IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received August 17, 2019, accepted August 29, 2019, date of publication September 6, 2019, date of current version September 20, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2939902

Prediction of City-Scale Dynamic Taxi Origin-Destination Flows Using a Hybrid Deep Neural Network Combined With Travel Time

ZONGTAO DUAN[1], KAI ZHANG[2], ZHE CHEN[1], ZHIYUAN LIU[2], LEI TANG[1], YUN YANG[1], AND YUANYUAN NI[1]

[1]School of Information Engineering, Chang'an University, Xi'an 710064, China
[2]Jiangsu Key Laboratory of Urban ITS, Jiangsu Province Collaborative Innovation Center of Modern Urban Traffic Technologies, School of Transportation, Southeast University, Nanjing 210018, China

Corresponding author: Zhe Chen (zchen@chd.edu.cn)

This work was supported in part by the Funds for Key Scientific and Technological Innovation Team of the Shaanxi Province, China, under Grant 2017KCT-29, in part by the Funds for Key Research and Development Plan Project of the Shaanxi Province, China, under Grant 2017GY-072 and Grant 2018GY-032, and in part by the Fundamental Research Funds for the Central Universities under Grant 300102249401.

ABSTRACT Predicting city-scale taxi origin-destination (OD) flows takes an important role in understanding passengers' travel demands as well as managing taxi operation and scheduling. But the complex spatial dependencies and temporal dynamics make this problem challenging. In this paper, a hybrid deep neural network prediction model based on convolutional LSTM (ConvLSTM) is proposed. For improving the prediction accuracy, the implicit correlation between travel time and OD flow is explored and they are combined as inputs of the prediction model. Moreover, in order to realize OD flows prediction at the road network level, and solve the problem that grid-based representation method cannot distinguish traffic flow at different heights, such as in multi-layer overpass areas, this paper presents a grid and road nested method to represent ODs. With the time of day partitioned into time slots, OD flows are extracted and predicted in both spatial and temporal domain. In the experiment, real taxi data are used to verify the proposed model and prediction method fully. And the experimental results show that the proposed model can effectively predict city-wide taxi OD flow, and outperforms the typical time sequence models and existing deep neural network models.

INDEX TERMS Prediction of dynamic origin-destination flows, hybrid deep neural network, demands for taxis, GPS trajectory of taxis.

I. INTRODUCTION

A. BACKGROUND

The purpose of predicting taxi origin-destination (OD) flows is to obtain the travel demand between ODs within a certain period. City-wide taxi OD flows can reflect the spatial-temporal distribution of taxi passengers' departure and arrival flows, and accurate prediction of it is an indispensable prerequisite for traffic modeling and planning.

Generally, the OD estimation can be divided into static and dynamic OD estimations based on the assumption of static or dynamic changes of traffic flows on the transportation links. Previously, due to the limited means of observation, most studies focused on static OD flow estimation, and it was usually assumed that the traffic flow on the road sections was stable during a relatively long observation period (e.g., 1 week, 1 month, or 1 year) [1]. For static OD flow estimation, various optimization and statistical models have been proposed, e.g., maximum entropy (information minimization) [2], [3], maximum likelihood [4], Bayesian reasoning [5], generalized least squares for networks without congestion [6], and bi-level programming for congested networks [7]. Since these methods are based on the assumption that the traffic flow is stable, they cannot be used for short-term prediction or circumstances

The associate editor coordinating the review of this manuscript and approving it for publication was Orazio Gambino.

with unstable traffic flow. In the actual urban traffic, affected by various uncertain factors, the traffic flow usually fluctuates in a short time, which does not fully satisfy the static hypothesis. Therefore, the research focus of OD flow estimation has been shifted to dynamic OD flow estimation.

The emergence and popularity of new sensor devices, especially positioning devices such as the GPS, have made dynamic OD flow estimation and prediction possible. Call dial record (CDR) data, positioning data on users' mobile devices, and GPS data on vehicles are used for estimation and prediction of OD flows [8]–[10]. The data cover a large spatial-temporal scope, involves a large amounts of samples, and can accurately obtain the user's location with a high sampling rate, so the departure time, origin, destination, arrival time, trip time, route trajectory, and even travel purposes of individuals can be relatively precisely recorded [11], which is very helpful for fine-grained estimation of OD flows in the spatial-temporal domain. Current studies on taxi OD flow prediction based on taxi trajectory generally use the method of traffic analysis zone (TAZ), which divides the urban area into several grids spatially, regards each grid as a region, and counts the number of trips from one region to another. However, this method fails to make full use of the accurate positioning ability of the GPS device, and it is still a coarse-grained statistical analysis method from the spatial perspective. To fully utilize the fine-grained characteristics of taxi trajectory in the spatial-temporal domain, we propose a grid and road nested OD representation method, and utilizes ConvLSTM, Conv2DTranpose, and SeparableConv2D to construct a hybrid deep neural network prediction model, which can be applied to explore the correlation between OD flow and travel time, and estimate the taxi OD flows on the city scale. To the best of our knowledge, there has been no study focusing on combining OD flows with travel time and performing fine-grained OD flow prediction in the spatial-temporal domain.

B. LITERATURE REVIEW

Traffic prediction has been extensively studied for many decades. However, due to the complex spatial dependencies and nonlinear temporal dynamics, it is still a challenging problem [12]. The research process has generally gone through two stages of development, i.e., the mathematical model-driven stage and the data-driven stage. The models can be roughly divided into parameterized models, non-parameterized models, and hybrid integration models [13].

Prediction of OD flow in the earlier stage was mainly driven by mathematical models, and the main body of the prediction model was often a parametric model. The typical models are statistical analysis models based on time series, e.g., the autoregressive integrated moving average (ARIMA) model and its variants [14]. The ARIMA model considers the intrinsic relevance of traffic variables over time and applies auto-regression for prediction [15]. For its simplicity in structure and only driven by endogenous variables, there is no need to consider other exogenous variables, but the time series or its difference should be stable. As a linear relationship model in nature, the ARIMA model is not suitable for describing the relationship between dynamic changes and nonlinear variables. In addition, Kalman filtering and its extension were also models commonly used in the prediction of traffic variables [16]–[18]. Kalman filtering was mainly constrained by the hypothesis that the noise followed a Gaussian distribution, and for the extended Kalman filtering, errors could easily occur during the linearization process, which could affect the accuracy of the solution. In terms of OD flow prediction, Ashok et al. used Kalman filtering for time-dependent OD flow prediction based on traffic flow and average velocity data observed on the road [19]. Based on bus card swiping data, Chen et al. used Kalman filtering to predict the short-term OD flow of bus lines [20]. The main drawback of the time series-based methods was that the prediction of traffic variables failed to take into account the spatial correlation of the road network; thus, some studies have been carried out to explore the influences of the spatial factors of the road network on the prediction performance [21].

Non-parameterized models mainly include k-nearest neighbor (KNN), support vector regression (SVR), support vector machine (SVM), and artificial neural networks [22]. Artificial neural networks (ANN), especially deep neural networks that have been rapidly developed in recent years, e.g., CNN and RNN, can describe complex nonlinear relationships between variables. Ren et al. developed an effective approach for non-linear and nonperiodic traffic speed prediction on urban road networks based on convolutional neural networks, achieved an average improvement of prediction accuracy from 0.9% to 3.3% for highly fluctuating road segments and up to 23.8% improvement in accuracy for individual road segment [23]. Lv et al. constructed a stacked autoencoder (SAE) deep network model for traffic flow prediction of local highway networks, indicating that deep networks could better capture the spatial-temporal correlation characteristics of data compared with the traditional shallow models such as the machine learning algorithms back propagation neural networks and SVM, and the prediction accuracy could be significantly improved [24]. Since CNN can extract the correlation of spatial features and RNN can extract the correlation of temporal features, the integration of the two has the ability to extract spatial-temporal features. Hence, it is widely applied for the spatial-temporal analysis of traffic flow. Yu et al. constructed spatiotemporal recurrent convolutional networks (SRCNs) to convert the running speeds of vehicles in urban road networks into images, predicted the vehicle running speed using image sequence analysis, and analyzed its superior performance in models like SVM and SAE [25]. Duan et al. used the hybrid deep neural network composed of CNN and long short-term memory (LSTM) to analyze the GPS trajectory data of taxis and forecast the citywide traffic flow. Considering the spatial-temporal characteristics of traffic flow changes, they extended the traditional traffic flow prediction

from road sections to the entire urban road network [26]. Moreover, De Brébisson et al. applied RNN in taxi destination prediction, predicting taxi destination based on the GPS trajectory of the taxi after it had just departed, and the results showed the excellent prediction performance of RNN [27]. Using deep learning technology to predict the evolution of congestion in large-scale transportation networks, Ma et al. proposed a recurrent neural network-restricted Boltzmann machine deep prediction model that applied GPS trajectory data to predict the formation, spread, and dissipation of traffic jams [28]. The prediction accuracy rate reached 88%, and its performances exceeded than that of BP NN and SVM by over 20%. Zheng et al. adopted deep residual networks to predict the flow of people in urban areas, achieving a better performance than that of prediction methods such as ARIMA and vector auto-regression (VAR) [29]. In order to take more exogenous dependencies into consideration, fusion deep network [30] and deep multi-view spatial-temporal network (DMVST-Net) [31] were proposed for predicting people's demands for online taxi-hailing services, which outperformed ARIMA and multiple layer perceptron. Li et al. proposed a diffusion convolutional recurrent neural network for the traffic flow prediction, and achieved a 12%-15% improvement on the prediction accuracy [32]. The above applications show that deep neural networks provides an effective new way for traffic prediction.

Deep learning technology was first used for network OD prediction in 2008, when Qian et al. introduced RNN technology to analyze wireless communication network traffic [33]. RNN performed better than the gravity model in predicting OD flows, showing the great potential of RNN in network analysis. Few studies have been carried out on traffic OD flow prediction in the current stage, and different traffic data were used in these studies. Toqué et al. used LSTM to analyze the swiping data of metro passengers to predict dynamic OD flows between metro network sites, the performance of which was improved compared with the traditional VAR and Calendar methods [34]. However, Toqué et al.'s method cannot be fully applied to the prediction of dynamic OD flows of taxis because the metro lines are relatively fixed, the road network has a limited scale relative to the road traffic network, and the running time between the metro stations is fixed.

In general, with the increasing enrichment of traffic data, non-parameterized prediction models have begun to play a role in nonlinear, static, and dynamic processes as well as spatial-temporal analysis, thus becoming the most effective traffic prediction models in the current phase. Besides, some of these models can also be combined with environmental factors to further improve the accuracy of prediction [35].

C. MOTIVATION

As aforementioned, the deep learning technology has been successfully applied in traditional traffic prediction [24], [34], [36], [37], while the spatial-temporal characteristics of traffic variables can be analyzed by constructing a hybrid deep neural network [15]. In addition, the increasing accumulation of traffic big data, e.g., trajectory data of urban taxis, has enabled the application of deep learning. Inspired by this, we tried to use deep learning technology for short-term prediction of city-scale taxi OD flows to explore the effectiveness of deep learning technology in prediction of OD flows.

