DeepTSP: Deep traffic state prediction model based on large-scale empirical data

(article starts on next page)

Communications in Transportation Research 1 (2021) 100012

Contents lists available at ScienceDirect

Communications in Transportation Research

journal homepage: www.journals.elsevier.com/communications-in-transportation-research

DeepTSP: Deep traffic state prediction model based on large-scale empirical data

Yang Liu [a,b,**], Cheng Lyu [b], Yuan Zhang [b], Zhiyuan Liu [b,*], Wenwu Yu [c], Xiaobo Qu [a]

[a] Department of Architecture and Civil Engineering, Chalmers University of Technology, Gothenburg, Sweden
[b] Jiangsu Key Laboratory of Urban ITS, Jiangsu Province Collaborative Innovation Center of Modern Urban Traffic Technologies, School of Transportation, Southeast University, China
[c] Jiangsu Provincial Key Laboratory of Networked Collective Intelligence, School of Mathematics, Southeast University, Nanjing, China

ARTICLE INFO

Keywords:
Large-scale traffic prediction
Traffic state propagation
Spatio-temporal data

ABSTRACT

Real-time traffic state (e.g., speed) prediction is an essential component for traffic control and management in an urban road network. How to build an effective large-scale traffic state prediction system is a challenging but highly valuable problem. This study focuses on the construction of an effective solution designed for spatio-temporal data to predict the traffic state of large-scale traffic systems. In this study, we first summarize the three challenges faced by large-scale traffic state prediction, i.e., scale, granularity, and sparsity. Based on the domain knowledge of traffic engineering, the propagation of traffic states along the road network is theoretically analyzed, which are elaborated in aspects of the temporal and spatial propagation of traffic state, traffic state experience replay, and multi-source data fusion. A deep learning architecture, termed as **Deep Traffic State Prediction** (DeepTSP), is therefore proposed to address the current challenges in traffic state prediction. Experiments demonstrate that the proposed DeepTSP model can effectively predict large-scale traffic states.

1. Introduction

Unprecedented urbanization has resulted in the expansion of the cities in terms of both population and area. In response to the challenges of mobility and sustainability, accurate traffic state prediction is required to guide the planning and management of urban transportation systems (Hoque et al., 2021). Apart from substantial economic and environmental values, improved prediction accuracy can also assist organizations and policymakers to better understand the underlying mechanism of how urban traffic runs and thus develop more rational transportation policies. Changes in traffic states reflect how travel demand fluctuates over time, which has already been modelled early back in the last century but in a simplified way due to data deficits (Manley and Cheng, 2018). Although recent advances in data science enable a more fine-grained analysis of traffic, the model complexity can be too high when applied in a large urban region. Therefore, accuracy, scalability and robustness should both be considered to support proactive and prompt decision-making such as adaptive traffic control strategies.

Essentially, traffic dynamics is a multivariate time series, and the prediction of future traffic should be made based on the changes in historical traffic states. One of the most classical methods for the time series prediction is the autoregressive integrated moving average (ARIMA) (Ahmed and Cook, 1979) and its numerous variants (Chen et al., 2019a; Milenković et al., 2016). However, the ARIMA-family models hold the assumption that the time series is stationary in time, and the prediction stage is usually computationally expensive when dealing with a large amount of data, especially when involving seasonal components. Beyond ARIMA, another line of research formulated the traffic prediction Problem on the basis of traffic assignment, including activity-based travel demand models (Liu et al., 2015; Wang et al., 2020). The growing popularity of machine learning encourages researchers to formulate the problem as a supervised learning problem. To capture the underlying nonlinear relationships, models like support vector machine (Castro-Neto et al., 2009), *k*-nearest neighbor (Cai et al., 2016), extreme value count model (Sohrabi et al., 2020), and stacked autoencoders (Lv et al., 2014) were adopted. However, most of the early works focused on the temporal dimension only, and their study objectives are highways or several road sections rather than urban road networks.

In recent works, the correlations in the spatial and temporal dimensions are extensively studied, which are often supported by emerging

* Corresponding author.
** Corresponding author. Department of Architecture and Civil Engineering, Chalmers University of Technology, Gothenburg, Sweden.
E-mail addresses: liuy@chalmers.se (Y. Liu), zhiyuanl@seu.edu.cn (Z. Liu).

https://doi.org/10.1016/j.commtr.2021.100012
Received 22 August 2021; Received in revised form 11 November 2021; Accepted 11 November 2021
Available online 11 December 2021

deep learning techniques (Chen et al., 2019b; Li et al., 2020; Zhan et al., 2020). Intuitively, regions that are spatially or semantically close (i.e., share similar land-use structures) incline to show similar traffic patterns. Therefore, convolutional neural network (CNN) (Yao et al., 2018; Zhang et al., 2016, 2017), which has witnessed great success in the field of computer vision, is frequently applied in traffic prediction tasks to handle the spatial correlations. Meanwhile, it can also model temporal patterns by stacking traffic maps as different channels. For example, Zhang et al. (2016) adopted a convolutional structure to capture the spatial dependency between regions. Traffic maps in different time slices were organized as channels to address the temporal relations. Instead of channels, Guo et al. (2019) directly dealt with the temporal dimension using 3D convolutions. They also designed a recalibration unit to capture the heterogeneity in spatial dependency. Different from many CNN-based models that generate traffic maps based on the geographic location of sensors, Dai et al. (2019) managed to rearrange those pixels according to correlation coefficients. Recurrent neural network (RNN) was combined with CNN by Ke et al. (2017) and Ma et al. (2020) to model the temporal dynamics of traffic. Yao et al. (2018) decomposed complex traffic patterns into three views, where the spatial information was handled by CNN, the temporal information was handled by a variant of RNN (e.g., long short term memory network), and the semantic dependency was extracted by structural embedding.

In addition, some researchers pointed out that a square grid structure is more suitable for data like images than urban road networks. For example, Ke et al. (2018) proposed a hexagonal CNN for taxi demand prediction as the hexagonal segmentation is superior to squares in terms of the neighborhood Definition, edge-to-area ratio, and isotropy. Some also argued that the topological structure of the road network can be better captured by graph-based models (Pan et al., 2020). Li et al. (2018) first introduced graph modeling into traffic prediction, and diffusion convolutions were applied to learn the spatial dependency of traffic flow. For modeling temporal dynamics, they embedded the diffusion convolution in the gated recurrent unit (GRU), a powerful variant of RNN. Similarly, Yu et al. (2018) also adopted the graph convolution operation for traffic prediction. Zhao et al. (2019) proposed a sequential combination of graph convolution and GRU. Lin et al. (2018) proposed a framework for bike demand prediction based on a graph convolution network to better capture relationships between docking stations. A graph filter was also designed to incorporate the heterogeneity between stations. Graph wavelet transform, which is flexible for extracting local spatial patterns for traffic prediction, was proposed by Cui et al. (2020) to improve common graph convolutional neural networks. Attention mechanism was included in the graph convolution framework proposed by Zhang et al. (2019b) to help determine the relevance of input information to traffic speed prediction.

Furthermore, multi-source data are also introduced in some literature. One motivation is to allow for external influencing factors, such as weather conditions and neighboring points-of-interest (Liao et al., 2018; Zhang et al., 2019a; Wessel, 2020). Some works even managed to leverage social media information to improve prediction accuracy (He et al., 2013). Another motivation is to impute the missing data in areas where no sensors are installed or in time slices when probes fail. For example, cellphone location data is potential to infer the traffic volume due to the high correlation between each other (Liu et al., 2019).

1.1. Challenges

Despite extensive research in this field, which aim at improving traffic conditions, fleet organization, utilization rate, and social welfare, the effective prediction of traffic states is still challenging in terms of data scale, granularity, and sparsity.

