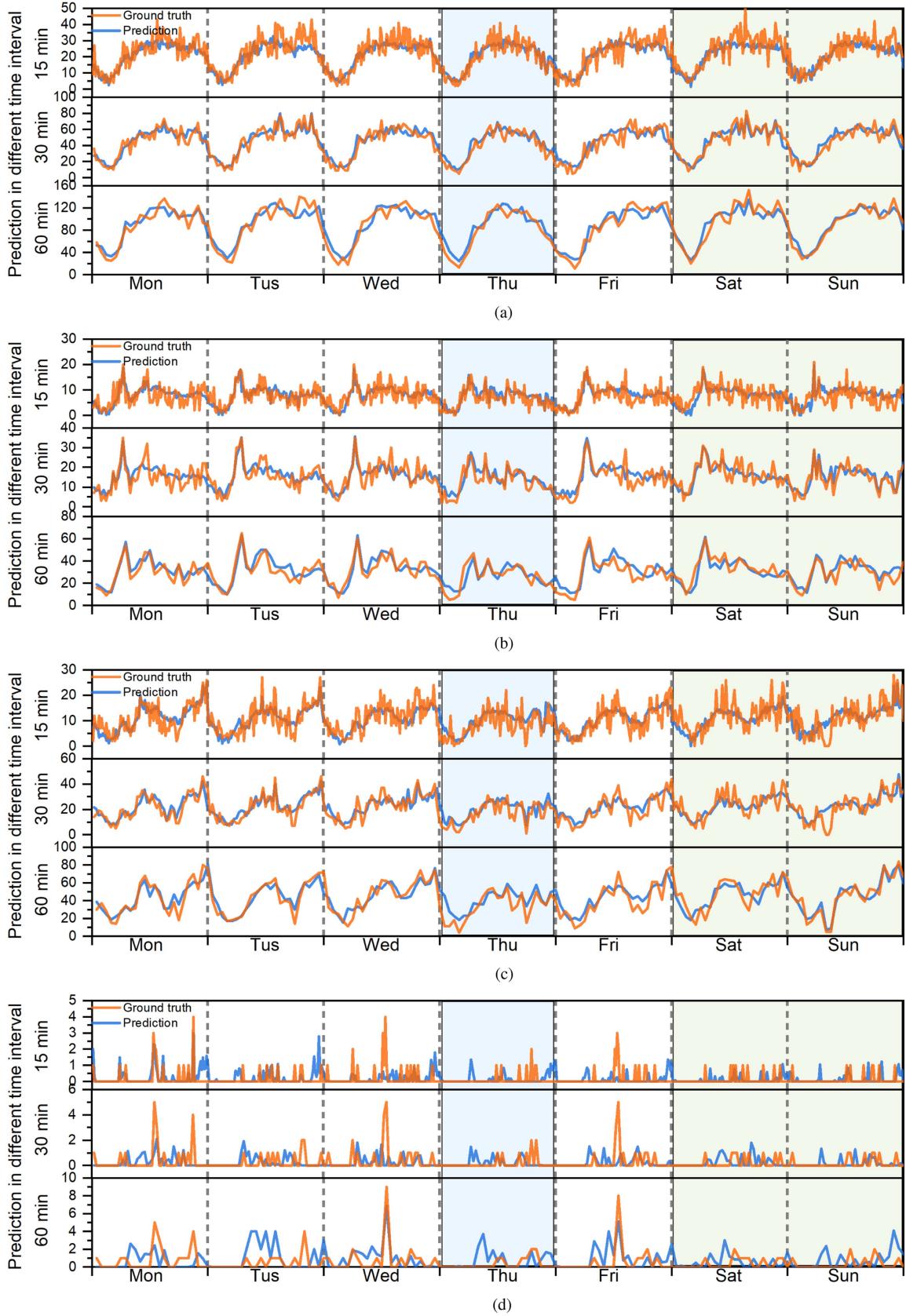


Fig. 9. The traffic flow prediction results of representative roads. The prediction curves and ground truth of four roads (trunk road, secondary road, tertiary road and service road) in one week are shown from (a) to (d). The light blue background represents there was a thunder shower on Thursday. The light green background represents the prediction results on weekends. (a) Trunk road prediction results. (b) Secondary road prediction results. (c) Tertiary road prediction results. (d) Service road prediction results.