As far as we know, there has not been any similar research using deep learning to predict urban taxi OD flows. The research conducted by Toqué *et al.* [34] is closest to our study. They used LSTM and the swiping data of metro passengers to predict the OD flows of the metro network by regarding the OD flows between different metro stations as a time series. However, our hybrid deep neural network not only takes into account the travel frequency but also the travel time of OD flows, exploring the correlation of OD flows from both the spatial and temporal perspectives to improve prediction accuracy.

Current OD flow prediction technologies are unable to accurately obtain the origins and destinations of travelers. For example, when mobile phone data are used for OD estimation, the accuracy of the location depends on the location of the base station [10]. Thus, the data analysis mostly reflects the distribution of OD flows between TAZs. Compared with bus card swiping data and mobile phone data, the trajectory data of taxis are characterized by better positioning performance and higher sampling frequency, based on which the origin and destination of each passenger can be accurately estimated. Usually, each origin and destination can be positioned to a specific point on the road where the taxi is running. Therefore, we considered implementing fine-grained OD flow analysis and integrating the TAZ with the road network so that the estimated OD flows can be refined to roads in TAZs, which is not only more conducive to providing detailed information about taxi demand but also helpful to estimate the traffic flow of roads. Thus, the problem is that the OD matrix will be very large and sparse, which will result in huge consumption of computing resources, so the dimension compression of OD matrices needs to be properly handled.

This study contains the following research objectives:

- To explore the OD flow prediction problem at the road network level, and at the same time, to solve the problem that the grid-based representation method cannot distinguish the traffic flow at different heights in multi-layer overpass areas;
- To design a hybrid deep neural network to explore the spatiotemporal correlation of OD flows;
- To explore the implicit correlation between OD flow and travel time and the influence of OD travel time on the prediction accuracy of OD flow in the deep network model.

And the main contributions of this study are as follows:

1) A fine-grained OD representation method is proposed to refine the expression of OD flow to the road network level. The method enables the traffic flow to have road attributes and overcomes the problem that the grid-based

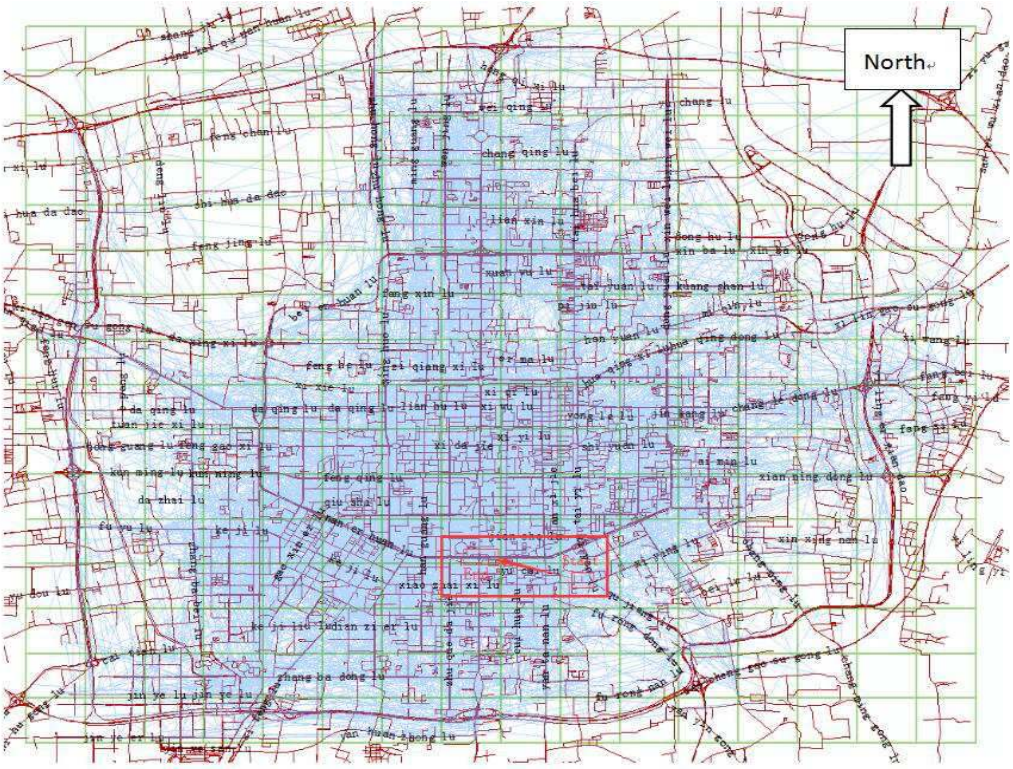


FIGURE 1. Grid division of the urban traffic network.

representation method cannot distinguish the traffic flow on the road at different heights in multi-layer overpass areas.

2) The proposed method can learn the spatial-temporal features of OD flow by using the ConvLSTM model, capture the sparse features of data with conv2DTranpose, and finally, obtain the spatial position relationship between the travel time OD matrix and the OD flow matrix by using SeparableConv2D point-by-point convolution.
3) Additionally, the proposed method makes use of the correlation between the travel time and travel frequency of OD flow, i.e., the travel time matrix and the travel frequency matrix are simultaneously fed into the deep network model to improve the accuracy of OD flow estimation.

D. ORGANIZATION OF THE PAPER

The subsequent sections are organized as follows: In Section II, the problem of fine-grained prediction of OD flows is defined. Section III introduces the method of predicting the OD flows using the hybrid deep neural network. This section first briefly introduces the proposed hybrid deep neural network, then the principle of three basic deep neural network blocks, i.e., ConvLSTM, conv2Dtranspose and SeparableConv2 are introduced in detail, and finally the OD flow prediction method combining travel time is described. In Section IV, a case study is conducted to compare the performance of the proposed method with the currently available OD flow prediction methods, offering test results based on a real dataset of taxis. In Section V, the conclusions from this study and the prospects for future research are presented.

II. PROBLEM DEFINITION AND OD DATA ANALYSIS

A. PROBLEM DEFINITION

In this section, we present the fine-grained OD representation and define the problem of OD flows prediction.

Traditional OD flow prediction is usually TAZ-based. With the improvement and popularization of data sensing technology, the traffic state can be perceived increasingly accurately, and more and more types of information can be perceived, making finer traffic analysis possible. The trajectory data of taxis used in this study contain information such as occupation status (available or unavailable), longitude and latitude of the GPS sampling point, GPS sampling time, and taxi ID. With these information, the pick-up point and drop-off point of each taxi trip can be extracted, thereby obtaining the OD of one trip [38]. Since the OD of trips can be aggregated to a specific road section in the road network, we refine the OD flow of the taxi to the road, which is more conducive to vehicle scheduling and the analysis of road network traffic flow, and it allows the driver to search for passengers on the road rather than in an area.

We first partition the urban road network into grids that are equally spaced horizontally and vertically. Assuming the urban area is divided into N parts in the horizontal direction and M parts in the vertical direction, a $N \times M$ rectangular grids are obtained. As shown in figure 1,

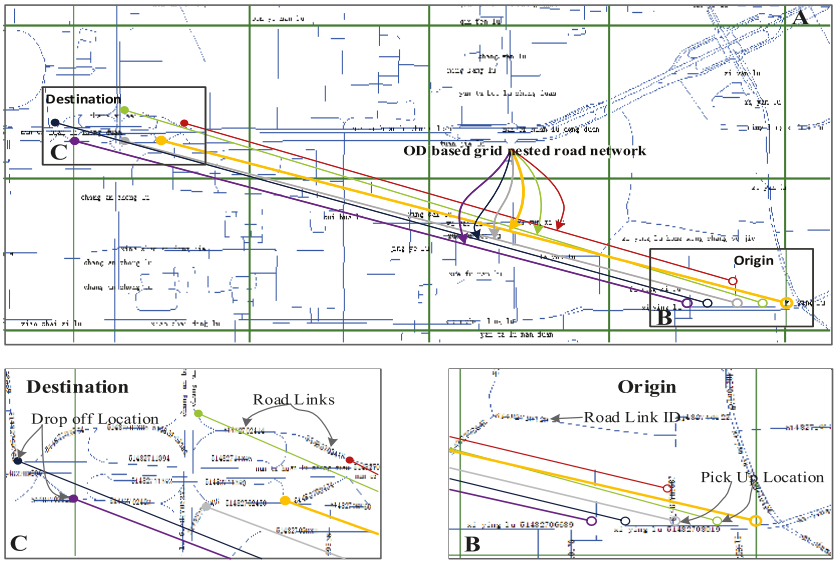


FIGURE 2. Description of OD flow with grid-nested road network. (A) Grid and road network nesting diagram. (B) Origins on road network. (C) Destinations on the road network.

each grid is represented by $Grid(n, m)$, where $(n, m) \in Z^+$(Z^+indicates a positive integer). The latitude range of Xi'an City is (34.181, 34.370), and the longitude range is (108.829, 109.076). In this paper, we divide the urban area into 16 $\times$ 16 grids, and the actual size of each grid is 1.314 km $\times$ 1.718 km.

Figure 2 shows the red area marked in figure 1. The figure gives an example of representing OD with the grid and road network nested method. Figure 2(A) shows that OD flow becomes sparse within a given time interval. When OD flow is represented with the road network, there is only one trip between origin and destination road sections; when the OD flow expression method is based on the traffic community, there are six trips between the origin grid and the destination grid. Figure 2(B) shows the starting point position at the level of the road network. Also, figure 2(C) shows the destination location at the level of the road network, in which each OD in the figure corresponds to the road network in different sections. The location in figure 2(C) is an interchange in Xi'an city, whose function is to connect two main roads. It is composed of a main north-south bridge and four directional ramps, and it is a double-deck long-strip clover leaf interchange. Therefore, compared with the grid-based method, our grid and road network nested method divides urban areas hierarchically, so that traffic flows with road attributes can effectively distinguish the traffic flows on roads at different heights in multi-layer overpass areas; thus, the expression of OD flows can be refined to the level of the road network. In the case that a dataset from a large number of taxis is available, it is feasible to use the grid and road nested network to represent OD flow.

As shown in figure 3, each grid contains a different number of road sections, and a triple can be defined as (n, m, rid), where (n, m) denotes the position coordinates of the grid, and rid is the identification number of a certain road section

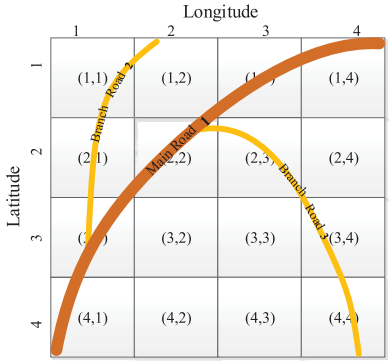


FIGURE 3. Grid and road nested representation.

in the grid $Grid(n, m)$. Here, each small grid has unique coordinates, the roads are highlighted in yellow, and each road section has a unique identification number.

All the road segments in the divided N×M rectangular grids are encoded with a number. For each road segment, a triplet is generated, and the generated triplets are recorded into the triplet set $R = \{(n, m, k)|1 \leqslant n \leqslant N, 1 \leqslant m \leqslant M, k \in Z^+\}$.