(a) Citywide traffic prediction is valuable in understanding the complex urban transportation system of the whole city, but is often dissuaded by its scale. Many megacities nowadays have grown to spread over thousands of square kilometers. Limited by the data availability, the study areas of many outstanding works on spatio-temporal prediction merely cover a small portion of a city (Pan et al., 2019; Yao et al., 2018; Zhang et al., 2019a). The main issues hindering the citywide traffic prediction are the low coverage of traffic sensors, including fixed sensors (e.g., RFID) and moving sensors (e.g., floating cars). Thanks to the efforts of the world's leading research institutions like the Institute of Advanced Research in Artificial Intelligence (IARAI),[1] the large-scale traffic data generated by over 100 million probes have made the prediction possible.

(b) Spatio-temporal data is sensitive to the granularity of data. The information loss of coarse data may exert unmeasurable negative impacts on model performance. Many studies on spatio-temporal data prediction that yield state-of-art performance split the study area into a low-resolution image. For example, the seminal work of Zhang et al. partitioned two cities (i.e., Beijing and New York) into a grid map of 16×8 subregions, where the size of each subregion is $1km \times 1km$. However, the intersections and links are hardly identifiable in such resolution. The topological information of the road network will be lost due to the coarse map segmentation. Therefore, it is extremely important to preserve fine-grained information. Provided that the size of the subregion is smaller, the network topology can be better kept.

(c) In the spatial dimension, although the high-resolution map segmentation is useful in preserving the information of network topology, it may introduce sparsity into the data. The values of traffic states in many subregions are zero, since not all of them contain roads where vehicles can run. For example, in the large-scale dataset used in this paper, only approximately 6% of pixels are covered by the road network. Sparsity also exists in the temporal dimension. The large size of the study area, high resolution of grid map, a limited number of floating cars, as well as unpredictable communication error and device malfunctioning, all contribute to a large number of missing values in the dataset. In contrast, few values are missing in typical traffic time series, and characteristics like peak hours and periodicity can be easily observed.

2. Contributions

The objective of this study is to address the large-scale traffic state prediction problem based on deep learning. Existing studies are mainly direct applications of machine learning techniques, in lack of thorough analysis of traffic state dynamics based on domain knowledge and physical interpretations. The fusion of interdisciplinary knowledge and experience is an emerging trend in traffic research, especially beneficial to building effective deep learning models. Additionally, these researches are often evaluated on part of a city rather than the entire area in a high resolution, hence limited value of practical applications. Compared with existing works, the contributions of this paper are twofold.

(a) Based on the domain knowledge of traffic engineering, the propagation of traffic states is theoretically analyzed. There have been studies working on combining classical traffic flow theories with deep learning (e.g., Shi et al., 2021), which, however, are limited to traffic state estimation rather than prediction. In this research, we manage to design the architecture of the deep learning network based on the analysis results. Several modules are designed and elaborated in the aspects of the temporal and spatial propagation of traffic state, etc. Aside from this, inspired by the experience replay in online reinforcement learning, we propose the Traffic State Experience Replay method in addressing

[1] https://www.iarai.ac.at/.

aperiodic traffic data. Faced with the challenge of data sparsity, a mask is designed to improve the efficiency of the model.

(b) The proposed deep learning framework is validated on a dataset much larger than those used in previous studies. Large-scale datasets like ImageNet have spawned the prosperity of computer vision techniques (Deng et al., 2009). The dataset used in this paper comes from NeurIPS Traffic4cast Challenge, which is the largest spatio-temporal data prediction task challenge to date and is remarkably potential to become the 'ImageNet' in the field of traffic engineering. The comparison between this dataset and several existing datasets for spatiotemporal prediction is shown in Table 1. TaxiBJ and BikeNYC first appeared in the study of Zhang et al. (2016). Yao et al. (2018) evaluated their model on the traffic dataset released by Didichuxing. Liang et al. (2019) researched crowd flow based on data from Beijing Happy Valley. In this study, we perform a fine-grained (100m × 100 m) traffic state prediction using the industrial-scale real world probe data in three cities, covering approximately six thousand square kilometers in total for the first time. Experiments at such a scale can significantly help the understanding of the citywide complex urban transportation system. Moreover, instead of the graph representation of road networks used in many recent studies, we adopt the grid representation to minimize the reliance on road network data. High-quality updated road maps are often unavailable, especially in the fast-evolving cities of emerging economies.

The remainder of this paper is organized as follows. In Section 2, the terms and problem in this study are first formulated, and the large-scale traffic dataset is then introduced. In Section 3, the motivations and the designs of the proposed modules and framework are analyzed and elaborated, while in Section 4, we evaluate the performance of our framework. Finally, the conclusions are given in Section 5.

3. Problem and data

3.1. Problem definition

Definition. high-resolution traffic map & spatio-temporal cell

The whole city can be seen as a high-resolution traffic map, each pixel of which represents a subregion of $p \times p$ square meters and is called a spatio-temporal cell. The high-resolution traffic map was derived from trajectories reported by a large fleet of probe vehicles over the course of a year. The sampling frequency for each frame is q minutes.

Problem. This problem is to use the given data to predict the speed at each spatio-temporal cell (pixel) separately in the next several time slices. The speed in all spatio-temporal cells at the t-th time slice of the d-th day can be denoted as a tensor $\boldsymbol{S}_d^t$.

Table 1
Comparison of datasets.

Data type	TaxiBJ	BikeNYC	Didi	HappyValley	This study
	Traffic flow	Traffic flow	Traffic demand	Crowd flow	Traffic flow, speed, direction
Time slice length	0.5 h	1 h	0.5 h	0.5 h	5 min
Traffic map size	(32,32)	(16,8)	(20,20)	(25,50)	(436,495)
Number of pixels	1024	128	400	1250	215820
Cell size	1km × 1 km	1km × 1 km	0.7km × 0.7 km	10m × 10 m	100m × 100 m
Study area coverage	1024 km^2	128 km^2	196 km^2	0.125 km^2	2158.2 km^2

3.2. Data description

The proposed methods in this paper will be evaluated on the industrial scale traffic state data provided by IARAI, which is prepared for the NeurIPS 2019, Traffic4cast Challenge. In the datasets, the historical traffic data of three cities, including Berlin, Istanbul, and Moscow, in 2018 were recorded. All of the frames contain three channels, representing the traffic flow, speed, and direction respectively. All the data of each city is transformed into a high-resolution traffic map, consisting of 436 × 495 pixels and covering over 2000 square kilometers. For each pixel, it records the traffic information in an area of 100 × 100 square meters. The data are aggregated on a 5-min basis, equivalent to a total of 288 frames per day. Moreover, the data of flow and speed has been standardized to the range from 0 to 255 through a min-max scaler. The heading angles of probes from due north are recorded as the directions in the data and are binned into four numbers, i.e., $[1, 85, 170, 255]$, indicating northeast, southeast, southwest, and northwest respectively. During aggregation, the direction bin that contains the largest number of records is labeled as the direction of traffic flow in a pixel in that time slice.

The spatial distributions of raw data of the three cities are illustrated in Fig. 1(a–c). Note that we map all the values of the raw dataset within the range from 0 to 255. An image from a standard digital camera will have a red, green, and blue channel. In this study, each type of data is then treated as one channel of an image (traffic flow, speed, and direction channel), which are visualized in Fig. 1(a–c). The speed variations of a pixel in Fig. 1(a) on a day is plotted in Fig. 1(d), where an abnormal spike can be found at time slice 248. These short spikes are usually caused by traffic congestion and can be propagated to neighboring subregions. For instance, a similar spike shows up at time slice 250 in an adjacent pixel, as plotted in Fig. 1(e). A model that is capable of learning the propagation mechanism of traffic state can better serve the demand for speed prediction in urban transportation networks. The prediction results are valuable in warning the intelligent transportation system about the potential congestion, hence a prompt reaction to it.