TABLE IV
THE PREDICTION ERRORS (MAE) UNDER DIFFERENT SCENARIOS ON TAXI DATASET

Time interval	Roads	Morning peak	Evening peak	Rainy day	Weekdays	Weekends
15 min	Road a	4.08	5.21	3.52	4.23	4.48
	Road b	2.32	2.57	1.82	2.10	2.66
	Road c	3.23	3.60	3.19	3.02	3.41
	Road d	0.09	0.27	0.23	0.25	0.19
	All roads	1.28	1.26	1.27	1.22	1.23
30 min	Road a	6.32	4.98	5.62	5.95	5.31
	Road b	2.89	3.16	2.81	3.21	2.67
	Road c	5.59	5.83	4.67	4.38	5.20
	Road d	0.48	0.64	0.4	0.43	0.32
	All roads	2.08	1.97	1.97	1.84	1.88
60 min	Road a	10.39	10.82	10.09	10.94	7.90
	Road b	3.64	5.35	5.21	5.16	4.82
	Road c	4.93	9.81	8.63	7.02	6.98
	Road d	1.44	0.71	0.75	0.91	0.81
	All roads	3.51	3.18	3.45	2.99	3.09

learning for each target road, the proposed model runs much slower than other methods in the training stage. However, it only takes 25 seconds in inference stage, which can meet the real-time requirements in practical applications.

5) *Visualization Examples of Prediction Results:* For an intuitive observation of the prediction, we further present the visualization of the predictions, the ground truth, and the prediction errors on the map. A specific Monday morning peak period (7:00 am to 8:00 am) on July 24, 2017, is selected for the visualization, which is shown in Fig. 8. Fig. 8(a) and (d) show the predictions, Fig. 8(b) and (e) show the real traffic flow of taxis and buses. Fig. 8(c) and (f) present the absolute errors for the two datasets respectively. It can be observed that, Limi-TFP achieves high accuracy for most roads (with the errors less than 20), only a small portion of roads close to the border of the road network have relatively higher errors. Such exceptions may due to the fact that some closely related roads are not included in the study area.

6) *Prediction Results Under Different Scenarios:* To further evaluate the effectiveness of Limi-TFP, we select some representative roads to test the traffic flow prediction results under different scenarios on taxi dataset, as shown in Fig. 9. We visualize the 15, 30 and 60 minutes ahead prediction results of four types of roads (i.e., trunk road in Fig. 9(a), secondary road in Fig. 9(b), tertiary road in Fig. 9(c) and service road in Fig. 9(d)) during one week, from Jul 24, 2017 (Monday) to Jul 30, 2017 (Sunday). Specially, there was a thunder shower on Thursday, Jul 27, 2017.

We can observe that the traffic flow pattern is influenced by many conditions:

- The traffic flow of different road types has different modes. The traffic flow of the selected trunk road is higher than that of other roads. The traffic flow of the service road is much lower and relatively discrete. The selected tertiary road is located in the central business district of the city, named “Hongqi Road,” so its traffic flow is higher than the selected secondary road.
- Weather conditions have important impacts on traffic patterns. A thunder shower happened on Thursday, we can see that the traffic flow of Thursday morning peak, evening peak and the whole day is lower than that of other days.

This phenomenon is particularly obvious in Fig. 9(b) and (c), which indicates that the traffic conditions of these two roads are more susceptible to weather factors.