Definition 1 (Time Slot): Each 15-minute interval is defined as a time slot, so each day can be divided into 96 (24×4) time slots. The set $S = \{s_j(o_t, d_t)|1 \leqslant j \leqslant 96)\}$ is used to represent all 96 time slots of a day, where s_j is the jth time slot, and o_t and d_t denote the start time and end time of the jth time slot, respectively.

Definition 2 (Trip): The expression $trip(o, d, s_j)$ indicates that the trip starts at time slot j from the origin o to the destination d.For each trip, a trip sign is defined as follows

$$trip\_sign(o, d, s_j, t) = \begin{cases} 1, & if\ o_t \leqslant t < d_t \\ 0, & otherwise \end{cases} \quad (1)$$

where t represents the start time of the trip.

Definition 3 (The Number of Trips Between ODs): The number of trips between ODs is defined as the number of trips from the origin road section to the destination road section in time slot j, which can be computed as follows:

$$trip\_frequency(o, d, s_j) = \sum_{t \in [o_t, d_t]} trip\_sign(o, d, s_j, \Delta t) \tag{2}$$

Definition 4 (OD Matrix): Let the number of elements in the triplet set R be C, corresponding to C road sections. The OD matrix can be expressed as follows:

$$Y^{(j)} = \begin{bmatrix} y_{1,1}^{(j)} & y_{1,2}^{(j)} & \cdots & y_{1,r}^{(j)} & \cdots & y_{l,C}^{(j)} \\ y_{2,1}^{(j)} & y_{2,2}^{(j)} & \cdots & y_{l,C}^{(j)} & \cdots & y_{l,C}^{(j)} \\ \vdots & \vdots & \cdots & \vdots & \cdots & \vdots \\ y_{l,1}^{(j)} & y_{l,2}^{(j)} & \cdots & y_{l,r}^{(j)} & \cdots & y_{l,C}^{(j)} \\ \vdots & \vdots & \cdots & \vdots & \cdots & \vdots \\ y_{C,1}^{(j)} & y_{C,2}^{(j)} & \cdots & y_{C,r}^{(j)} & \cdots & y_{C,C}^{(j)} \end{bmatrix} \tag{3}$$

where $1 \leqslant l \leqslant C$, $1 \leqslant r \leqslant C$, whose size is $C \times C$, represents the OD matrix corresponding to the time slot j, where $y_{l,r}^{(j)}$ denotes the number of trips, with the lth triplet as the origin road section and the rth triple as the destination road section.

The problem of fine-grained OD flow prediction can be defined based on the above definitions.

Problem 1: Given the historical OD flows observation matrices $\left\{ \boldsymbol{Y}^{(i)} | i = 0, 1, \cdots, t-1 \right\}$, the OD matrix of the tth time slot, $\boldsymbol{Y}^{(t)}$, need to be predicted.

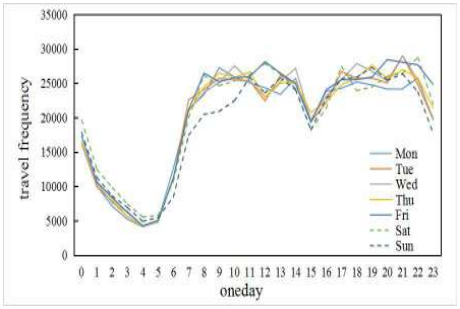
The origin or destination in the OD matrices represents a road in a certain grid. The essence of this representation is to characterize OD flows from one road section to another. Such fine-grained representation allows for a more accurate representation of passengers' travel demands. Moreover, for relatively long road section crossing multiple grids, grid constraint can avoid the ambiguity in representing flows on the same road section.

B. OD DATA'S FEATURE ANALYSIS

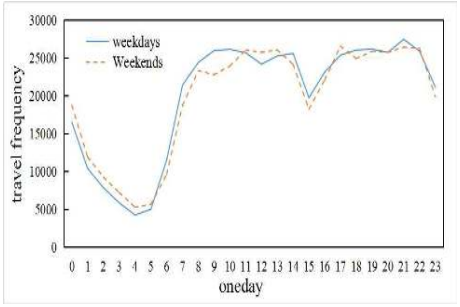
In this section, we analyze the correlation between travel frequency and travel time by comparing travel frequency and travel time in a different time slot on weekdays and weekends.

1) THE FREQUENCY OF TRIPS

Here we compare OD travel frequency on weekdays and weekends. As shown in figure 4(a), the travel frequency of all taxis in the Third Ring of Xi'an on a working day is shown by the solid line for weekdays and the dotted line for weekends. By comparing the travel frequency of passengers on weekdays and weekends (every 15 min), we can find that the travel frequency of passengers on weekends lags behind that on weekdays. Especially during the period from 6:00 to 11:30, passengers travel more frequently on weekdays



(a) One week



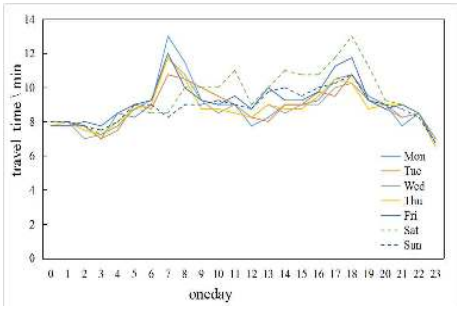
(b) Weekdays and weekends

FIGURE 4. Travel frequency.

than on weekends. A possible reason is that most people take a taxi to work place in the morning on weekdays, so there are more vehicles on a limited road and the speed is relatively slower. Citizens have a fixed lifestyle on working days, such as workers commuting between their residential area and work area, and students traveling between home and school. Therefore, the demand for taxis is relatively fixed. In addition, in figure 4(b), the solid blue line is the average travel frequency for five days in the week, and the orange dotted line is the average travel frequency for two days on the weekend. It is obvious that during the period from 6:00 to 7:00 and from 9:00 to 11:00, the frequency of travel on weekdays is more than that on weekends, and from 11:00 to 13:00, the frequency of travel on weekends is more than that on weekdays. In addition, during the periods from 17:00 to 18:00, 19:30 to 20:00, and 22:30 to 23:30, the travel frequency on weekends is more than that on weekdays. According to Wang's study [39], residents have no fixed lifestyle on weekends, and most of them choose to stay at home in the morning and go out for activities after 11:00.

2) THE DURATION OF TRIPS

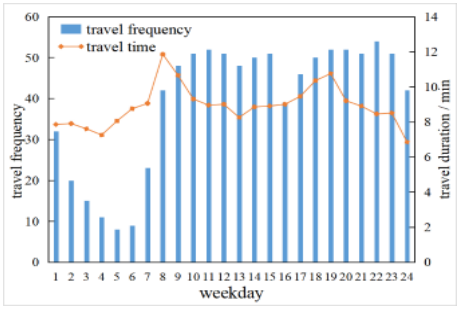
In this section, we compare the OD travel time starting at different time slots on weekdays and weekends. As shown in figure 5(a), the duration of taxi trips on weekdays is indicated by the solid line, and that on weekends is indicated by the dotted line. During the period from 6:30 to 8:30, the travel time on weekdays is longer than that on weekends. Residents have a fixed lifestyle on working days. For example, on working days, people mainly commute between residential areas and work areas, especially during morning and



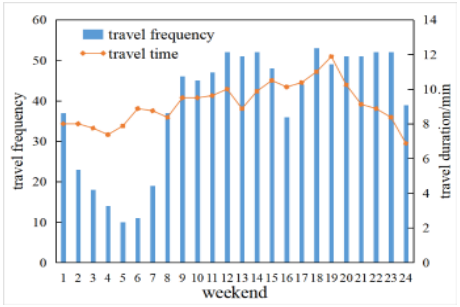
(a) A week

(b) Weekdays and weekends

FIGURE 5. The duration of trips.



(a) Weekdays



(b) Weekends

FIGURE 6. The analysis of the relationship between travel time and travel frequency.

evening rush hours, which is likely to cause road congestion. During the period from 8:30 to 22:00, the average travel time on weekends is relatively long, probably because residents do not have a fixed lifestyle on weekends. For example, most residents prefer to take a taxi to go shopping or eat out after 11:00 on weekends, causing road congestion. To analyze the travel difference of residents on weekdays and weekends, we calculate the average travel time of each time slot on weekdays and weekends. As shown in figure 5(b), we can see that the average travel time of passengers across five days of a week shows a double-peak feature, while the average travel time of passengers across two days of a weekend shows a single-peak feature.

3) THE RELATIONSHIP BETWEEN TRAVEL TIME AND TRAVEL FREQUENCY

Generally speaking, when the departure place and the destination are determined, the more the travel frequency and vehicle density on the road increase, and the lower the vehicle traffic speed becomes. Correspondingly, travel time will also be lengthened. The Pearson correlation coefficient is a pairwise comparison. Therefore, the travel time corresponding to the OD stream with the period N during weekdays and weekends is denoted as x, and the travel frequency is denoted as y. The calculating formula for the Pearson correlation coefficient is

$$r = \frac{N\sum x_i y_i - \sum x_i \sum y_i}{\sqrt{N\sum x_i^2 - (\sum x_i)^2}\sqrt{N\sum y_i^2 - (\sum y_i)^2}} \tag{4}$$

The closer the Pearson correlation coefficient approaches to 1, the stronger the correlation is. Conversely, the closer the correlation coefficient approaches to 0, the weaker the correlation would be. As shown in Table 1, the greater the absolute value of the Pearson correlation coefficient, the stronger the correlation. In figure 6(a), the Pearson correlation coefficient of travel time and travel frequency on weekdays calculated in this paper is 0.4447. Figure 6(b) shows that the Pearson correlation coefficient of the two on weekends is 0.584. By looking up the Pearson correlation coefficient in Table 1, we find that the travel time and the travel frequency are moderately correlated.

III. METHODOLOGY

A. OVERVIEW OF THE METHODOLOGY

As shown in figure 7, the proposed OD flow prediction modeling process can be roughly divided into four parts, namely, data preprocessing, OD information extraction, data organization, and prediction modeling. The purpose of data preprocessing is to remove invalid data and correct GPS data errors through map matching. Another major task of preprocessing is to conduct grid division of urban areas and establish the expression of the grid-nested road network of track points. The OD information extraction includes extracting the number of trips and the average trip time between the ODs from taxis' trajectory data. The extracted OD demands and trip times are organized into matrices. For the convenience of subsequent calculation, the OD matrix and the trip time matrix are stored in the h5py library according to time slots and then compressed. Prediction modeling is the core part of this study, which will construct a hybrid deep neural network based on ConvLSTM.

TABLE 1. Pearson correlation coefficient.