4. Framework overview

4.1. Motivation

As illustrated in Fig. 2, the urban road network is represented as a graph, where the nodes stand for junctions and the edges stand for links. For the ease of understanding the propagation of traffic states, the following analysis on congestion propagation is presented. When traffic congestion occurs in node a of the network, it will propagate upstream, leading to congestion in nodes nearby. Inspired by traffic flow theory, to deduce which node will be affected by the congestion in the near future, we have to recognize the flow of each direction in node $\boldsymbol{a}$. Further, by considering the distances between node a and neighboring nodes, the time when the congestion might happen can be deduced.

4.2. System framework

The graph representation of the road network can well preserve the topological information of the road network. However, such modeling method heavily relies on high-quality updated road maps, which are often unavailable, especially for megacities covering thousands of square kilometers (especially for their suburban areas). On top of the availability, the representation of the whole urban road network will result in millions of nodes. Currently, it is still a challenge regarding the computation of large-scale graphs.

In response to this problem, we compromise to divide the city into a high-resolution traffic map, where each pixel aggregates the information of traffic states inside a subregion of $p \times p$ square meters. Adequately small pixels (i.e., subregions) can preserve the topological information to some extent, without the need for high-quality up-to-date road maps.

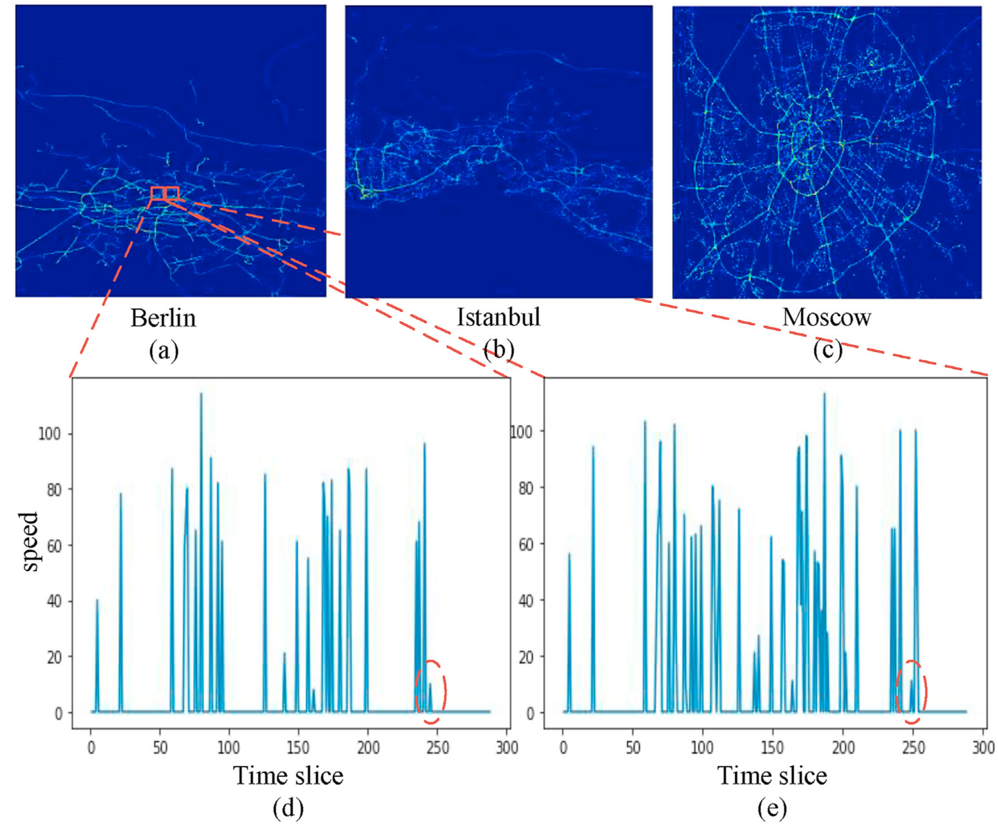

Fig. 1. The spatial distribution of raw data.

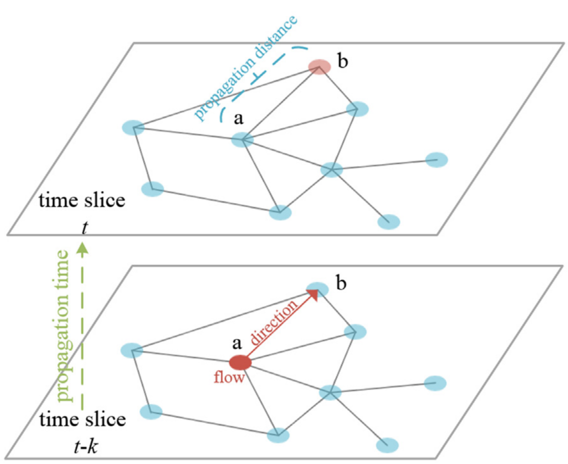

Fig. 2. Illustration of traffic congestion propagation.

Fig. 3 presents the framework of our solution, which models the temporal propagation of traffic state, spatial propagation of traffic state, etc. In the next sections, the detailed designs of the proposed framework are analyzed and elaborated.

4.3. Temporal propagation of traffic state

In general, traffic state prediction is essentially a time-series prediction problem. For one-dimensional time-series prediction, a typical solution is to transform the problem into a supervised learning problem with the help of feature engineering. In our framework, many features are designed based on the domain knowledge of traffic engineering, and these features are later input into the model as different channels.

Typical traffic time series data are intrinsically continuous rather than discrete. Fig. 4 shows a representative time series plot of the average speed changes on a day in Berlin. Normally, the temporal dependency is critical in time series prediction, because the traffic state at a specific time slice has a close resemblance to the traffic state at the adjacent time slice. Hence, normally, to predict the speed in at time t of day d, denoted by $\boldsymbol{S}_d^t$, speeds at $t-1, t-2, \ldots$, and $t-k$ will be treated as input features, represented as $[\boldsymbol{S}_d^{t-1}, \boldsymbol{S}_d^{t-2}, \ldots, \boldsymbol{S}_d^{t-k}]$.

However, as shown in Fig. 5, only a few time slices have speed record at a pixel (where the zeros in the figure imply that there is no floating car passing a specific subregion at that time slice) since the penetration of floating cars is limited. Based on the analysis of Fig. 2, the traffic state will be propagated to other subregions along the links of road network after k time slices. Therefore, to predict the city-wide speed at time t, we should include the speed snapshots at time $t-1$, $t-2$, and $t-k$ in the input feature map, so that the model can learn such propagation of traffic state in the temporal dimension. It should be noted that, though represented as $[\boldsymbol{S}_d^{t-1}, \boldsymbol{S}_d^{t-2}, \ldots, \boldsymbol{S}_d^{t-k}]$, they indicate the propagation of traffic state in this paper.

4.4. Spatial propagation of traffic state

One important attribute of the data is the spatial correlation, which means that the traffic state in a specific subregion can be affected by adjacent subregions. Similar to what happens in the graph of a road network, when congestion occurs in pixel a of the traffic map, it might propagate to pixels nearby along the links. The flows of every direction in pixel a are required to deduce which and when neighboring pixels (subregions) will be congested after k time slices, as illustrated in Fig. 6.

