ARCHIVES OF TRANSPORT *ISSN (print):* *0866-9546*
Volume 61, Issue 1, 2022 ***e-ISSN (online):*** *2300-8830*
DOI: *10.5604/01.3001.0015.8148*

SPATIOTEMPORAL ATTENTION MECHANISM-BASED MULTISTEP TRAFFIC VOLUME PREDICTION MODEL FOR HIGHWAY TOLL STATIONS

Zijing HUANG[1], Peiqun LIN[2], Xukun LIN[3], Chuhao ZHOU[4], Tongge HUANG[5]

[1, 2, 3, 4] *School of Civil Engineering and Transportation, South China University of Technology, Guangzhou, China*

[3] *Department of Transportation of Guangdong Guangzhou, China*

[5] *Institute of Software, Chinese Academy of Sciences, Beijing, China*

Abstract:

As the fundamental part of other Intelligent Transportation Systems (ITS) applications, short-term traffic volume prediction plays an important role in various intelligent transportation tasks, such as traffic management, traffic signal control and route planning. Although Neural-network-based traffic prediction methods can produce good results, most of the models can't be explained in an intuitive way. In this paper, we not only proposed a model that increase the short-term prediction accuracy of the traffic volume, but also improved the interpretability of the model by analyzing the internal attention score learnt by the model. we propose a spatiotemporal attention mechanism-based multistep traffic volume prediction model (SAMM). Inside the model, an LSTM-based Encoder-Decoder network with a hybrid attention mechanism is introduced, which consists of spatial attention and temporal attention. In the first level, the local and global spatial attention mechanisms considering the micro traffic evolution and macro pattern similarity, respectively, are applied to capture and amplify the features from the highly correlated entrance stations. In the second level, a temporal attention mechanism is employed to amplify the features from the time steps captured as contributing more to the future exit volume. Considering the time-dependent characteristics and the continuity of the recent evolutionary traffic volume trend, the timestamp features and historical exit volume series of target stations are included as the external inputs. An experiment is conducted using data from the highway toll collection system of Guangdong Province, China. By extracting and analyzing the weights of the spatial and temporal attention layers, the contributions of the intermediate parameters are revealed and explained with knowledge acquired by historical statistics. The results show that the proposed model outperforms the state-of-the-art model by 29.51% in terms of MSE, 13.93% in terms of MAE, and 5.69% in terms of MAPE. The effectiveness of the Encoder-Decoder framework and the attention mechanism are also verified.

Keywords: *ITS, traffic volume forecasting, attention mechanism, highway station, model interpretation*

To cite this article:

Huang, Z., Lin, P., Lin, X., Zhou, C., Huang, T. (2022). Spatiotemporal attention mechanism-based multistep traffic volume prediction model for highway toll stations. Archives of Transport, 61(1), 21-38. DOI: https://doi.org/10.5604/01.3001.0015.8148

Contact:
1) cth_zijing@mail.scut.edu.cn [https://orcid.org/0000-0002-8437-2244]; 2) pqlin@scut.edu.cn [https://orcid.org/0000-0002-1812-0620]; 3) linxukun@126.com; 4) deus0916@outlook.com [https://orcid.org/0000-0001-7406-4187]; 5) tongge@iscas.ac.cn [https://orcid.org/0000-0002-5534-4623]

1. Introduction

1.1. Background

The continuous construction of highway infrastructure has kept the highway mileage growing in China for decades. The country's car ownership is also increasing. As the traffic demand continues to grow, large cities in China and around the world have to deal with the traffic congestion problem, especially for highways as they are supposed to provide a high level of service. Serious traffic congestion problems often occur on highway during important festivals or holidays. To help improve traffic management strategies, it is important to precisely predict traffic volume. Hence, with the development of Intelligent Transportation System (ITS), large amount of reliable travel data has been collected, many data-driven methods have been proposed to solve the traffic volume prediction problem based on data generated from ITS.

1.2. Related Work

Previous research for traffic forecasting models can be classified into two categories: parametric models and nonparametric models.

Parametric models have fixed parameters computed based on prior knowledge of the research field. The autoregressive integrated moving average (ARIMA) model is a typical traffic volume forecasting parametric model. Attempting to make predictions based on the continuity of recent traffic trends, some traffic prediction models using ARIMA have been proposed (Karlaftis & Vlahogianni, 2011; Williams & Hoel, 2003; Giraka & Selvaraj, 2020). ARIMA requires the time series to be highly stationary, while traffic volume series with small time intervals or small geographic coverage often fluctuates. Kalman filtering approach, which is another typical parametric model, is widely applied for real-time short-term traffic flow prediction (Guo et al., 2014; L. Zhang et al., 2011; Zhou et al., 2019) and is especially suitable for handling large-scale traffic flow data. Traffic simulation models are also in the parametric model category, but they require a huge amount of computing power, which is not suitable for dealing with real-time prediction tasks.