Pearson correlation coefficient	0.0−0.2	0.2−0.4	0.4−0.6	0.6−0.8	0.8−1.0
Correlation degree	very weak or no	weak	moderate	strong	highly strong

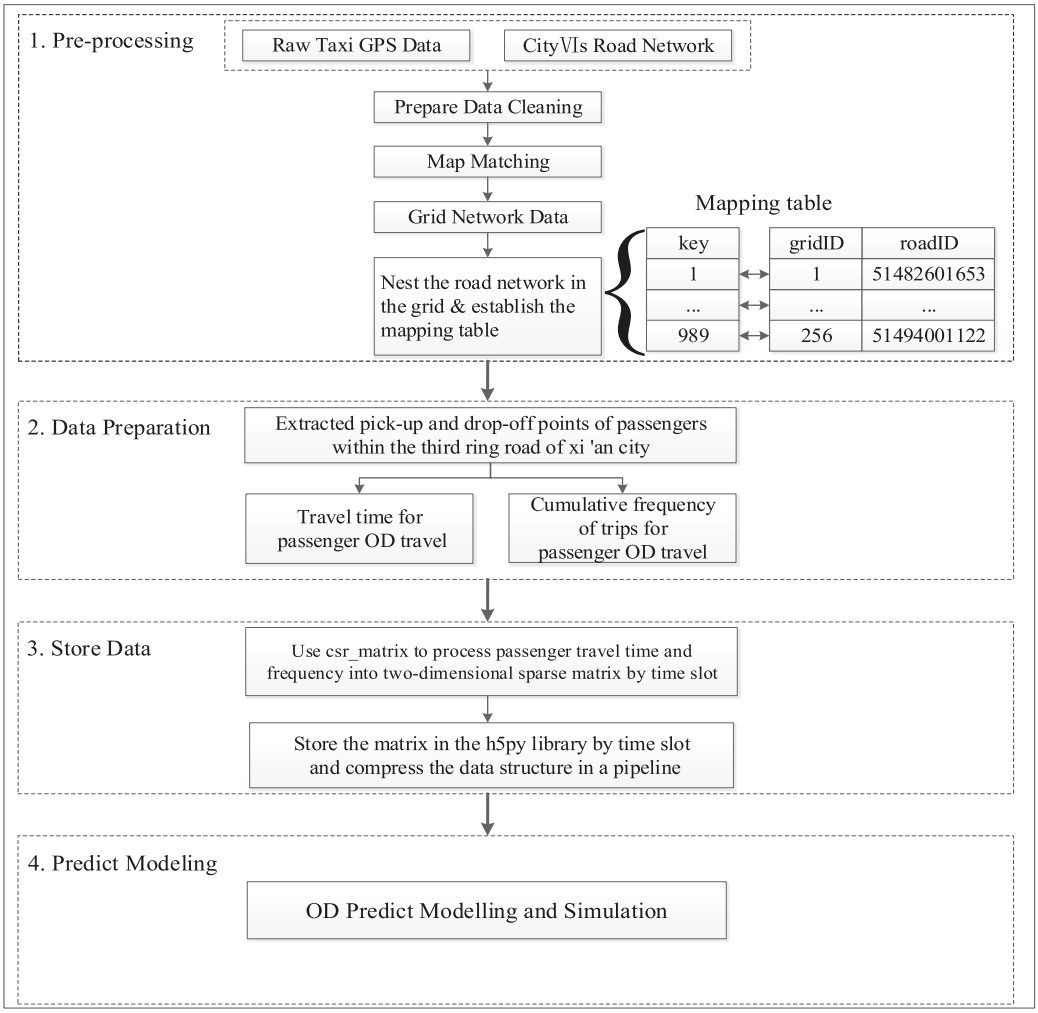


FIGURE 7. Overview of the methodology.

B. DEEP NEURAL NETWORKS

1) CONVLSTM

The ConvLSTM model, which is the combination of CNN and LSTM, was first proposed by Shi et al. for precipitation nowcasting [40]. Each input feature of a ConvLSTM network is a three-dimensional spatiotemporal tensor, where the first two dimensions are the spatial dimensions. The structure diagram of ConvLSTM is shown in figure 8. In its network structure, the output values of each gate, and the state value of the memory unit can be represented by the following group of formulas:

$$\begin{aligned} i_t &= \sigma\left(W_{xi} * x_t + W_{hi} * h_{t-1} + W_{ci} \bigodot c_{t-1} + b_i\right) \\ f_t &= \sigma\left(W_{xf} * x_t + W_{hf} * h_{t-1} + W_{cf} \bigodot c_{t-1} + b_f\right) \\ o_t &= \sigma\left(W_{xo} * x_t + W_{ho} * h_{t-1} + W_{co} \bigodot c_t + b_o\right) \\ h_t &= o_t \bigodot tanh(c_t) \end{aligned} \tag{5}$$

where i_t, f_t and o_t denote the output of the input gate, the forgotten gate, and the output gate, respectively; c_t is the output of the memory units; h_t indicates the current state of LSTM network; W_{xi}, W_{xf}, W_{xo}, W_{xc} is the weight matrix connecting x_t to the three gate units and the memory units; W_{hi}, W_{hf}, W_{ho}, W_{hc} is the weight matrix connecting h_{t-1} to the three gate units and the memory units; W_{ci}, W_{cf} is the weight matrix connecting C_{t-1} to the three gate units and the memory units; b_i, b_f, b_o, b_c stands for the bias of the three gate units and the memory units; $\bigodot$ represents the Hadamard product; $*$ is the convolution operation; and σ and $tanh\left(\cdot\right)$ are nonlinear activation functions, which are expressed as follows:

$$\sigma = \frac{1}{1 + e^{x}} \tag{6}$$

$$tanh\left(x\right) = \frac{e^{x} - e^{-x}}{e^{x} + e^{-x}} \tag{7}$$

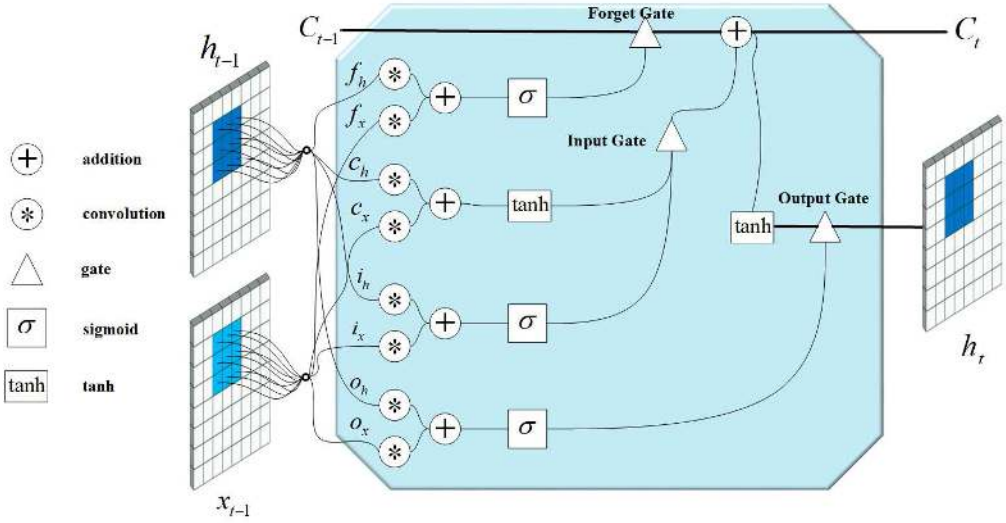


FIGURE 8. The model diagram of ConvLSTM.

Liu et al. and Yuan et al. have proved that ConvLSTM has good properties for traffic flow prediction [41], [42] and demonstrated that the local spatial characteristics of traffic flow could be captured with the convolutional part of ConvLSTM. In addition, the LSTM portion can capture the temporal correlation of traffic flow. Therefore, a ConvLSTM layer is used in this paper to learn the spatial and temporal characteristics of OD flow.

In this paper, the input of ConvLSTM layer is a matrix as shown in equation (3), which consists of a series of vectors $\boldsymbol{Y}^j = \left(\boldsymbol{Y}_1^j, \boldsymbol{Y}_2^j, \cdots, \boldsymbol{Y}_C^j\right)$. Each element in the vector represents the OD flow between origin and destination, i.e. $\boldsymbol{Y}_i^j = \left(y_{i,1}^j, y_{i,2}^j, \cdots, y_{i,C}^j\right)$. A one-dimensional convolution and pooling processes are performed on each $\boldsymbol{Y}_i^j$ and thus the spatial characteristics between near spots are extracted. After then, these results are used as the input to the LSTM layer, finally, the temporal characteristics are also extracted.

2) CONV2DTRANSPOSE

Conv2DTranspose [43] is widely used in image processing field. In the Conv2DTranspose network layer, the spatial vector y^j consisting of N images $y_1^i, \cdots, y_N^i$ is used as the input. Each channel of the image $y_{n,c}^i$ is represented as a linear sum of K latent feature maps z_k^i convolved with filters $f_{k,c}$.

$$\sum_{k=1}^{K} z_k^i * f_{k,c} = y_{n,c}^i \tag{8}$$

where $y_{n,c}^i$ is an $N_r \times N_c$ image and the filters are $H \times H$, then the latent feature maps are $(N_r + H - 1) \times (N_c + H - 1)$ in size. Since equation (8) is a non-deterministic system and cannot obtain a unique solution, a regular term needs to be introduced to capture the sparsity in the feature, and the cost function is shown in equation (9).

$$C_1\left(y^i\right) = \frac{\lambda}{2} \sum_{c=1}^{N} || \sum_{k=1}^{K} z_k^i \bigoplus f_{k,c} - y_c^i ||_2^2 + \sum_{k=1}^{K} ||z_k^i||_p \tag{9}$$

where λ is a constant used to balance the reconstructed relative error of y_i and the sparsity of the feature map z_k^i. It is worth mentioning that the sparse norm is the p-norm of the vectorization version of the matrix. Some sparse regularization norms $||Z_k^i||_p$ on the reconstructed item can capture the sparse features of the features. In this paper, the Conv2DTranspose layer is used to capture the sparse features of OD flow.

3) SEPARABLECONV2D

SeparableConv2D consists of two smaller convolutional filters: depth convolution and point-by-point convolution. As shown in figure 9, the process of convolution in SeparableConv2D can be illustrated as follows. First, the depth convolution of the 3×3 filter window is used to convolve the vectors of each input channel respectively, and the normalization operation is carried out. Then, the conventional convolution of 1×1 window is used for point-by-point convolution. Finally, the calculated channel vector is projected onto the new channel space through the depth convolution. Verified by Kaiser *et al.* [44], SeparableConv2D is able to learn the spatial correlation of different channels.