Different from common tasks relating to computer vision, the extracted features serving the traffic prediction have to be endowed with more solid physical meaning and interpretability. The receptive field of the model needs to be large enough to capture the traffic states in distant

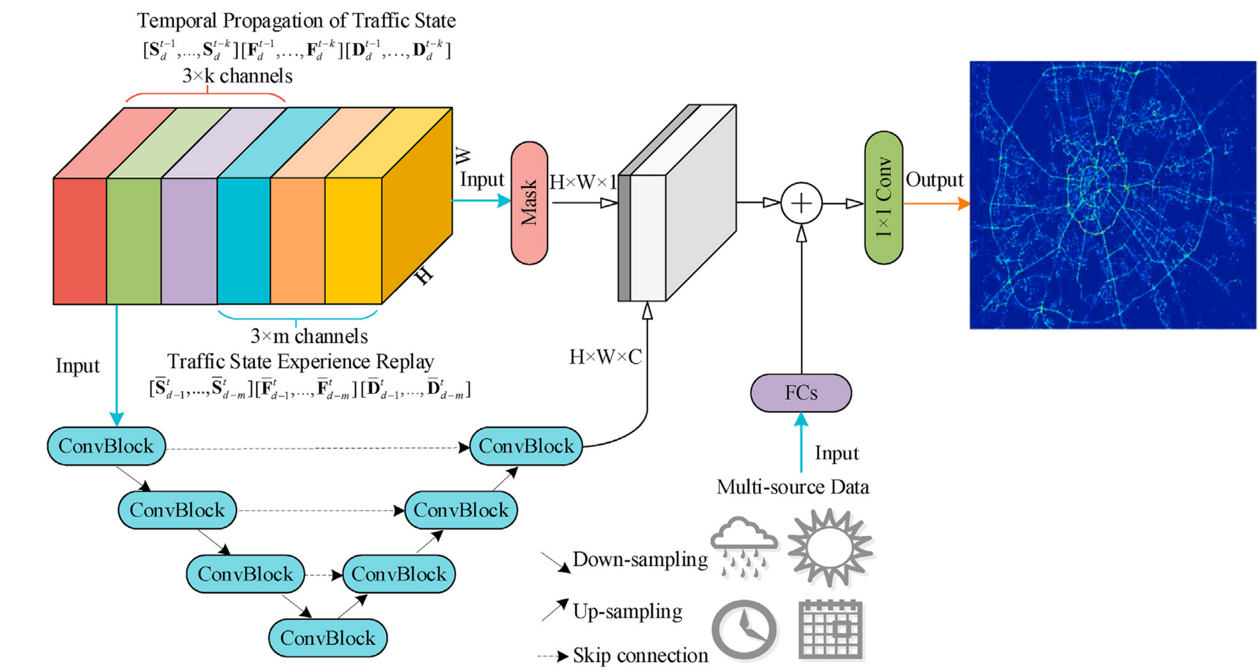

Fig. 3. System framework. FC: Fully-connected layer; ConvBlock: It consists of the repeated application of two 3 × 3 convolutions layers, each followed by a batch normalization (BN) and a rectified linear unit (ReLU). Mask: A module we designed to solve data sparsity (see Section *Use Mask Instead of Spatial Attention*).

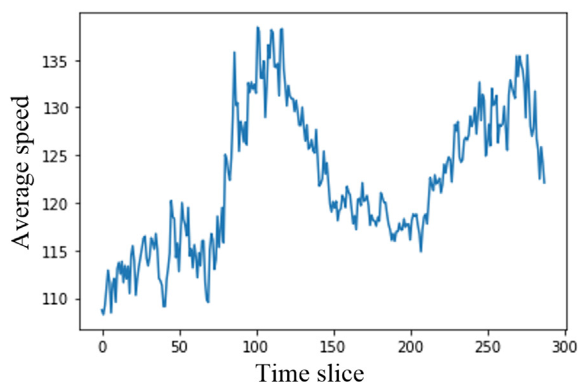

Fig. 4. Average speed on a day in Berlin.

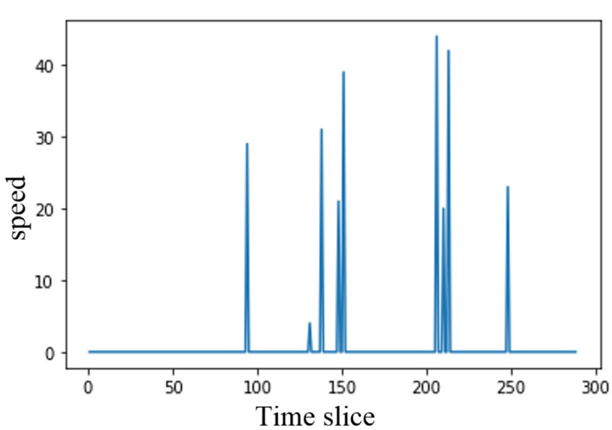

Fig. 5. Speed on a day at a pixel in Berlin.

subregions. Additionally, researchers working on analytical models found that the global average of traffic variables is one of the most significant factors contributing to the prediction. Information like this

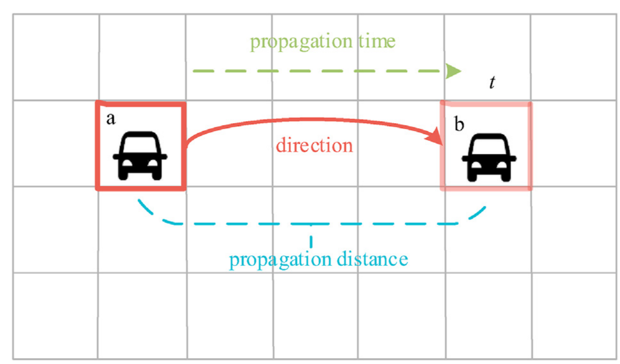

Fig. 6. Illustration of spatial propagation of traffic state.

cannot be learned by the model if the receptive field is not adequately large. Existing state-of-the-art models on traffic prediction can all yield accurate results but are limited to small study areas, hence small-scale traffic maps. In their cases, the receptive field can be large enough with several convolutional layers applied. Generally, the simplest approaches to increase the receptive field include enlarging the size of convolutional kernels and increasing the number of convolutional layers. However, considering the size of traffic maps in this study, i.e., 436 × 495 pixels, neither large convolutional kernel size nor more convolutional layers are applicable, both of which are computational demanding. Therefore, to increase the receptive field, the pooling operation is a more feasible way. Since this task is similar to the pixel-wise image segmentation, we design our framework referring to the structure of U-Net (Ronneberger et al., 2015), and many improvements are made based on the domain knowledge of the urban transportation network.

4.5. *Traffic state experience replay*

Periodicity is one of the common characteristics of traffic time series data. In other words, fixed patterns might repetitively occur as the traffic states change. For example, commuters can find that the metro always reaches a crowded state between 8:00 a.m. and 10:00 a.m., i.e., the morning peak on workdays. If we take a look at the daily variations in passenger flow, the similarity between workdays can be easily observed. Likewise, we may also expect similar patterns to be found in the flow changes of floating cars, as well as other vehicles on the urban road network.