Unlike parametric models, non-parametric models are not built based on certain theoretical assumptions, and the parameters are not fixed in these methods. To perform satisfactory outcomes, state-of-the-art non-parametric traffic prediction models tend to be data-driven. The K-nearest neighbour (KNN) model is applied to forecast traffic flow by adding the weighted sum of the volume of the same period of the K days with the closest recent pattern (B. Sun et al., 2018; Z. Wang et al., 2019; D. Xu et al., 2020; Li et al., 2012). Learning algorithms (2020) applied an adaptive KNN method for imputing missing traffic data, which can also be utilized for prediction problems. The support vector machine (SVM), a popular machine learning method for time series prediction, is also widely applied to solve traffic prediction problems (Hong, 2011; Ge et al., 2019; Qiming et al., 2017; Feng et al., 2019). Another typical non-parametric approach employed for traffic prediction is the Bayesian model (Ghosh et al., 2007; J. Wang et al., 2014; S. Sun et al., 2006; Yu & Cho, 2008; Park et al., 2018).

Neural networks, being quite widely used for traffic prediction in recent years, are also in the non-parametric model family. These models are often regarded as more flexible compared to traditional statistical models and outperform classical statistical models in accuracy (Karlaftis & Vlahogianni, 2011). Traffic congestion (2019) concluded that Artificial Neural Networks(ANN) outperformed the other investigated algorithms for predicting traffic conditions. Connor et al. (1994) used a recurrent neural network (RNN) for time series prediction in the early stage. Ma et al. (2015) proposed an architecture using Long Short-Term Memory (LSTM) network, an improved variant of RNN, to capture nonlinear traffic dynamics, especially long temporal dependency. Z. Zhao et al. (2017) proposed a novel traffic forecast model based on an LSTM, where a two-dimensional network composed of many memory units is deployed to consider the temporal-spatial correlation in a traffic system. Ma et al. Ma et al. (2017) converted traffic dynamics to images via a two-dimensional time-space matrix and used a convolutional neural network (CNN) to learn traffic as images. Huang et al. Huang et al. (2014) proposed a deep belief network (DBN) model to learn effective features in an unsupervised fashion. Lv et al. (2015) applied a deep architecture model using stacked auto-encoders (SAEs) as building blocks to represent traffic features.

More recent studies combine different models to capture both spatial and temporal correlation. Luo et al. (2019) used a KNN to select the most related neighbouring stations, capture their spatial features,

and used an LSTM to mine the temporal variability. Feng et al. (2019) incorporated the spatial-temporal correlation information with an adaptive multi-kernel SVM (AMSVM) to predict short-term traffic flows. Predictive analytics (2016) improved SVM by introducing Particle Swarm Optimization (PSO) to find the optimal parameters. J. Xu et al. (2018) proposed a deep learning architecture where a CNN and an LSTM are used to extract the spatial and temporal characteristics, respectively. Lin et al. (2019) employed a sparse autoencoder to extract the spatial features via fully connected layers and used an LSTM to capture the temporal features. Lu et al. (2020) designed a graph LSTM (GLSTM) framework, in which RNN learns temporal relations and graph neural networks (GNNs) integrate graph structured and node-attributed features. Du et al. (2020) deployed a one-dimensional convolutional neural network (ID CNN) to capture the local trend features and used gated recurrent units (GRU) to capture the long temporal dependencies.

Although neural network-based methods have been proven to be promising and robust for traffic prediction (Lv et al., 2015), the internal mechanism of most above models remains unclear like a black box, making them less convincing in presentation. Thus, this is the main gap we try to fill in this paper.

1.3. Contributions

The contributions of this study mainly lie in two aspects: (1) Spatiotemporal attention mechanism is applied to determine the input features' contributions by learning the spatial and temporal correlations. Although spatial and temporal features are commonly used in recent studies, the proposed structure is built with more comprehensive consideration. Through an in-depth inspection of the micro traffic evolution process and the macro traffic evolutionary pattern (see Sections 2.1, 2.3, and 2.4), three forms of correlation mechanism are illustrated, which are the local spatial correlation, the local temporal correlation, and the global spatial correlation, respectively. (2) Distinguished from the related work, the proposed model generates explicit scalar scores that measure the degree of attention that the trained model learnt via the attention mechanism. These scores indicate how much the entrance volume of different stations or different historical periods contributes to making the prediction. The scores can then be verified by being compared with traffic characteristics obtained from historical data (see Section 4.2), or explained with prior knowledge (see Section 4.3). Thus, the proposed model has improved the interpretability of neural-network-based traffic prediction approaches.