In this paper, we use SeparableConv2D to study the relationship between the OD matrix of travel time and the OD matrix of travel frequency. The matrix of our input model is 2×989×989(989×989 is the size of the matrix; 2 is the number of input channels, referring to travel time and travel frequency). Here, there are two 3×3 convolution kernels to

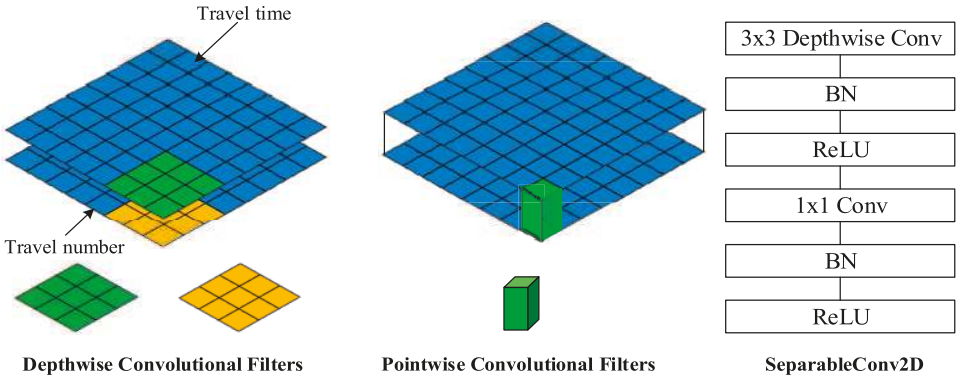


FIGURE 9. The structure diagram of SeparableConv2D.

convolve with the two channels and the size of the output is 2×989×989. Then, the pointwise convolution layer is responsible for combining, that is, to combine the results of the depthwise convolution layers. The above two steps illustrate the relationship between travel time and travel frequency, and the specific formula is as follows:

$$travel\_frequency_{t_n} = \sum_{i=0}^{n-1} \left[\alpha \times travel\_frequency_{t_n}\right] + \sum_{i=0}^{n-1} \left[\beta \times travel\_time_{t_n}\right] \quad (10)$$

where α and β are the learnable parameters of travel time and travel frequency, respectively, during model training.

C. HYBRID DEEP NEURAL NETWORK

In the actual road network, the changes in traffic variables are not only related to time (e.g., tidal phenomena on weekdays) but also to space. Since the roads are connected to each other, the traffic variables between road sections are also relevant. Usually, changes in traffic variables on a certain road section will affect the traffic variables of adjacent roads, and such effects will spread over time, propagating from near areas to far areas. Therefore, when forecasting traffic variables, it is necessary to explore the dependence of changes in traffic variables on time and space at the same time.

In this paper, we use a combination of deep learning models to improve prediction effectiveness and accuracy. According to the characteristics and performance descriptions of the three deep learning models in Section III-B, we apply the ConvLSTM model to learn the spatiotemporal features of the OD stream. For the challenge of the sparsity problem in fine-grained OD flow prediction, we use Conv2DTranpose to capture the sparse features of the data. Finally, SeparableConv2D is used to convolve the travel time OD matrix with the travel frequency OD matrix point-by-point in order to extract spatial correlation features. On this basis, a new OD flow prediction network model CLTS (ConvLSTM-Conv2DTranpose-SeparableConv2D) is designed. The model framework structure is shown in figure 10. The entire model consists of two channels; one feeds the travel frequency matrix between ODs, and the other feeds the travel time matrix between ODs. Each channel consists of two layers of ConvLSTM, one layer of Batch Normalization, and one layer of Conv2DTranpose.

It is worth mentioning that at each time interval t, we will think of the two-layer matrix $C \times C$ containing OD position and neighborhood road network information around it as a tensor $y_t^i \in \mathbb{R}^{C\times C\times 2}$ (with two channels, namely, the travel time matrix and the travel frequency matrix, as shown in figure 10). The matrix can be represented as a time series vector $Y_t = (Y_1, Y_2, Y_3, \cdots, Y_t)$, in which each element represents the traffic flow data of all points including the adjacent regions of the predicted points at the same time.

Here we input OD flow data into the model according to time sequence. First, ConvLSTM is used to capture the spatiotemporal characteristics of the OD flow. Second, Conv2D transpose is used to capture the sparse feature of the OD flow. Third, the results are sent to SeparableConv2D to learn the relationship between the travel time OD matrix and the travel frequency OD matrix. Finally, we feed y_t^i to the connection network to obtain the final prediction value $\hat{y}_{t+1}^i$ of each region. The final prediction function is defined as follows:

$$\hat{y}_{t+1}^i = \sigma\left(W_f y_t^i + b_f\right) \quad (11)$$

where W_f and b_f are learnable parameters, and $\sigma(x)$ is a scaled exponential linear units (SELU) activation function [45], which allows the network to have a self-normalization function. After the activation function, the value distribution of the variable is automatically normalized to zero mean and unit variance. As the demand value is normalized, the output value of the model has limited to the [0,1] interval, and later we will normalize the prediction to obtain the actual demand value.

IV. EMPIRICAL STUDY

In this section, the effectiveness of the proposed taxi OD flow prediction model, CLTS, is evaluated. And in order to show its advance, we compare our model to the existing time sequence prediction models and other deep neural network models.

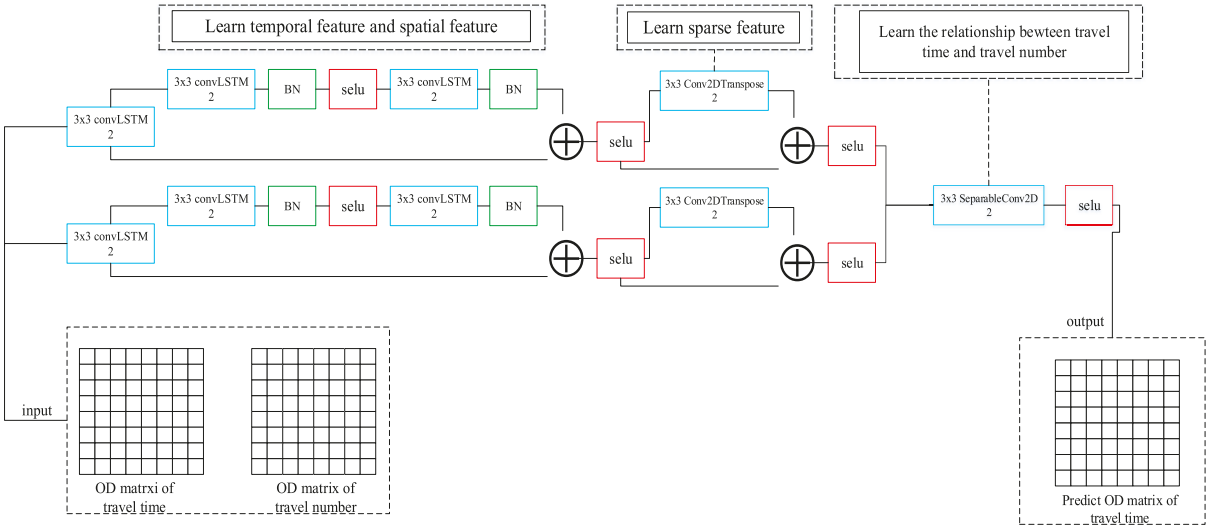


FIGURE 10. Frame structure of the hybrid deep neural network.

A. DATASET AND PLATFORM

In the experiments, we use two datasets for performance evaluation. One is the real taxi trajectory data of Xi'an City, and the other one is the taxicab GPS data in Beijing.[1]

The real trajectory data of more than 10,000 taxis of Xi'an City are used as the test data in the experiment. These data are acquired inside the Third Ring of the city, which covers almost 2/3 area of the city, including various urban functional areas. The road network within this region contains about 12,000 road sections and over 12,800 road nodes. Each taxi is equipped with a GPS data acquisition device that collected GPS data once every 30 s, and the daily operation time is more than 18 hours on average. The GPS data used in this study is collected from September 1, 2016, to September 10, 2016, with the total number of records exceeding 0.243 billion, in which 80% of the data is used as training data and the remaining 20% is used as testing data.

Beijing taxicab GPS data includes two types of crowd flows, i.e. in-flow and out-flow. In-flow means the total traffic of crowds entering a region from other place during a given time interval. On the contrary, out-flow denotes the total traffic of crowds leaving a region for other places during a given period. Both flows track the transition of crowds among regions and can be measured by the number of pedestrians, or the number of cars driven on roads, or the number of people traveling in public transportation systems (e.g., metro, bus), or all of them together if data is available. We choose data collected during the last four weeks as the testing data, and all others as the training data.

Two types of computing platforms used in the experiment are the high-performance computing cluster and the GPU deep learning workstation. The high-performance computing cluster is configured with nine computing nodes, each of which is equipped with a 2.6 GHz dual CPU with a memory of 32 GB. Distributed data processing software systems such as Spark and Hbase are installed for preprocessing and cluster analysis of taxi data. The GPU deep learning workstation is equipped with an Intel Core i7-6500U CPU with a memory of 64 GB and two GTX 1080Ti GPU accelerator cards for deep network training and testing of the performance of OD flow prediction.

B. DATA PREPROCESSING

The area within the Third Ring of the urban area is divided into 256 (16 × 16) grids of the same size. The choice of grid size takes the size and sparsity of the OD flow matrix into account, avoiding oversize and sparseness, which we will discuss in detail in the Model Sensitivity Test in Section IV-D-3.

By querying the attributes of the road network, we know that there are 24,027 road sections in the whole urban area. If all these road sections being considered, the OD matrix will be particularly large and sparse. In this article, we confine the research area within the Third Ring and only study the OD flows of main roads. In this area, there are 912 trunk roads, and some of them pass through multiple grids. For the roads that span multiple grids, the sections of it within different grids will be regarded as different road segments. For example, the main road named A spans the grid 1 and grid 2, when the grid and road nested method is used, the two sections in grid 1 and grid 2 will be coded independently. As shown in figure 11, we conduct 16×16 grid division for the research area, and finally form an OD matrix with the size of 989×989. According to definitions (1) and (2), the travel time and travel frequency of taxi passengers in each time slot are calculated. The calculation results are stored in the corresponding OD travel time matrix and the OD travel frequency matrix, that is, the 2×989×989 three-dimensional matrix, in which 2 represents the dimension and 989×989 represents the size of a single matrix.

[1] https://github.com/lucktroy/DeepST/tree/master/data/TaxiBJ

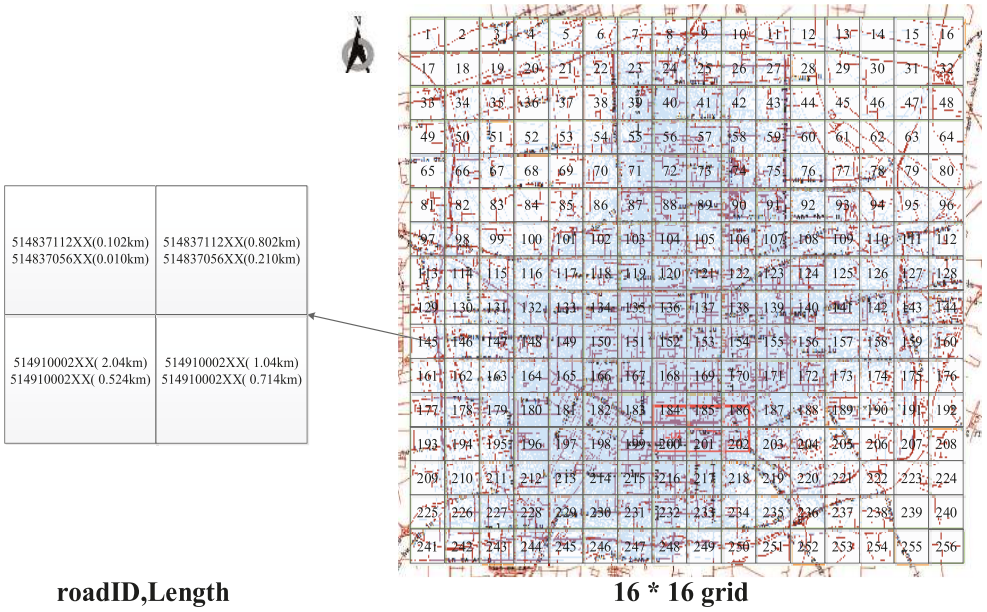


FIGURE 11. Grid scaling at different scales.