Nevertheless, as has been pointed out in the previous section, the low penetration of floating cars leads to sparsity in data. As a result, the periodicity cannot be observed from the plots. Inspired by the experience replay in online reinforcement learning, through which agents can recall the experiences in the past, we adopt this concept in addressing aperiodic traffic data. Historical traffic states in near time slices will be replayed for the model to learn the propagation of states. In detail, considering that the propagation of traffic state is a continuous process, the speeds in k neighboring time slices of day $d-1$ can be used as inputs, denoted by $[\boldsymbol{S}_{d-1}^{t-k}, \boldsymbol{S}_{d-1}^{t+k}]$. If the traffic state is witnessed to propagate up here in historical data, the model can then reuse the experiences from the past. Also, the traffic states in previous m days can be reused for experience replay and can be represented as follows,

$$\left[\boldsymbol{S}_{d-1}^{t-k}, \boldsymbol{S}_{d-1}^{t+k}\right] \cup \left[\boldsymbol{S}_{d-2}^{t-k}, \boldsymbol{S}_{d-2}^{t+k}\right] \cdots \cup \left[\boldsymbol{S}_{d-m}^{t-k}, \boldsymbol{S}_{d-m}^{t+k}\right] \tag{1}$$

Note that the number of input channels is $(2 \times k + 1) \times m$. When m and k are large, there will be too many channels to input, which needs to be further compressed. For example, if we are going to compress traffic states in the m-th day, average global pooling operations can be applied to $[\boldsymbol{S}_{d-m}^{t-k}, \boldsymbol{S}_{d-m}^{t+k}]$ along the channel dimension (i.e., calculating the average speed of $2 \times k + 1$ time slices). Then, the number of channels will be reduced from $2 \times k + 1$ to 1. This process can be formulated as follows,

$$\boldsymbol{S}_{d-m}^{t} = global\_pooling\left(\left[\boldsymbol{S}_{d-m}^{t-k}, \boldsymbol{S}_{d-m}^{t+k}\right]\right) \tag{2}$$

where $\boldsymbol{S}_{d-m}^{t}$ denotes the compressed traffic states. The channels in Eq. (1) can be represented as follows after compression,

$$\left[\boldsymbol{S}_{d-1}^{t}, \boldsymbol{S}_{d-2}^{t}, \ldots, \boldsymbol{S}_{d-m}^{t}\right] \tag{3}$$

4.6. Multi-source data fusion

The fusion of multi-source data is frequently reported to be useful to assist traffic prediction. In our model, we introduce external information (e.g., weather condition, holiday, the day of week, the time of day) into the models. The external features are embedded into x-dimensional vector through two fully connected layers and are added to the U-Net.

Apart from the external information, traffic flow and direction, which can be regarded as another two sources of data, are also utilized to facilitate the prediction of speed. To predict the speed at time slice t, denoted by $\boldsymbol{S}_d^t$, the flow and direction at time slice $t-1, t-2, \ldots, t-k$ will be included in the input features, represented as $[\boldsymbol{F}_d^{t-1}, \boldsymbol{F}_d^{t-2}, \ldots, \boldsymbol{F}_d^{t-k}]$ and $[\boldsymbol{D}_d^{t-1}, \boldsymbol{D}_d^{t-2}, \ldots, \boldsymbol{D}_d^{t-k}]$. Each frame in the data aggregates the traffic variables in 5 min. When the data of the previous 1 h are used to learn the temporal propagation of traffic state, there will be 12×3 channels, where the number 3 represents speed, flow, and direction.

For experience replay, the traffic flow and direction in k adjacent time slices on the previous m days will be used, represented as follows,

$$\left[\boldsymbol{F}_{d-1}^{t-k}, \boldsymbol{F}_{d-1}^{t+k}\right] \cup \left[\boldsymbol{F}_{d-2}^{t-k}, \boldsymbol{F}_{d-2}^{t+k}\right] \cup \cdots \cup \left[\boldsymbol{F}_{d-m}^{t-k}, \boldsymbol{F}_{d-m}^{t+k}\right] \tag{4}$$

$$\left[\boldsymbol{D}_{d-1}^{t-k}, \boldsymbol{D}_{d-1}^{t+k}\right] \cup \left[\boldsymbol{D}_{d-2}^{t-k}, \boldsymbol{D}_{d-2}^{t+k}\right] \cup \cdots \cup \left[\boldsymbol{D}_{d-m}^{t-k}, \boldsymbol{D}_{d-m}^{t+k}\right] \tag{5}$$

Then the data in each day is compressed as one channel, represented as $[\overline{\boldsymbol{F}}_{d-1}^{t}, \overline{\boldsymbol{F}}_{d-2}^{t}, \ldots, \overline{\boldsymbol{F}}_{d-m}^{t}]$ and $[\overline{\boldsymbol{D}}_{d-1}^{t}, \overline{\boldsymbol{D}}_{d-2}^{t}, \ldots, \overline{\boldsymbol{D}}_{d-m}^{t}]$. When the data of the previous two weeks are used, there will be 14×3 channels.

4.7. Use mask instead of spatial attention

To address the data sparsity, an intuitive solution is by introducing spatial attention, as illustrated in Fig. 7. The attention mechanism was motivated by the visual system of the human. Only parts of the visual field will be paid attention to when observing the surrounding environment, which allows humans to exclude the distractions and focus on what really matters. This mechanism is also useful in this task, since the topology of the road network can be implicitly learned by the model as a spatial attentional matrix. All attention weights are from zero to one, indicating the probability that a specific spatio-temporal cell is covered by a road. However, we should caution the use of spatial attention since it will lead to a sharp increase in the training time.

It should be noted that the inputs contain a number of channels of multiple data sources, including speed, flow, and direction, from various time slices. If the input values of a pixel are all zero at any time slice and any data source, it is highly possible that the corresponding subregion is not covered by the urban road network. Therefore, faced with the challenge of data sparsity, a mask is designed to improve the efficiency of the model. We extend the U-Net with a new branch, which calculates the average for all channels of the inputs and obtains a feature map M of size $H \times W \times 1$. Here, H and W denote the length and width of the input images respectively, and 1 is the number of channels. Pixels that have values (i.e., normalized values of speed, flow, and direction information) greater than zero in the feature map M are numbered as 1, indicating that this pixel is covered by roads in reality. The mask can be obtained as formulated below,

$$mask_{i,j} = \begin{cases} 1, & if \;\; M_{i,j} > 0 \\ 0, & if \;\; M_{i,j} = 0 \end{cases} \tag{6}$$

where $M_{i,j}$ denotes the value at pixel (i,j). The mask implies the topology of the road network and is stacked to the input channels of U-Net.

5. Experiment and result

In many applications of intelligent transportation systems, practitioners usually care more about easily interpretable factors, for example, the traffic density and mean speed in a subregion. Therefore, pixel-wise mean absolute error (MAE) is used to evaluate the performance of our model:

$$MAE = \frac{1}{m} \sum_{i=1}^{m} \left| \boldsymbol{S}^i - \widehat{\boldsymbol{S}}^i \right| \tag{7}$$

where $\boldsymbol{S}^i$ is the observed speed during the i-th time slice, $\widehat{\boldsymbol{S}}^i$ is the predicted values, and m is the total predicted samples.

We set the speeds during 7–9 a.m. and 5–7 p.m. as the targets as the traffic states in peak hours worth more attention for urban traffic

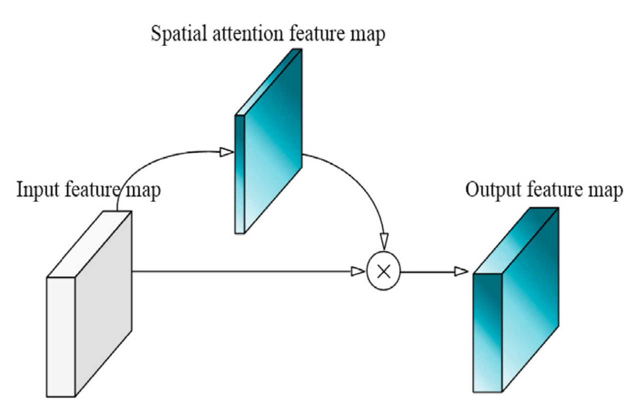

Fig. 7. Illustration of spatial attention.

management. The data in December 2018 is selected as the test set, while the others (from January to November 2018) are used to construct the training set.

Based on the analysis in Section 3.4 (Spatial Propagation of Traffic State), which find that the receptive field of the model needs to be large enough to capture the traffic states in distant subregions. we design a novel framework referring to the structure of U-Net (Ronneberger et al., 2015). Moreover, many improvements are made based on the domain knowledge on the urban transportation network, which are reflected in the following proposed modules, i.e., Temporal Propagation of Traffic State, Periodicity of Traffic System, Traffic State Experience Replay, and Multi-source Data Fusion.