2. Correlation mechanism analysis

Considering the micro traffic evolution process, the exit flow of a toll station in a certain period originates from the vehicles that entered the highway network via toll stations in a certain area over a certain period. These vehicles then interact with the adjacent vehicles, together affecting the traffic flow state, such as the speed, volume and density, of the current road section, and consequently affect the total travel time needed to reach the target station. Therefore, it can be deduced that there is causation between the historical entrance volume of certain toll stations and the exit volume of the target toll station.

However, this paper does not attempt to build a model that fits the micro traffic evolution process. The evolution of traffic is a highly complex stochastic process. Micro-traffic simulation models such as the Car-following Model (Yang & Koutsopoulos, 1996; Newell, 2002; X. Zhao & Gao, 2005; Xiao et al., 2020) and the Cell Transmission Model (Huang et al., 2008; Hu, Wang, & Lu, 2010; Hu, Wang, & Sheng, 2010; Ji et al., 2009; Xie et al., 2013) require large numbers of parameters and huge amounts of calculations to obtain results with acceptable accuracy, which is highly time-consuming and therefore impractical for real-time prediction tasks.

Before introducing the model, the correlation mechanism analysis is presented below. The analysis and experiments were applied to the data collected by the highway toll collection system of Guangdong Province. The system has recorded information including the license plate number, the entrance time or the exit time, the road ID and the station ID, of every entrance or exit event within the entire highway network of Guangdong Province. Because of the tolling propose, the collected data is highly precise and reliable. The spatial distributions of all 565 highway toll stations of the highway system are shown in Figure 1. The stations are densely located in the Pearl River Delta Economic Zone, while sparsely located in the less-developed areas. The study area, marked

in red in Figure 1, covering many logistics warehouses and the city airport of the provincial capital, locates in the major transportation hub of the province. The highway network in this area is among the densest in the province.

2.1. Local spatial correlation

Although the exit flow of the target station originates from the historical entrance flow of its upstream stations, only a small proportion of the historical entrance traffic will exit via the target station by observing the historical data. Figure 2 shows the entrance volume of the stations within a range of approximately 60 km from Zhongluotan station (ID: 52-25), and the proportion of the entrance vehicles heading for Zhongluotan station, in September 2019. The maximum proportion shown is only 3%, which might be explained with the complex traffic connections between stations and the small buffer zone of each station.

Figure 2 shows that the stations with the largest proportions of their entrance volume heading for Zhongluotan, marked by the circles with the darkest colour, are located north to the target station. However, the exit volumes contributed from these stations may still be not significant because the scales of their entrance volumes are among the smallest. Meanwhile, the exit volumes contributed by the stations within the area marked by red dashed line may be relatively large because: (1) the scales of their entrance volumes are among the biggest; and (2) the proportions of the entrance vehicles heading for the target station are just slightly smaller than the biggest.

From the demonstration above, it can be concluded that the local spatial correlation between the historical entrance volumes of nearby stations and the future exit volume of the target station is influenced by both the scale of the entrance volume and the proportion of the entrance vehicles heading for the target station.