Figure 12 is a visual illustration of the actual OD flow matrix, where the color indicates the value of OD travel time or frequency. The brighter the color, the larger the value. Zhang S analyzed the influence of grid size on travel distance distribution. They found that no matter which scale they choose, the spatial distribution of the trip-length is still near the diagonal of the matrix [46]. As seen from figure 12, the OD stream near the diagonal of the OD matrix belongs to a short-range passenger (less than or equal to 6 km), and the OD flow far from the diagonal belongs to a long-haul passenger (greater than 6 km). As shown in figure 12, the spatial distribution of the travel frequency and travel time is similar, except that the order of magnitude is different.

For a single time slot, if each element in the matrix uses a two-byte representation, approximately 7 Mbytes of storage space is required, such that 960 time slots require approximately 6,720 Mbytes of storage capacity. In order to alleviate the storage pressure and reduce the memory requirements in operation, the matrix needs to be compressed. To be specific, the data are stored by using the data partitioning method of the HDF5 filtering channel, in which the matrix is divided into blocks of comparable size, with each block having a size of approximately 10 kB to 1 MB.

C. MODEL TRAINING

The OD flow prediction model constructed in this study is an end-to-end system that requires supervised learning to train model parameters before use.

The optimized Adam (Adaptive Moment Estimation) gradient descent algorithm is used in the network training. This algorithm dynamically adjusts the learning rate of each parameter by using the first-moment estimation and the second-moment estimation of the gradient. After the offset correction, the learning step of each iterative parameter has a certain range. Instead of generating a large learning step because of a large gradient, the values of the parameters are relatively stable, which is beneficial to the stable extraction of characteristics of the OD flow.

In addition, it should be noted that in this study, a custom loss function is used to train the network, which is defined as follows:

$$loss = \frac{1}{N}\sqrt{\sum_{i=1}^{N}\left[(X_i - Y_i)^2 + \omega_i\alpha\right]} \tag{12}$$

where X_i is the actual value and Y_i is the predicted value; α is the absolute value of $X_i - Y_i$; and the value of ω_i is

$$\omega_i = \begin{cases} 0, & if\ Y > 0.5 \\ 1, & if\ \text{otherwise} \end{cases} \tag{13}$$

The regular term $\omega_i\alpha$ is added to this loss function based on the Euclidean distance to ensure the following conditions:

1) During the rush hours in the morning, at noon, and in the evening, the traffic flow is large, the data of the OD matrix are relatively dense, and the prediction results are comparatively accurate. After normalization, the predicted $Y_i > 0.5$, so there is no need to add supplementary weight in the loss function.
2) During the non-rush hours in the early morning, the traffic flow is small, the data of the OD matrix are relatively sparse, and there are a large amount of zero-value data in the matrix, which will affect the accuracy of the prediction. After normalization, the predicted $Y_i \leqslant 0.5$. At this time, it is necessary to add supplementary weight to the loss function to strengthen the impact on the training of network parameters.

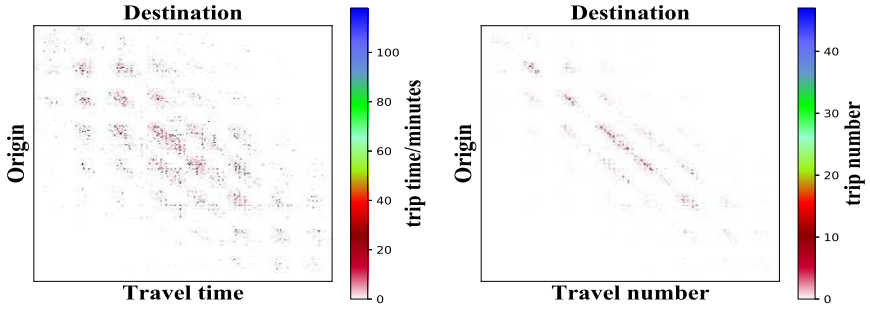


FIGURE 12. Sparse OD matrix with time slot of 15 minutes.

D. PERFORMANCE ANALYSIS

1) MEASUREMENTS AND PARAMETERS

a: EVALUATION INDICES

In order to evaluate the accuracy of the prediction model, the root mean square error (RMSE), mean absolute error (MAE) and mean absolute percentage error(MAPE) are used as the evaluation indicators to analyze the performance of the model, and they are defined as follows:

$$RMSE = \left[\frac{1}{N}\sum_{i=1}^{N}\left(y_i - \hat{y}_i\right)^2\right]^{\frac{1}{2}} \tag{14}$$

$$MAE = \frac{1}{N}\sum_{i=1}^{N}|y_i - \hat{y}_i| \tag{15}$$

$$MAPE = \sum_{i=1}^{N}\frac{|y_i - \hat{y}_i|}{y_i} \times \frac{100}{N} \tag{16}$$

where y_i and $\hat{y}_i$ represent the true OD traffic flow and the predicted OD traffic flow, respectively, and N is the total number of accurately predicted OD flows.

b: MODEL PARAMETER SETTINGS

In the experiment, the Keras2 deep learning framework is applied, in which all the parameters of the model take empirical values; the number of filters in each layer is 2; the kernel size is (3, 3); the learning rate is 0.0002; the number of iterations is set to 140; the attenuation parameter is 0.9; and the batch size is set to 4.

2) PERFORMANCE EVALUATION

a: COMPARE WITH COMMON TRAFFIC FLOW MODELS

Seven models (HA, ARIMA, MLP, SAEs, GRU, LSTM and SRCNs) are selected for performance comparison with CLTS in order to verify the validity of the proposed model. We will introduce the seven models separately:

(1) Historical average (HA). The current traffic flow is predicted by the historical inflow and outflow in the corresponding period. In this paper, the corresponding periods are all historical periods, especially between 7:30 and 7:40 in the morning.

(2) Auto-regressive integrated moving average (ARIMA) [14]. The data structure of the traffic flow time series is decomposed into linear autocorrelation structure, and the moving average and autoregressive components are combined with the modeling of time series.

(3) Multi-layer perceptron (MLP). The MLP is a common neural network prediction model, which continuously trains and learns historical data in a certain period, constructs the optimal model structure between input and expected output, and applies this optimal structure for prediction.

(4) SAEs [24] is a neural network consisting of multiple layers of autoencoder, whose model inputs are encoded into dense or sparse representations before being fed into the next layer. More clearly, considering SAEs with L layer, the first layer is trained like an autoencoder, with the training set as inputs. After obtaining the first hidden layer, the output of the k-th hidden layer is used as the input of the (k+1)-th hidden layer. In this way, multiple autoencoder can be stacked hierarchically. They put a logistic regression layer on top of the network for supervised traffic flow prediction.

(5) GRU [47] has been utilized to capture non-linear traffic behavior efficiently. We use two layers of GRU and a linear regression layer is applied to the output layer of the GRU.

(6) LSTM [34]. This model is suitable for processing and predicting important events with a relatively long interval and delay in time series. In this paper, LSTM is used to process one-dimensional OD flow data.

(7) SRCNs [25] consists of three layers: patch extraction/representation, non-linear mapping and reconstruction. Filters of spatial sizes 9×9, 1×1, and 5×5 were used respectively.

As shown in Table 2, it can be seen from the prediction results of various models that the prediction errors of HA, MLP, and ARIMA are relatively larger. Among them, MLP, ARIMA, SAEs and GRU belong to a shallow network and cannot memorize the OD flow with a relatively long time sequence. Although LSTM can learn the characteristics of OD flow with longer time sequence, it cannot learn the spatial characteristics of OD flow. SRCNs has the smallest RMSE, which shows that it can learn the spatial characteristics of OD flows, but it does not work very well in learning the temporal features. Compared with the other seven models

TABLE 2. Model comparison.

Model	RMSE	MAE	MAPE
HA	1.218	0.102	74.1979%
ARIMA [14]	2.6627	2.079153	14.37674%
MLP	2.52	1.308509	29.709133%
SAEs [24]	2.687849	2.032406	25.146635%
GRU [47]	2.958639	2.035636	19.881214%
LSTM [34]	3.050568	2.097614	19.683256%
SRCNs [25]	**0.000188**	0.074952	6.649495%
CLTS	0.064674	**0.002793**	**5.401400%**

TABLE 3. Comparison on public dataset.

Model	RMSE	MAE	MAPE
SAEs	2.116182	1.606162	22.505910%
GRU	2.116182	1.606162	20.024778%
LSTM	2.229644	1.660732	20.588726%
SRCNs	**0.003010**	0.955625	11.553283%
CLTS	0.16605	**0.82215**	**10.8161%**

in short-term traffic flow prediction, CLTS gives the best performance in MAE and MAPE and ranks only second to SRCNs in RMSE. It is mainly because the CLTS model takes into account the spatiotemporal feature of OD flow, can learn the spatial characteristics and temporal regularity of citywide traffic flow, capture the sparse characteristics of OD flow, and learn the relationship between travel time and travel frequency of OD flow.

b: PERFORMANCE ON THE PUBLIC DATASET

Table 3 shows the comparison results of our model with other baselines on the Beijing crowd flow dataset. Beijing crowd flows consists of two different types of crowd flows, i.e., in-flow and out-flow. As it can be seen that the prediction errors of SAEs, GRU and LSTM are relatively larger. Although, the RMSE of SRCNs is lower than other models, its MAPE is the highest, and its MAE is bigger than CLTS. Compared with the other models in short-term traffic flow prediction, MAE and MAPE of the CLTS model indicate the minimum error, 0.82215 and 10.8161%, respectively.

In general, the performance of our model outperforms the baseline models in OD flows prediction task for its ability to learn the spatial and temporal characteristics of OD flows simultaneously.

3) MODEL SENSITIVITY ANALYSIS

In this section, sensitivity analysis is performed on CLTS. In this process, four parameters are investigated, including loss function, grid size, whether combined with travel time, and length of the time slot.

a: LOSS FUNCTION

The smaller the loss function is, the better a model's robustness will be. In our model, a custom loss function is used, as defined in equation (12). As shown in figure 13, when the training number in our model reaches 15 Epochs, the fitting error between the predicted value and the true value of OD flow starts to decrease; when it reaches 20 Epochs, the error reaches the minimum.