- In Fig. 9(c), the traffic flow pattern shows obvious commercial street characteristics. It has obvious evening peak, the traffic flow at night is significantly higher than that during daytime in general. It is sensitive to weather conditions, the abnormal weather makes the traffic flow drop significantly.
- The traffic flow shows the property of periodicity, and the traffic flow patterns are different in weekdays and weekends. The traffic flow during morning peak is relatively low on weekends, especially on Sunday. In Fig. 9(c), we can observe that the traffic flow in Sunday evening is significantly higher than that on weekdays.

In Fig. 9, we can observe that the prediction curves have high fitting degree with ground truth in different scenarios. The trend and order of magnitude of the predicted values are consistent with the ground truth, even the ground truth curves are quite complicated. In addition, we use MAE to evaluate the performance of Limi-TFP in morning peak, evening peak, rainy day, weekdays and weekends. The MAE of the selected four roads and all roads in different time interval is shown in Table IV. We can see that although MAE increases with the increase of the real traffic flow, in most cases, it is relatively low and within an acceptable range.

In summary, Limi-TFP performs reasonable prediction results under different complicated scenarios, including different road types, different weather conditions, different POI distribution, different day of the week and different time periods, which indicates that the proposed model can handle complex traffic conditions in different time interval.

VI. CONCLUSION

In this paper, we proposed a novel model named Limi-TFP. Different from existing methods, the proposed model only requests monitoring the traffic conditions of a limited number of roads to achieve citywide scale traffic flow prediction. Such a method allows us to significantly improve the efficiency of traffic forecasting at a lower cost. Experiments on two real-world datasets demonstrated that, with proper

selection of the monitored roads, taking only 5% of the roads as the monitored roads would allow us to achieve the better performance and stronger tolerance of data loss compared to the state-of-the-art algorithms. In future studies, we will evaluate the proposed model on other datasets and extend for multi-step prediction on different topological structures.