Table 2 details the three datasets we use, where each dataset contains three kinds of traffic data. We use a Min-Max normalization method to scale the traffic data into the range $[-1, 1]$.

In Table 3, we only tested the modules of this framework utilizing the information of speed, since speed is our primary focus, and most existing studies did not incorporate the combinative effects of speed, flow, and direction. Note that the large size of the traffic map in this study results in over 200 thousand pixels for a single time slice. Predicting future traffic state for each pixel using models like LightGBM is hardly applicable, nor is predicting traffic state pixel by pixel using ARIMA-family models. Therefore, the performances of different variants are tested in contrast with the following two baselines, and the results are shown in Table 3. Then, in Table 4, the final model is compared to a state-of-the-art deep learning model.

Historical Data 1: Predict using the historical data at the same time last week.

Historical Data 2: Predict using the historical data at the previous time slice.

The basic model is DeepTSP-*T*, where the suffix '*T*' indicates the usage of the ***Temporal Propagation of Traffic State*** module. When predicting the city-wide speed at time t, the speed snapshots at time $t-1, t-2$, and $t-k$ are included in the input feature map, such that the model can learn the propagation of traffic state along the temporal dimension.

Based on the basic model, DeepTSP-*TP* further added the module ***Periodicity of Traffic System***, which allows for the speeds at the same time of day in the past.

DeepTSP-*TE* added the module ***Experience Replay of Traffic State***, to the basic model. Instead of the periodical data, the historical traffic states in near time slices are input to the model.

Multi-source data are fed into DeepTSP-*TE* to form another variant, DeepTSP-*TEM*. In this model, holidays are represented by an indicator variable where 1 implies it is a holiday. The numbers 0–6 represent the day of week from Monday to Sunday respectively, and the time of day is split into 288 time slices (5 min for each) and represented by the number 0–287. The weather conditions are also labeled according to the local weather reports. In Moscow, there are in total 36 recorded weather conditions, e.g., fog, overcast and chilly, denoted by 0–35. The data is updated every 3 h. For Berlin and Istanbul, the weather information is updated twice an hour and contains 100 types of weather conditions, denoted by 0–99.

The performances of different variants are tested and compared in terms of MAE to determine the efficacy of the designed modules. It can be observed from Table 3 that using the history data alone cannot give satisfactory results. The errors of all proposed model variants (DeepTSP) are significantly lower than that of Historical Data. Hence, in spite of the sparsity of data and the lack of periodicity in traffic time series, we can still effectively predict future traffic states based on the designs proposed in Section 3.

The MAE of the basic model (DeepTSP-*T*) on the dataset of Berlin, Istanbul, and Moscow are 4.491, 3.999, and 6.001 respectively. The introduction of the Periodicity of Traffic System module (DeepTSP-*TP*) slightly improved the results of the basic model on three datasets to 4.461, 3.967, and 5.975. Compared with the periodical information, the replayed experience (DeepTSP-*TE*) shows a larger contribution to the improvement of the basic model, which reaches the MAE of 4.422, 3.918, and 5.935 respectively on the three cities. Finally, we also experimented with auxiliary multi-source data, such as weather conditions, but no significant improvement is observed. It is inferred that this can be attributed to the size of the study area and the limit of auxiliary data sources. For a megacity spreading over thousands of square kilometers, dozens of monitoring stations are often distributed in various locations, since the weather conditions can be completely different in distant parts of a city. Unfortunately, the weather information provided in the dataset is obtained from one of these stations, which is not representative of an entire city. Moreover, the prediction horizon in this task is 5 min. This also requires a prompt update of the weather information. In Moscow, where the data is updated every 3 h, the result of DeepTSP-*TEM* is even the same as that of DeepTSP-*TE*.

Though studies on the spatio-temporal prediction of traffic abound, their application contexts are different from ours. For example, Liang et al. partitioned their study area as a relatively fine-grained grid map of 64×64 pixels. Rather than predicting the future, their objective was to infer the fine-grained crowd flow through super-resolving a given coarse map of 32×32 pixels (Liang et al., 2019). Pan et al. designed a novel meta-knowledge learner, whereby the traffic characteristics of nodes and links can be embedded (Pan et al., 2019). As aforementioned, however, the graph-based models heavily rely on reliable base maps. Considering the size of study areas in our study and the availability of updated maps, one of our primary goals was to relieve the model from the dependency on the topological information of road networks. Therefore, the performance of the proposed model is tested with the following well-known generic models:

ST-ResNet is one of the state-of-the-art models for spatio-temporal prediction (Zhang et al., 2017) and is adopted used in comparison with the proposed framework. It should be stated that the two datasets used by

Table 2
Dataset description.

Dataset	Berlin	Istanbul	Moscow
Traffic data type	Speed, flow, direction	Speed, flow, direction	Speed, flow, direction
Time slice length	5 min	5 min	5 min
Traffic map size	(436,495)	(436,495)	(436,495)
Spatio-temporal cell size	100m $\times$ 100 m	100m $\times$ 100 m	100m $\times$ 100 m
Study area coverage	2158.2 km^2	2158.2 km^2	2158.2 km^2
Weather conditions	100 types	100 types	36 types
Weather update interval	30 min	30 min	3 h

Table 3
Performance comparison of MAE for different models.

Model	Berlin	Istanbul	Moscow
Historical Data 1	6.951	6.255	9.358
Historical Data 2	6.396	5.729	8.466
DeepTSP-*T*	4.491	3.999	6.001
DeepTSP-*TP*	4.461	3.967	5.975
DeepTSP -*TE*	4.422	3.918	5.935
DeepTSP -*TEM*	4.418	3.914	5.935

Table 4
Performance comparison of MAE for different models.

Model	Berlin	Istanbul	Moscow
ST-ResNet (speed, flow, direction)	4.552	3.994	6.032
DeepTSP (speed)	4.418	3.914	5.935
DeepTSP (speed, flow)	4.406	3.891	5.904
DeepTSP (speed, flow, direction)	4.300	3.885	5.897

Table 5
Performance comparison of MAE for different parameters.

Setting	Input channels	Berlin	Istanbul	Moscow
k = 12, m = 7	3 × 12 + 3 × 7	4.316	3.899	5.929
k = 12, m = 14	3 × 12 + 3 × 14	4.300	3.885	5.897
k = 12, m = 21	3 × 12 + 3 × 21	4.311	3.888	5.902
k = 24, m = 14	3 × 24 + 3 × 14	4.304	3.899	5.920

ST-ResNet are split as grid maps of 32 × 32 and 16 × 8 respectively. Moreover, since the resolution is low, there is not any issue relating to sparsity or the lack of apparent periodicity. We also extend the model by introducing features relating to traffic flow and vehicle direction.

Two variants are built based on DeepTSP-*TEM*, which only utilizes the information of speed. The first variant utilizes the information about traffic flow and speed, while the second variant further includes the information of direction.

It is worth noting that the difficulty in this task is much greater than the traditional spatio-temporal prediction of urban traffic due to the three challenges aforementioned. The time series appear to be more like a chaotic sequence since the periodicity can hardly be observed. From Table 4, we can observe that the accuracy improves to 4.406, 3.891, and 5.904 when the flow was incorporated in DeepTSP (speed), i.e., the DeepTSP-*TEM* in Table 3. In addition, the direction information is helpful in assisting traffic state prediction, further improving the MAE to 4.300, 3.885, and 5.897 respectively. ST-ResNet (speed, flow, direction) is also able to accurately predict the traffic state but is constrained by the limited receptive field. When dealing with a large-scale dataset, e.g., the dataset in this study with spatio-temporal data of 436 × 495 pixels, the propagation of traffic state in the spatial dimension cannot be effectively captured without using operations like pooling to enlarge the receptive field.