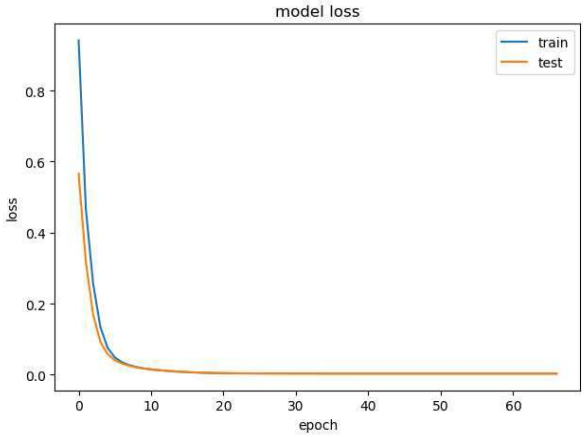


FIGURE 13. Loss function of model.

b: COMPARISON BETWEEN DIFFERENT GRID SIZE AND WITH OR WITHOUT TRAVEL TIME INFORMATION

We follow the method of Zhang et al. to analyze the effects of grid size on taxi OD flow prediction [46]. According to statistics, more than 10,000 taxis in Xi'an have about 3,500 passengers every 15 min. Taking Xi'an Third Ring as an example,if the selected grid size is too large, each grid will contain too much road sections, with just considering main roads, the corresponding OD flow matrix will lose some spatial information of those branch road, resulting in low accuracy. However, if the selected grid size is too small, and each grid covers less or no roads, the corresponding OD matrix is large and too sparse, hence increasing the complexity of the model calculation. Although the spatial information of the road network is preserved, the sparseness problem is still difficult to solve. Based on the above considerations, we not only test the performance of the prediction model with different grid size, but also explore the influence of travel time on predicting travel frequency. The results are shown in Table 4.

Figure 14 shows the visualization of the results of Table 4. From the perspective of grid size, we can see that the RMSE and MAE corresponding to the 24×24 grid are smaller than those of the 8×8 grid and the 16×16 grid. However, the nonzero data of the 24×24 grid only account for 1.15%. However, the nonzero data of the 24×24 grid only account for 1.15%. We know that if such sparse data are directly fed into the model, the training result will be noisy due to the influence of more zero elements, resulting in low training effect. In Table 4, the term "16×16_989" in the grid size column indicates that 989 road segments, which come from 912 main roads, were selected using 16×16 grid as the basis by the method of grid and road nested method, which preserves the spatial relationship between main roads. As can be seen from Table 4, the RMSE and MAE values of the

TABLE 4. Impacts of grid size and travel time on prediction performance.

Grid size	OD matrix size	RMSE only travel frequency	MAE only travel frequency	RMSE with travel time	MAE with travel time	Non- zero ratio
8×8	64×64	1.694105	0.480440	1.245685	0.272000	100%
16×16	256×256	0.628235	0.077143	0.450634	0.042292	5%
24×24	576×576	0.184321	0.014426	0.135494	0.008613	1.15%
$16 \times 16\_989$	989×989	**0.086399**	**0.005010**	**0.064674**	**0.002793**	0.357%

TABLE 5. Comparison experiment on weekdays and weekends.

Time slot	Data scale	Weekdays (RMSE)	Weekdays (MAE)	Weekdays (MAPE)	Weekends (RMSE)	Weekends (MAE)	Weekends (MAPE)
15 min	960 samples	0.064674	0.002793	0.000168	0.072214	0.003051	0.000198
30 min	480 samples	0.104066	0.005612	0.000166	0.119162	0.006098	0.000224
60 min	240 samples	0.192124	0.015548	0.000216	0.217251	0.015778	0.000284

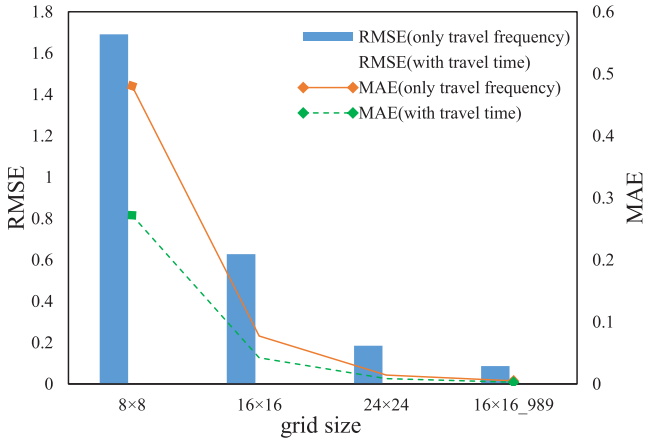


FIGURE 14. Impacts of grid size and travel time on prediction performance.

grid-nested road network method (the last 2 rows) are lower than those of the gridding method only (the first 3 rows), which are 0.064674 and 0.002793, respectively. This proves that the grid and road nested road method can effectively preserve the spatial relationship of the road segments and has higher precision when estimating the OD flows.

We have verified the correlation between travel frequency and travel time in Section II-B, so we combine the OD travel frequency matrix and its corresponding OD travel time matrix in the model training. It can be seen from Table 4 that after the combination of the travel time matrix, the RMSE and MAE of the model test are significantly reduced compared with those without the combination. This shows that the proposed prediction model has learned the relationship between travel time and travel frequency as shown in equation (10) during the training process.

c: VERIFY THE STABILITY OF THE MODEL WITH DIFFERENT LENGTH OF TIME SLOT

To further verify the stability of the algorithm, trajectory data obtained from the actual sampling (sampling time interval is 30 s) are aggregated, and low-frequency sampling datasets with sampling intervals of 15 min, 30 min, and 60 min are constructed. Then, the algorithm in this paper is used to predict the low-frequency datasets, calculate the accuracy of prediction, and analyze the prediction difference between the high-frequency sampling trajectory data and the low-frequency sampling data.

By comparing the OD flow prediction on weekdays with that on weekends, it can be found in Table 5 that the proposed model performs better in predicting OD flows on work days than on weekends. We believe that this is closely related to people's travel patterns. On work days, the travel patterns of residents are relatively fixed and their demands for taxis are relatively stable as well. For example, people tend to go to work from their home and come back home from their work places at the same time each day. Nonetheless, the RMSE is relatively high on weekends when people's travel needs are more diverse, their travel does not follow a fixed pattern, and their demands for taxis are quite variable. The prediction performance of the algorithm proposed in the paper on OD flow at the 15 min low-frequency sampling interval is better than that of the low-frequency sampling data at 30 min and 60 min, which reflects that the influence of the adjacent time-slot OD flow on the current OD flow is stronger than that of the far separated OD flow.

V. CONCLUSION AND DISCUSSION

In this study, we propose a taxi OD flows prediction model based on a hybrid deep neural network to predict fine-grained OD flows under city scale, and jointly use OD travel time and travel frequency in the prediction.

The main contributions are as follows: First, a multi-granular urban space division method with grid-nested road network is proposed, which allows OD flow prediction to reach the road network level and solves the problem that the grid division cannot distinguish traffic flow at different heights in multi-level overpass areas. Second, the ConvLSTM-based hybrid deep neural network prediction

model CLTS is used to learn the spatial and temporal dependence of OD flow. Third, the correlation between OD travel time and travel frequency is explored to train the deep network prediction model by jointly using the OD travel time matrix and the travel frequency matrix. Based on the real taxi GPS trajectory data of Xi'an and Beijing, the performance of the proposed CLTS prediction model is evaluated and compared with existing prediction models (i.e., HA, ARIMA, MLP, SAEs, GRU, LSTM and SRCNs). The results show that the proposed prediction model has better prediction performance.

Undeniably, with the OD flows expression granularity decreasing, the finer-grained OD matrix can more accurately reflect the travel demands and distribution of urban passengers; but at the meantime, it will make the corresponding OD matrix sparser, which brings great challenges to the model training. In this study, only 912 main roads within the Third Ring of Xi'an City are discussed, ignoring most of the branch roads. In future work, the branch information will be taken into consideration, and more external factors, such as weather conditions, are going to be introduced into the network prediction model to further improve the prediction accuracy of OD flows.

In addition, the optimization of functional layer combinations, such as SeparableConv2D+Conv2DTranspose, ConvLSTM+SeparableConv2D, etc., will be considered in the future study to further improve the performance of the prediction model.

ACKNOWLEDGMENT

The authors would like to thank the valuable advice of reviewers and Xi'an Transportation Information Centre for providing the GPS trajectory data of taxis.

REFERENCES

[1] H. J. Van Zuylen and L. G. Willumsen, "The most likely trip matrix estimated from traffic counts," *Transp. Res. B, Methodol.*, vol. 14, no. 3, pp. 281–293, 1980.

[2] A. G. Wilson, "A family of spatial interaction models, and associated developments," *Environ. Planning A, Economy Space*, vol. 3, no. 1, pp. 1–32, Mar. 1971.

[3] C. Xie, K. M. Kockelman, and S. T. Waller, "A maximum entropy-least squares estimator for elastic origin–destination trip matrix estimation," *Transp. Res. B, Methodol.*, vol. 45, no. 9, pp. 1465–1482, Nov. 2011.

[4] T. Siripirote, A. Sumalee, H. W. Ho, and W. H. K. Lam, "Statistical approach for activity-based model calibration based on plate scanning and traffic counts data," *Transp. Res. B, Methodol.*, vol. 78, pp. 280–300, Aug. 2015.

[5] V. Khabarov and A. Tesselkin, "Method for estimating origin-destination matrices using Markov models," in *Proc. 11th Int. Forum Strategic Technol. (IFOST)*, Novosibirsk, Russia, Jun. 2016, pp. 389–393.

[6] H. Shao, W. H. K. Lam, A. Sumalee, A. Chen, and M. L. Hazelton, "Estimation of mean and covariance of peak hour origin–destination demands from day-to-day traffic counts," *Transp. Res. B, Methodol.*, vol. 68, pp. 52–75, Oct. 2014.

[7] W. H. K. Lam and Z. X. Wu, "Estimation of transit passenger origin-destination matrices from passenger counts in congested transit networks," in *Schedule-Based Dynamic Transit Modeling: Theory and Applications*. Boston, MA, USA: Springer, 2004, pp. 175–196.

[8] F. Calabrese, M. Diao, G. Di Lorenzo, J. Ferreira, and C. Ratti, "Understanding individual mobility patterns from urban sensing data: A mobile phone trace example," *Transp. Res. C, Emerg. Technol.*, vol. 26, pp. 301–313, Jan. 2013.

[9] Q. Ge and D. Fukuda, "Updating origin–destination matrices with aggregated data of GPS traces," *Transp. Res. C, Emerg. Technol.*, vol. 69, pp. 291–312, Aug. 2016.

[10] M. S. Iqbal, C. F. Choudhury, P. Wang, and M. C. González, "Development of origin–destination matrices using mobile phone call data," *Transp. Res. C, Emerg. Technol.*, vol. 40, pp. 63–74, Mar. 2014.

[11] L. Alexander, S. Jiang, M. Murga, and M. C. González, "Origin–destination trips by purpose and time of day inferred from mobile phone data," *Transp. Res. C, Emerg. Technol.*, vol. 58, pp. 240–250, Sep. 2015.

[12] A. Abadi, T. Rajabioun, and P. A. Ioannou, "Traffic flow prediction for road transportation networks with limited traffic data," *IEEE Trans. Intell. Transp. Syst.*, vol. 16, no. 2, pp. 653–662, Apr. 2015.

[13] J. W. C. Van Lint and C. P. I. J. Van Hinsbergen, "Short term traffic and travel time prediction models," in *Transportation Research Circular—Artificial Intelligence Applications to Critical Transportation Issues*, R. Chowdhury and S. Sadek, Eds. Washington, DC, USA: National Academies Press, 2012, pp. 22–41.