REFERENCES

- [1] X. Song, Y. Guo, N. Li, and L. Zhang, "Online traffic flow prediction for edge computing-enhanced autonomous and connected vehicles," *IEEE Trans. Veh. Technol.*, vol. 70, no. 3, pp. 2101–2111, Mar. 2021.
- [2] L. Zhu, F. R. Yu, Y. Wang, B. Ning, and T. Tang, "Big Data analytics in intelligent transportation systems: A survey," *IEEE Trans. Intell. Transp. Syst.*, vol. 20, no. 1, pp. 383–398, Jan. 2019.
- [3] Y. Lv, Y. Duan, W. Kang, Z. Li, and F.-Y. Wang, "Traffic flow prediction with Big Data: A deep learning approach," *IEEE Trans. Intell. Transp. Syst.*, vol. 16, no. 2, pp. 865–873, Apr. 2015.
- [4] S. K. Bhoi, P. K. Sahu, M. Singh, P. M. Khilar, R. R. Sahoo, and R. R. Swain, "Local traffic aware unicast routing scheme for connected car system," *IEEE Trans. Intell. Transp. Syst.*, vol. 21, no. 6, pp. 2360–2375, Jun. 2020.
- [5] S. Çolak, A. Lima, and M. C. González, "Understanding congested travel in urban areas," *Nature Commun.*, vol. 7, no. 1, pp. 1–8, 2016.
- [6] Z. Li, I. Kolmanovsky, E. Atkins, J. Lu, D. P. Filev, and J. Michelini, "Road risk modeling and cloud-aided safety-based route planning," *IEEE Trans. Cybern.*, vol. 46, no. 11, pp. 2473–2483, Nov. 2016.
- [7] C. Li, W. Yue, G. Mao, and Z. Xu, "Congestion propagation based bottleneck identification in urban road networks," *IEEE Trans. Veh. Technol.*, vol. 69, no. 5, pp. 4827–4841, May 2020.
- [8] M. Chen, X. Yu, and Y. Liu, "PCNN: Deep convolutional networks for short-term traffic congestion prediction," *IEEE Trans. Intell. Transp. Syst.*, vol. 19, no. 11, pp. 3550–3559, Nov. 2018.
- [9] Y. Yang, Y. Xu, J. Han, E. Wang, W. Chen, and L. Yue, "Efficient traffic congestion estimation using multiple spatio-temporal properties," *Neurocomputing*, vol. 267, pp. 344–353, 2017.
- [10] B. M. Williams and L. A. Hoel, "Modeling and forecasting vehicular traffic flow as a seasonal ARIMA process: Theoretical basis and empirical results," *J. Transp. Eng.*, vol. 129, no. 6, pp. 664–672, 2003.
- [11] S. V. Kumar and L. Vanajakshi, "Short-term traffic flow prediction using seasonal ARIMA model with limited input data," *Eur. Transport Res. Rev.*, vol. 7, no. 3, 2015, Art. no. 21.
- [12] G. P. Zhang, "Time series forecasting using a hybrid ARIMA and neural network model," *Neurocomputing*, vol. 50, pp. 159–175, 2003.
- [13] X. Ma, Z. Tao, Y. Wang, H. Yu, and Y. Wang, "Long short-term memory neural network for traffic speed prediction using remote microwave sensor data," *Transp. Part C: Emerg. Technol.*, vol. 54, pp. 187–197, 2015.
- [14] B. Yang, S. Sun, J. Li, X. Lin, and Y. Tian, "Traffic flow prediction using LSTM with feature enhancement," *Neurocomputing*, vol. 332, pp. 320–327, 2019.
- [15] Z. Wang, X. Su, and Z. Ding, "Long-term traffic prediction based on LSTM encoder-decoder architecture," *IEEE Trans. Intell. Transp. Syst.*, vol. 22, no. 10, pp. 6561–6571, Oct. 2021.
- [16] F. Zhao, G.-Q. Zeng, and K.-D. Lu, "EnLSTM-WPEO: Short-term traffic flow prediction by ensemble LSTM, NNCT weight integration, and population extremal optimization," *IEEE Trans. Veh. Technol.*, vol. 69, no. 1, pp. 101–113, Jan. 2020.
- [17] L. Zhao et al., "T-GCN: A temporal graph convolutional network for traffic prediction," *IEEE Trans. Intell. Transp. Syst.*, vol. 21, no. 9, pp. 3848–3858, Sep. 2020.
- [18] S. Guo, Y. Lin, N. Feng, C. Song, and H. Wan, "Attention based spatial-temporal graph convolutional networks for traffic flow forecasting," in *Proc. 33rd AAAI Conf. Artif. Intell.*, 2019, pp. 922–929.
- [19] C. Chen et al., "Gated residual recurrent graph neural networks for traffic prediction," in *Proc. 33rd AAAI Conf. Artif. Intell.*, 2019, pp. 485–492.
- [20] Y. Liang, S. Ke, J. Zhang, X. Yi, and Y. Zheng, "GeoMAN: Multi-level attention networks for geo-sensory time series prediction," in *Proc. 27th Int. Joint Conf. Artif. Intell.*, 2018, pp. 3428–3434.