To predict the speed in at time t of day d, the complete input channels to the DeepTSP can be formulated as:

$$Input = concat\Big(\big[\boldsymbol{S}_d^{t-1}, \boldsymbol{S}_d^{t-2}, \ldots, \boldsymbol{S}_d^{t-k}\big], \big[\boldsymbol{F}_d^{t-1}, \boldsymbol{F}_d^{t-2}, \ldots, \boldsymbol{F}_d^{t-k}\big], \big[\boldsymbol{D}_d^{t-1}, \boldsymbol{D}_d^{t-2}, \ldots, \boldsymbol{D}_d^{t-k}\big], \left[\boldsymbol{S}_{d-1}^{t}, \boldsymbol{S}_{d-2}^{t}, \ldots, \boldsymbol{S}_{d-m}^{t}\right], \left[\boldsymbol{F}_{d-1}^{t}, \boldsymbol{F}_{d-2}^{t}, \ldots, \boldsymbol{F}_{d-m}^{t}\right], \left[\boldsymbol{D}_{d-1}^{t}, \boldsymbol{D}_{d-2}^{t}, \ldots, \boldsymbol{D}_{d-m}^{t}\right]\Big) \quad (8)$$

where $concat()$ represents the concatenation of tensors into a tensor with a larger dimension.

In Table 5, the performances of various hyper-parameter settings (k and m) are presented. Specifically, if the propagation process of traffic states within 1 h is input, the value of k will be 12. If the traffic states within one week are replayed, the value of m will be 7. Larger k and m are not necessarily beneficial to the prediction performance, as the increased number of channels can bring about huge computational pressure, and the additional features can be redundant. After the experiments, the most appropriate hyper-parameters are set as 12 for k and 14 for m.

6. Conclusion

This paper addresses problem of large-scale fine-grained traffic state prediction. A deep learning architecture is proposed to handle the challenge of data scale, granularity, and sparsity. Based on the domain knowledge of traffic engineering, the propagation of traffic state is analyzed, and four modules are specially designed and elaborated in the aspects of the temporal propagation of traffic state, spatial propagation of traffic state, etc. Experiments are performed on the industrial-scale real world probe data in three cities. The experimental results show the effectiveness of our framework in the prediction of large-scale fine-grained traffic states. For future work, integrating prior knowledge from physical traffic flow models, like the LWR model, into the deep learning framework should be further explored. It is worth investigating how the knowledge on these physical laws pertaining to traffic flow can be better incorporated in data-driven models while retaining the computational efficiency.

Declaration of competing interest

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

Acknowledgment

This study is supported by the Distinguished Young Scholar Project (No. 71922007) of the National Natural Science Foundation of China, and supported in part by the Jiangsu Provincial Key Laboratory of Networked Collective Intelligence under Grant BM2017002. This study is part of a project that has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No. 101025896.

References

Ahmed, M.S., Cook, A.R., 1979. Analysis of freeway traffic time-series data by using Box–Jenkins techniques. Transport. Res. Rec. 722, 1–9.
Cai, P., Wang, Y., Lu, G., Chen, P., Ding, C., Sun, J., 2016. A spatiotemporal correlative k-nearest neighbor model for short-term traffic multistep forecasting. Transport. Res. C Emerg. Technol. 62, 21–34.
Castro-Neto, M., Jeong, Y.S., Jeong, M.K., Han, L.D., 2009. Online-SVR for short-term traffic flow prediction under typical and atypical traffic conditions. Expert Syst. Appl. 36 (3), 6164–6173.
Chen, E., Ye, Z., Wang, C., Xu, M., 2019a. Subway passenger flow prediction for special events using smart card data. IEEE Trans. Intell. Transport. Syst. 1–12.
Chen, X., Zhang, S., Li, L., Li, L., 2019b. Adaptive rolling smoothing with heterogeneous data for traffic state estimation and prediction. IEEE Trans. Intell. Transport. Syst. 20 (4), 1247–1258.
Cui, Z., Ke, R., Pu, Z., Ma, X., Wang, Y., 2020. Learning traffic as a graph: a gated graph wavelet recurrent neural network for network-scale traffic prediction. Transport. Res. C Emerg. Technol. 115, 102620.
Dai, X., Fu, R., Zhao, E., Zhang, Z., Lin, Y., Wang, F.Y., Li, L., 2019. DeepTrend 2.0: a light-weighted multi-scale traffic prediction model using detrending. Transport. Res. C Emerg. Technol. 103, 142–157.
Deng, J., Dong, W., Socher, R., Li, L.J., Li, F.F., 2009. ImageNet: A Large-Scale Hierarchical Image Database. IEEE Conference on Computer Vision & Pattern Recognition, Miami, FL, pp. 248–255.
Guo, S., Lin, Y., Li, S., Chen, Z., Wan, H., 2019. Deep spatial-temporal 3D convolutional neural networks for traffic data forecasting. IEEE Trans. Intell. Transport. Syst. 20 (10), 3913–3926.
He, J., Shen, W., Divakaruni, P., Wynter, L., Lawrence, R., 2013. Improving traffic prediction with tweet semantics. In: Proceedings of the Twenty-Third International Joint Conference on Artificial Intelligence, pp. 1387–1393.
Hoque, J.M., Erhardt, G.D., Schmitt, D., Chen, M., Wachs, M., 2021. Estimating the uncertainty of traffic forecasts from their historical accuracy. Transport. Res. Pol. Pract. 147, 339–349.
Ke, J., Yang, H., Zheng, H., Chen, X., Jia, Y., Gong, P., Ye, J., 2018. Hexagon-based convolutional neural network for supply-demand forecasting of ride-sourcing services. IEEE Trans. Intell. Transport. Syst. 1–14.
Ke, J., Zheng, H., Yang, H., Chen, X., 2017. Short-term forecasting of passenger demand under on-demand ride services: a spatio-temporal deep learning approach. Transport. Res. C Emerg. Technol. 85c, 591–608.
Li, W., Wang, J., Fan, R., Zhang, Y., Ban, X.J., 2020. Short-term traffic state prediction from latent structures: accuracy vs. efficiency. Transport. Res. C Emerg. Technol. 111, 72–90.
Li, Y., Yu, R., Shahabi, C., Liu, Y., 2018. Diffusion Convolutional Recurrent Neural Network: Data-Driven Traffic Forecasting, *International Conference on Learning Representations*, pp. 1–16. Vancouver, B.C., Canada.
Liang, Y., Ouyang, K., Jing, L., Ruan, S., Liu, Y., Zhang, J., Rosenblum, D.S., Zheng, Y., 2019. UrbanFM: inferring fine-grained urban flows. In: Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining - KDD '19, pp. 3132–3142.
Liao, B., Zhang, J., Wu, C., McIlwraith, D., Chen, T., Yang, S., Guo, Y., Wu, F., 2018. Deep sequence learning with auxiliary information for traffic prediction. In: The 24th ACM SIGKDD International Conference (London, United Kingdom).
Lin, L., He, Z., Peeta, S., 2018. Predicting station-level hourly demand in a large-scale bike-sharing network: a graph convolutional neural network approach. Transport. Res. C Emerg. Technol. 97, 258–276.
Liu, P., Liao, F., Huang, H., Timmerhans, H., 2015. Dynamic Activity-Travel Assignment in Multi-State Supernetworks, *the 21st International Symposium on Transportation and Traffic Theory*. Kobe, Japan.
Liu, Z., Liu, Y., Meng, Q., Cheng, Q., 2019. A tailored machine learning approach for urban transport network flow estimation. Transport. Res. C Emerg. Technol. 108, 130–150.