[14] E. I. Vlahogianni, M. G. Karlaftis, and J. C. Golias, "Short-term traffic forecasting: Where we are and where we're going," *Transp. Res. C, Emerg. Technol.*, vol. 43, pp. 3–19, Jun. 2014.

[15] X. Ma, Z. Dai, Z. He, J. Ma, Y. Wang, and Y. Wang, "Learning traffic as images: A deep convolutional neural network for large-scale transportation network speed prediction," *Sensors*, vol. 17, no. 4, p. 818, Apr. 2017.

[16] G. J. Grindey, S. M. Amin, E. Y. Rodin, and A. Garcia-Ortiz, "A Kalman filter approach to traffic modeling and prediction," *Proc. SPIE*, vol. 3207, pp. 234–241, Jan. 1998.

[17] J. Guo, W. Huang, and B. M. Williams, "Adaptive Kalman filter approach for stochastic short-term traffic flow rate prediction and uncertainty quantification," *Transp. Res. C, Emerg. Technol.*, vol. 43, pp. 50–64, Jun. 2014.

[18] P. Jiao, R. Li, T. Sun, Z. Hou, and A. Ibrahim, "Three revised Kalman filtering models for short-term rail transit passenger flow prediction," *Math. Problems Eng.*, vol. 2016, Mar. 2016, Art. no. 9717582.

[19] K. Ashok and M. E. Ben-Akiva, "Estimation and prediction of time-dependent origin-destination flows with a stochastic mapping to path flows and link flows," *Transp. Sci.*, vol. 36, no. 2, pp. 184–198, May 2002.

[20] X. Chen, S. Guo, L. Yu, and B. Hellinga, "Short-term forecasting of transit route OD matrix with smart card data," in *Proc. 14th Int. IEEE Conf. Intell. Transp. Syst. (ITSC)*, Oct. 2011, pp. 1513–1518.

[21] D. Deng, C. Shahabi, U. Demiryurek, L. Zhu, R. Yu, and Y. Liu, "Latent space model for road networks to predict time-varying traffic," in *Proc. 22nd ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, San Francisco, CA, USA, Aug. 2016, pp. 1525–1534.

[22] M. G. Karlaftis and E. I. Vlahogianni, "Statistical methods versus neural networks in transportation research: Differences, similarities and some insights," *Transp. Res. C*, vol. 19, no. 3, pp. 387–399, Jun. 2011.

[23] S. Ren, B. Yang, L. Zhang, and Z. Li, "Traffic speed prediction with convolutional neural network adapted for non-linear spatio-temporal dynamics," in *Proc. 7th ACM SIGSPATIAL Int. Workshop Anal. Big Geospatial Data*, Seattle, WA, USA, Nov. 2018, pp. 32–41.

[24] 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.

[25] H. Yu, Z. Wu, S. Wang, Y. Wang, and X. Ma, "Spatiotemporal recurrent convolutional networks for traffic prediction in transportation networks," *Sensors*, vol. 17, no. 7, p. 1501, 2017.

[26] Z. Duan, Y. Yang, K. Zhang, Y. Ni, and S. Bajgain, "Improved deep hybrid networks for urban traffic flow prediction using trajectory data," *IEEE Access*, vol. 6, pp. 31820–31827, 2018.

[27] A. De Brébisson, É. Simon, A. Auvolat, P. Vincent, and Y. Bengio, "Artificial neural networks applied to taxi destination prediction," 2015, *arXiv:1508.00021*. [Online]. Available: https://arxiv.org/abs/1508.00021

[28] X. Ma, H. Yu, Y. Wang, and Y. Wang, "Large-scale transportation network congestion evolution prediction using deep learning theory," *PLoS ONE*, vol. 10, no. 3, Mar. 2015, Art. no. e0119044.

[29] J. Zhang, Y. Zheng, and D. Qi, "Deep spatio-temporal residual networks for citywide crowd flows prediction," in *Proc. 21rst AAAI Conf. Artif. Intell.*, New Orleans, LA, USA, Feb. 2017, pp. 1655–1661.

[30] J. Ke, H. Zheng, H. Yang, and X. Chen, "Short-term forecasting of passenger demand under on-demand ride services: A spatio-temporal deep learning approach,"*Transp. Res. C, Emerg. Technol.*, vol. 85, pp. 591–608, Dec. 2017.

[31] H. Yao, F. Wu, J. Ke, X. Tang, Y. Jia, S. Lu, P. Gong, J. Ye, and Z. Li, "Deep multi-View spatial-temporal network for taxi demand prediction," in *Proc. 22nd AAAI Conf. Artif. Intell.*, New Orleans, LA, USA, Feb. 2018, pp. 2588–2595.

[32] Y. Li, R. Yu, C. Shahabi, and Y. Liu, "Diffusion convolutional recurrent neural network: Data-driven traffic forecasting," 2017, *arXiv:1707.01926*. [Online]. Available: https://arxiv.org/abs/1707.01926
[33] F. Qian, G. Hu, and J. Xie, "A recurrent neural network approach to traffic matrix tracking using partial measurements," in *Proc. 3rd IEEE Conf. Ind. Electron. Appl.*, Jun. 2008, pp. 1640–1643.
[34] F. Toqué, E. Côme, M. K. El Mahrsi, and L. Oukhellou, "Forecasting dynamic public transport Origin-Destination matrices with long-Short term Memory recurrent neural networks," in *Proc. IEEE 19th Int. Conf. Intell. Transp. Syst. (ITSC)*, Nov. 2016, pp. 1071–1076.
[35] A. M. Nagy and V. Simon, "Survey on traffic prediction in smart cities," *Pervas. Mobile Comput.*, vol. 50, pp. 148–163, Oct. 2018.
[36] N. G. Polson and V. O. Sokolov, "Deep learning for short-term traffic flow prediction," *Transp. Res. C, Emerg. Technol.*, vol. 79, pp. 1–17, Jun. 2017.
[37] Y. Tian, K. Zhang, J. Li, X. Lin, and B. Yang, "LSTM-based traffic flow prediction with missing data," *Neurocomputing*, vol. 318, pp. 297–305, Nov. 2018.
[38] Z. Chen, "Research on extraction and analysis of taxi passenger travel characteristics based on big data," M.S. thesis, School Inf. Eng., Chang'an Univ., Xi'an, China, 2017.
[39] Y. Wang, Y. Zheng, and Y. Xue, "Travel time estimation of a path using sparse trajectories," in *Proc. 20th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, New York, NY, USA, USA, Aug. 2014, pp. 25–34.
[40] X. Shi, Z. Chen, H. Wang, D.-Y. Yeung, W.-K. Wong, and W.-C. Woo, "Convolutional LSTM network: A machine learning approach for precipitation nowcasting," in *Proc. Adv. Neural Inf. Process. Syst. (NIPS)*, Montreal, Canada, 2015, pp. 802–810.
[41] Y. Liu, H. Zheng, X. Feng, and Z. Chen, "Short-term traffic flow prediction with Conv-LSTM," in *Proc. 9th Int. Conf. Wireless Commun. Signal Process., (WCSP)*, Nanjing, China, Oct. 2017, pp. 1–6.
[42] Z. Yuan, X. Zhou, S. Wang, and T. Yang, "Hetero-ConvLSTM: A deep learning approach to traffic accident prediction on heterogeneous spatio-temporal data," in *Proc. 24th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, London, U.K., Aug. 2018, pp. 984–992.
[43] M. D. Zeiler, D. Krishnan, G. W. Taylor, and R. Fergus, "Deconvolutional Networks," in *Proc. Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, Jun. 2010, pp. 2528–2535.
[44] L. Kaiser, A. N. Gomez, and F. Chollet, "Depthwise separable convolutions for neural machine translation," 2017, *arXiv:1706.03059*. [Online]. Available: https://arxiv.org/abs/1706.03059
[45] G. Klambauer, T. Unterthiner, A. Mayr, and S. Hochreiter, "Self-normalizing neural networks," in *Proc. Adv. Neural Inf. Process. Syst. (NIPS)*, Long Beach, CA, USA, 2017, pp. 971–980.
[46] S. Zhang, D. Zhu, X. Yao, X. Cheng, X. He, and Y. Liu, "The scale effect on spatial interaction patterns: An empirical study using taxi O-D data of Beijing and Shanghai," *IEEE Access*, vol. 6, pp. 51994–52003, 2018.
[47] R. Fu, Z. Zhang, and L. Li, "Using LSTM and GRU neural network methods for traffic flow prediction," in *Proc. 31st Youth Academic Annu. Conf. Chin. Assoc. Automat., (YAC)*. Wuhan, China: Institute of Electrical and Electronics Engineers, Nov. 2016, pp. 324–328.

ZONGTAO DUAN was born in Baoji, Shaanxi, China, in 1977. He received the Ph.D. degree in computer science from Northwestern Polytechnical University, China, in 2006.

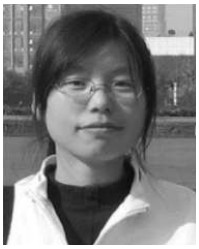
He was a Postdoctoral Research Fellow with the University of North Carolina, USA, from 2009 to 2010. He is currently a Professor with the School of Information Engineering, Chang'an University, China. His research interest includes context-aware computing in transportation. He is a member of CCF, CCF High Performance Computing, and the Pervasive Computing Technical Committee.

KAI ZHANG received the M.Sc. degree in computer technology, in 2019, under the supervision of Prof. Z. Duan. He is currently pursuing the Ph.D. degree with Southeast University, Nanjing, China, under the supervision of Prof. Z. Liu. His research interests include traffic data analysis, dynamic traffic modeling, and machine learning.

ZHE CHEN was born in Xi'an, Shaanxi, China, in 1969. He received the Ph.D. degree in computer science from Northwestern Polytechnical University, Xi'an, China, in 2005. He is currently an Associate Professor with the School of Information Engineering, Chang'an University, Xi'an. His research interests include computer vision, machine learning, and intelligent transportation systems. He is a member of ACM and the China Computer Federation.

ZHIYUAN LIU received the Ph.D. degree in transportation engineering from the National University of Singapore (NUS), in 2011. He is currently a Professor with the School of Transportation, Southeast University, Nanjing, China.

LEI TANG was born in Jiangyou, Sichuan, China, in 1983. She received the Ph.D. degree in computer science and technology, in 2012. She was a Visiting Researcher with the Chair of Information Systems, Mannheim University, Germany, from 2009 to 2010. She is currently with the School of Information Engineering, Chang'an University, China. Her research interests include the areas of intelligent transportation systems and pervasive computing. She is a member of ACM and the China Computer Federation.



YUN YANG received the Ph.D. degree in control theory and control engineering from the South China University of Technology, Guangzhou, China, in 2007. She is currently an Associate Professor with the School of Information and Engineering, Chang'an University, Xi'an, China. Her research interests include computer vision, machine learning, and intelligent transportation systems.

YUANYUAN NI received the B.Eng. degree in communication engineering, in 2017. She is currently pursuing the M.Sc. degree with Chang'an University, Xi'an, China, under the supervision of Prof. Z. Duan and Associate Prof. Y. Yang. Her research interests include urban computing, machine learning, and artificial neural networks.