- [21] X. Wang et al., "Traffic flow prediction via spatial temporal graph neural network," in *Proc. Web Conf.*, 2020, pp. 1082–1092.
- [22] R. Dai, S. Xu, Q. Gu, C. Ji, and K. Liu, "Hybrid spatio-temporal graph convolutional network: Improving traffic prediction with navigation data," in *Proc. 26th ACM SIGKDD Int. Conf. Knowl. Discov. Data Mining*, 2020, pp. 3074–3082.
- [23] Z. Lv, J. Xu, K. Zheng, H. Yin, P. Zhao, and X. Zhou, "LC-RNN: A deep learning model for traffic speed prediction," in *Proc. 27th Int. Joint Conf. Artif. Intell.*, 2018, pp. 3470–3476.
- [24] Z. Pan, Y. Liang, W. Wang, Y. Yu, Y. Zheng, and J. Zhang, "Urban traffic prediction from spatio-temporal data using deep meta learning," in *Proc. 25th ACM SIGKDD Int. Conf. Knowl. Discov. Data Mining*, 2019, pp. 1720–1730.
- [25] Q. Huang, Y. Yang, Y. Xu, F. Yang, Z. Yuan, and Y. Sun, "Citywide road-network traffic monitoring using large-scale mobile signaling data," *Neurocomputing*, vol. 444, pp. 136–146, 2021.
- [26] Q. Shi et al., "Block hankel tensor ARIMA for multiple short time series forecasting," in *Proc. AAAI Conf. Artif. Intell.*, 2020, pp. 5758–5766.
- [27] J. Zhang, Y. Zheng, D. Qi, R. Li, X. Yi, and T. Li, "Predicting citywide crowd flows using deep spatio-temporal residual networks," *Artif. Intell.*, vol. 259, pp. 147–166, 2018.
- [28] T. Chen and C. Guestrin, "XGBoost: A scalable tree boosting system," in *Proc. 22nd ACM SIGKDD Int. Conf. Knowl. Discov. Data Mining*, 2016, pp. 785–794.
- [29] J. Zhang, Y. Zheng, and D. Qi, "Deep spatio-temporal residual networks for citywide crowd flows prediction," 2016, *arXiv:1610.00081*.
- [30] X. Ma, Z. Dai, Z. He, J. Ma, Y. Wang, and Y. Wang, "Learning traffic as images: A deep convolutional neural network for large-scale transportation network speed prediction," *Sensors*, vol. 17, no. 4, 2017, Art. no. 818.
- [31] Y. Zhang, Y. Zhou, H. Lu, and H. Fujita, "Traffic network flow prediction using parallel training for deep convolutional neural networks on spark cloud," *IEEE Trans. Ind. Inform.*, vol. 16, no. 12, pp. 7369–7380, Dec. 2020.
- [32] C. Chen, K. Li, S. G. Teo, X. Zou, K. Li, and Z. Zeng, "Citywide traffic flow prediction based on multiple gated spatio-temporal convolutional neural networks," *ACM Trans. Knowl. Discov. Data*, vol. 14, no. 4, pp. 1–23, 2020.
- [33] Y. Hechtlinger, P. Chakravarti, and J. Qin, "A generalization of convolutional neural networks to graph-structured data," 2017, *arXiv:1704.08165*.
- [34] B. Yu, H. Yin, and Z. Zhu, "Spatio-temporal graph convolutional networks: A deep learning framework for traffic forecasting," in *Proc. 27th Int. Joint Conf. Artif. Intell.*, 2018, pp. 3634–3640.
- [35] C. Zheng, X. Fan, C. Wang, and J. Qi, "GMAN: A graph multi-attention network for traffic prediction," in *Proc. AAAI Conf. Artif. Intell.*, 2020, pp. 1234–1241.
- [36] A. Grover and J. Leskovec, "node2vec: Scalable feature learning for networks," in *Proc. 22nd ACM SIGKDD Int. Conf. Knowl. Discov. Data Mining*, 2016, pp. 855–864.
- [37] X. Rong, "word2vec parameter learning explained," 2014, *arXiv:1411.2738*.
- [38] A. Ng et al., "Sparse autoencoder," *CS294A Lecture notes*, vol. 72, no. 2011, pp. 1–19, 2011.
- [39] A. Koeswadiy, R. Soua, and F. Karay, "Improving traffic flow prediction with weather information in connected cars: A deep learning approach," *IEEE Trans. Veh. Technol.*, vol. 65, no. 12, pp. 9508–9517, Dec. 2016.
- [40] Q. Huang, Y. Yang, Z. Yuan, H. Jia, L. Huang, and Z. Du, "The temporal geographically-explicit network of public transport in changchun city, northeast China," *Scientific Data*, vol. 6, 2019, Art. no. 190026.
- [41] C.-H. Wu, J.-M. Ho, and D.-T. Lee, "Travel-time prediction with support vector regression," *IEEE Trans. Intell. Transp. Syst.*, vol. 5, no. 4, pp. 276–281, Dec. 2004.
- [42] S. Fang, Q. Zhang, G. Meng, S. Xiang, and C. Pan, "GSTNet: Global spatial-temporal network for traffic flow prediction," in *Proc. 28th Int. Joint Conf. Artif. Intell.*, 2019, pp. 2286–2293.