Lv, Y., Duan, Y., Kang, W., Li, Z., Wang, F.Y., 2014. Traffic flow prediction with big data: a deep learning approach. IEEE Trans. Intell. Transport. Syst. 16 (2), 865–873.
Ma, D., Song, X., Li, P., 2020. Daily traffic flow forecasting through a contextual convolutional recurrent neural network modeling inter-and intra-day traffic patterns. IEEE Trans. Intell. Transport. Syst. 1–10.
Manley, E., Cheng, T., 2018. Exploring the role of spatial cognition in predicting urban traffic flow through agent-based modelling. Transport. Res. Pol. Pract. 109, 14–23.
Milenković, M., Švadlenka, L., Melichar, V., Bojović, N., Avramović, Z., 2016. SARIMA modelling approach for railway passenger flow forecasting. Transport 1–8.
Pan, S., Hu, R., Fung, S.-F., Long, G., Jiang, J., Zhang, C., 2020. Learning graph embedding with adversarial training methods. IEEE Trans. Cybern. 50 (6), 2475–2487.
Pan, Z., Liang, Y., Wang, W., Yu, Y., Zheng, Y., Zhang, J., 2019. Urban traffic prediction from spatio-temporal data using deep meta learning. In: Proceedinds of the 25th ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD'19). Anchorage, AK, USA, pp. 1720–1730.
Ronneberger, O., Fischer, P., Brox, T., 2015. U-net: convolutional networks for biomedical image segmentation. In: Navab, N., Hornegger, J., Wells, W.M., Frangi, A.F. (Eds.), International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI 2015). Springer International Publishing, Munich, Germany, pp. 234–241.
Shi, R., Mo, Z., Huang, K., Di, X., Du, Q., 2021. Physics-informed Deep Learning for Traffic State Estimation arXiv preprint arXiv:2101.06580.
Sohrabi, S., Paleti, R., Balan, L., Cetin, M., 2020. Real-time prediction of public bike sharing system demand using generalized extreme value count model. Transport. Res. Pol. Pract. 133, 325–336.
Wang, D., Liao, F., Gao, Z., Rasouli, S., Huang, H., 2020. Tolerance-based column generation for boundedly rational dynamic activity-travel assignment in large-scale networks. Transport. Res. E Logist. Transport. Rev. 141, 102034.
Wessel, J., 2020. Using weather forecasts to forecast whether bikes are used. Transport. Res. Pol. Pract. 138, 537–559.
Yao, H., Wu, F., Ke, J., Tang, X., Jia, Y., Lu, S., Gong, P., Ye, J., Li, Z., 2018. Deep multi-view spatial-temporal network for taxi demand prediction. In: *Thirty-Second AAAI Conference on Artificial Intelligence (AAAI-18)*, New Orleans. Louisiana, USA, pp. 2589–2595.
Yu, T., Yin, H., Zhu, Z., 2018. Spatio-temporal graph convolutional networks: a deep learning framework for traffic forecasting. In: Proceedings of the Twenty-Seventh International Joint Conference on Artificial Intelligence, pp. 3634–3640. Stockholm, Sweden.
Zhan, X., Li, R., Ukkusuri, S.V., 2020. Link-based traffic state estimation and prediction for arterial networks using license-plate recognition data. Transport. Res. C Emerg. Technol. 117, 102660.
Zhang, J., Zheng, Y., Qi, D., 2017. Deep Spatio-Temporal Residual Networks for Citywide Crowd Flows Prediction. AAAI Conference on Artificial Intelligence, San Francisco, California, USA, pp. 1655–1661.
Zhang, J., Zheng, Y., Qi, D., Li, R., Yi, X., 2016. DNN-based prediction model for spatio-temporal data. In: Proceedings of the 24th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems - GIS '16. Burlingame, California, pp. 1–4.
Zhang, J., Zheng, Y., Sun, J., Qi, D., 2019a. Flow prediction in spatio-temporal networks based on multitask deep learning. IEEE Trans. Knowl. Data Eng. 1, 1.
Zhang, Z., Li, M., Lin, X., Wang, Y., He, F., 2019b. Multistep speed prediction on traffic networks: a deep learning approach considering spatio-temporal dependencies. Transport. Res. C Emerg. Technol. 105, 297–322.
Zhao, L., Song, Y., Zhang, C., Liu, Y., Li, H., 2019. T-GCN: a temporal graph convolutional network for traffic prediction. IEEE Trans. Intell. Transport. Syst. 1–11.

Yang Liu received his Ph.D. degree in Transportation Engineering from the School of Transportation, Southeast University, Nanjing, China, in 2021. He is currently working at the Department of Architecture and Civil Engineering, Chalmers University of Technology as a Marie Curie Fellow.

His research interest is machine learning and its applications in intelligent transportation systems. He serves as a Young Editor for two top-tier academic journals including, the Innovation (Cell Press' flagship general journal) and IEEE/CAA Journal of Automatica Sinica. Dr. Yang Liu is experienced in the practice of AI techniques and has also won several world prizes in AI competitions organized by leading international AI conferences or research institutes, including the 1st place of KDD Cup, the most well-known algorithm competition in data mining.

Cheng Lyu received his B.E. and M.E. degrees in Transportation Engineering in the School of Transportation at Southeast University, Nanjing, China. He is currently working toward to a Ph.D. degree at Technical University of Munich, Germany. His research interests include transportation big data analysis and modeling, and intelligent transportation systems.

Yuan Zhang received her M.S. degree in Computer Technology in the School of Computer Science and Engineering at Nanjing University of Science and Technology, Nanjing, China. She is now pursuing a Ph.D. degree in Cyber Science and Engineering in the School of Cyber Science and Engineering at Southeast University, Nanjing, China. Her research interests include transportation big data analysis, transportation network planning, and intelligent transportation systems.

Zhiyuan Liu received his Ph.D. degree in Transportation Engineering in 2011 from National University of Singapore. He is currently a professor at School of Transportation in Southeast University, and director of the Research Center for Complex Transport Networks. His research interests include transport network modeling, public transport, and intelligent transport system. In these areas, Dr. Liu has published over 70 journal papers.

Wenwu Yu received the B.Sc. degree in information and computing science and M.Sc. degree in applied mathematics from the Department of Mathematics, Southeast University, Nanjing, China, in 2004 and 2007, respectively, and the Ph.D. degree from the Department of Electronic Engineering, City University of Hong Kong, in 2010. Currently, he is the Founding Director of Laboratory of Cooperative Control of Complex Systems and the Deputy Associate Director of Jiangsu Provincial Key Laboratory of Networked Collective Intelligence, an Associate Dean in the School of Mathematics, and a Full Professor with the Young Endowed Chair Honor in Southeast University, China.

Dr. Yu held several visiting positions in Australia, China, Germany, Italy, the Netherlands, and USA. His research interests include multi-agent systems, complex networks and systems, distributed optimization, smart grids, intelligent transportation systems. Dr. Yu serves as an Editorial Board Member of several flag journals, including IEEE Trans. on Circuits and Systems II, IEEE Trans. on Industrial Informatics, IEEE Trans. on Systems, Man, and Cybernetics: Systems. Science China Information Sciences, Science China Technological Sciences, etc.

He was listed by Clarivate Analytics/Thomson Reuters Highly Cited Researchers in Engineering in 2014–

2018. He publishes about 100 SCI journal papers with more than 10000

citations. He was awarded a National Natural Science Fund for Excellent Young Scholars in 2013, the National Ten Thousand Talent Program for Young Top-notch Talents in 2014, and the Cheung Kong Scholars Programme of China for Young Scholars in 2016. Dr. Yu is also the recipient of the Second Prize of State Natural Science Award of China in 2016.

Xiaobo Qu received the B.Eng. degree from Jilin University, Changchun, China, the M.Eng. degree from Tsinghua University, Beijing, China, and the Ph.D. degree from the National University of Singapore, Singapore. He is currently a Chair Professor of Urban Mobility Systems at Chalmers University of Technology, Gothenburg, Sweden. His research is focused on integrating emerging technologies into urban transport system. He is an elected member of Academia Europaea – the Academy of Europe.