QiuYang Huang received the Ph.D. degree in computer science and technology with Jilin University, Changchun, China, in 2021, and the B.E. and M.E. degrees in software engineering from Jilin University, in 2013 and 2016, respectively. He is currently a Postdoc with the college of transportation, Jilin University. He is also a Visiting Scholar with the School of Electrical and Electronic Engineering, Nanyang Technological University, Singapore. His research interests include applications of data mining, urban computing, and intelligent transportation systems.



Hongfei Jia is currently a Professor and the Ph.D. supervisor with the college of transportation, Jilin University, Changchun, China. He has hosted two projects supported by the National Natural Science Foundation, one project by the National 863 plan, one subproject by the National Key Technology R&D Program, one subproject by the State Key Program of the National Natural Science Foundation, and one project by the Foreign Exchange and Cooperation Program of the National Natural Science Foundation. He is a Member of the National Standardization Technical Committee of Comprehensive Transportation, Road Transportation and Engineering Teaching Steering Committee of higher Education institutions of the Ministry of Education. Besides, he is a Senior Member of China Highway and Transportation Society and Standing Committee Member of the Automatic Driving Working Committee of China Highway and Transportation Society.



Yuanbo Xu received the B.E., M.E., and Ph.D. degrees in computer science and technology from Jilin University, Changchun, China, in 2012, 2015, and 2019, respectively. He is currently an Associate Professor with the College of Computer Science and Technology Jilin University, Changchun. He is also a Visiting Scholar with Rutgers, New Brunswick, NJ, USA. His research interests include applications of data mining, recommender system, and mobile computing. He has authored or coauthored some research results on journals, such as TKDE, TMM, TNNLS and conference as INFOCOM, ICDE, ICDM, IWQoS.



Yongjian Yang is currently a Professor and the Ph.D. supervisor with Jilin University, Changchun, China, Vice Dean of Software College of Jilin University, Changchun, China, Director of Key lab under the Ministry of Information Industry, Standing Director of Communication Academy, a Member of the Computer Science Academy of Jilin Province. His research interests include network intelligence management, wireless mobile communication and services, research and exploitation for next generation services foundation, and key productions on wireless mobile communication. He participated three projects of NSFC, 863 and funded by National Education Ministry for Doctoral Base Foundation. He has authored 12 projects of NSFC, key projects of Ministry of Information Industry, Middle and Young Science and Technology Developing Funds, Jilin provincial programs, Shenzhen, Zhuhai, and Changchun.



Gaoxi Xiao (Senior Member, IEEE) received the B.S. and M.S. degrees in applied mathematics from Xidian University, Xian, China, in 1991 and 1994, respectively, and the Ph.D. degree in computing from The Hong Kong Polytechnic University, Hong Kong, in 1998. From 1994 to 1995, he was an Assistant Lecturer with Xidian University. In 1999, he was a Postdoctoral Research Fellow with Polytechnic University, Brooklyn, NY, USA, and from 1999 and 2001, a Visiting Scientist with the University of Texas at Dallas, Richardson, TX, USA. In 2001, he joined the School of Electrical and Electronic Engineering, Nanyang Technological University, Singapore, where he is currently an Associate Professor. His research interests include complex systems and complex networks, communication networks, smart grids, system resilience, and risk management. Dr. Xiao is/was an Associate Editor or the Guest Editor for the IEEE TRANSACTIONS ON NETWORK SCIENCE AND ENGINEERING, *PLOS One*, and Advances in Complex Systems, and a Technical Program Committee Member for numerous conferences, including the IEEE International Conference on Communications and the IEEE Global Communications Conference.