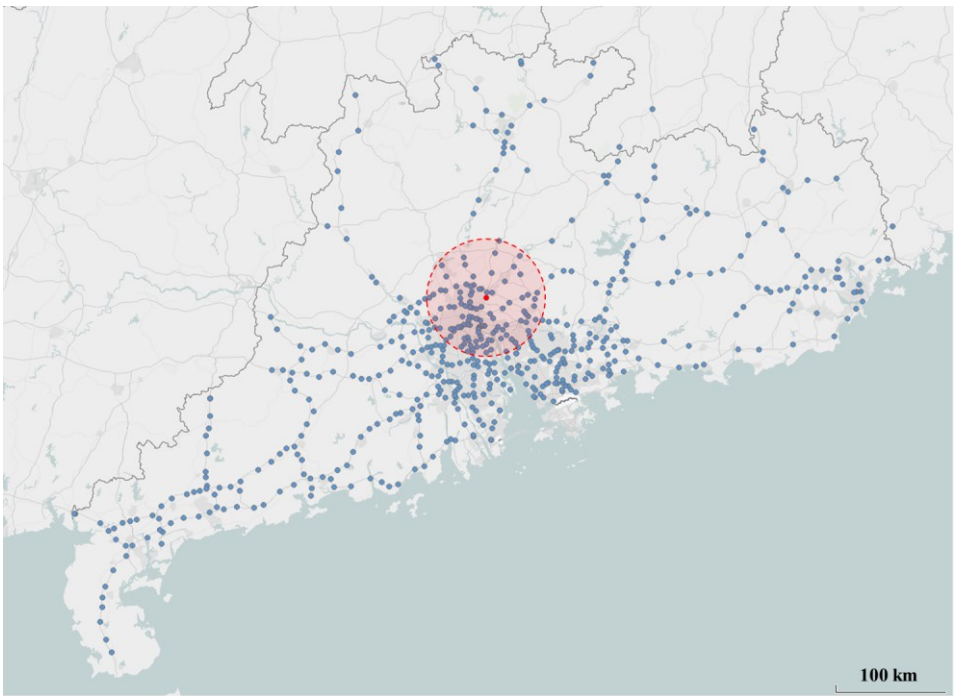

Fig. 1. Spatial distribution of all highway toll stations in Guangdong Province

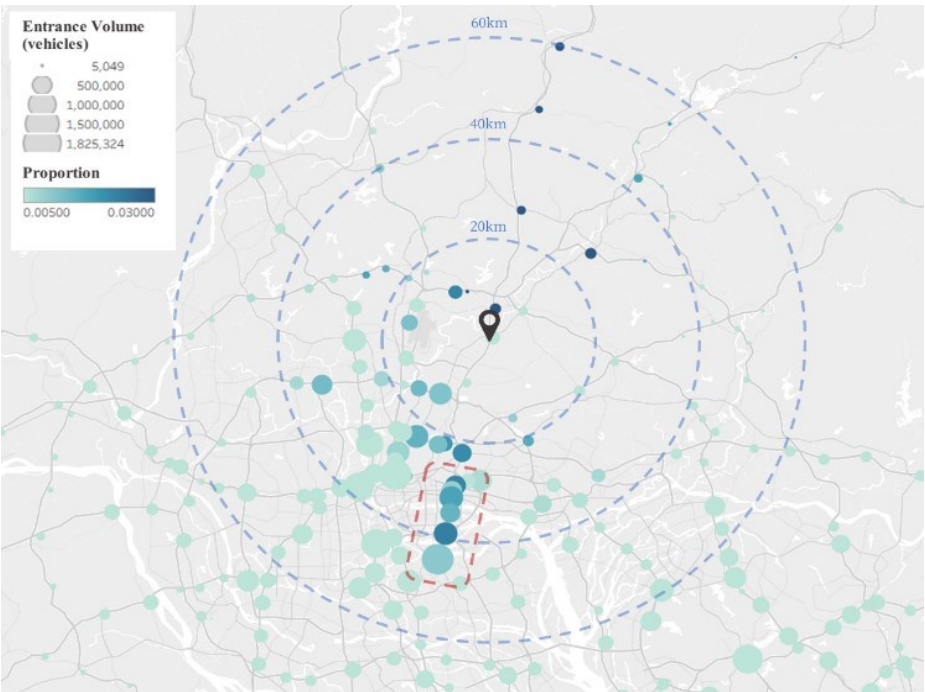

Fig. 2. Demonstration of the factors that affect the local spatial correlation

2.2. Time-varying characteristics of the local spatial correlation

Because of the regularity of working and entertaining behaviours, the traffic demand keeps changing throughout the day, resulting in the fluctuation of the exit volume. By plotting the hourly proportion of vehicles that enter via Tuhua station to those that exit via Zhongluotan station over a 3-day period (Figure 3), a conclusion can be drawn that the fluctuation pattern for each day is quite similar. The 3 days all exhibit steep crests at around 4 o'clock and 7 o'clock, and a mild crest at around 13 o'clock. The proportion continues to gently decrease afterwards until around 23 o'clock, when a slight increase is exhibited. The above results show that the contribution of the entrance station to the target station is time-varying with a certain degree of periodicity.

2.3. Local temporal correlation

The travel time of a highway trip heavily depends on the distance between OD stations. Figure 4 shows the average travel time for trips with the target station as the destination and stations within a certain range as origins. The majority of the cases shown follow the pattern that the longer the distance to the target station is, the longer the mean travel time. However, distance is not linearly associated with travel time. From a microscopic perspective, the travel time of a vehicle is affected by the traffic flow state of all the road sections that it has gone through. The traffic flow state of a road is related to many unobserved factors, including the weather, congestion level, number of lanes, and so on. As the travel distance gets longer, these factors accumulate, causing the uncertainty of each travel to increase, enlarging the dispersion of the travel time distribution. Here, the dispersion level is measured by the standard deviation, which is shown in Figure 4 by the lightness of the colour. As shown in the figure, in general, the dispersion level increases as the entrance station gets farther away from the target station. Then the travel time distribution of vehicles that enters during a certain time period can be converted to the distribution of the arrival volume of the same fleet over time, which is illustrated by Figure 5.

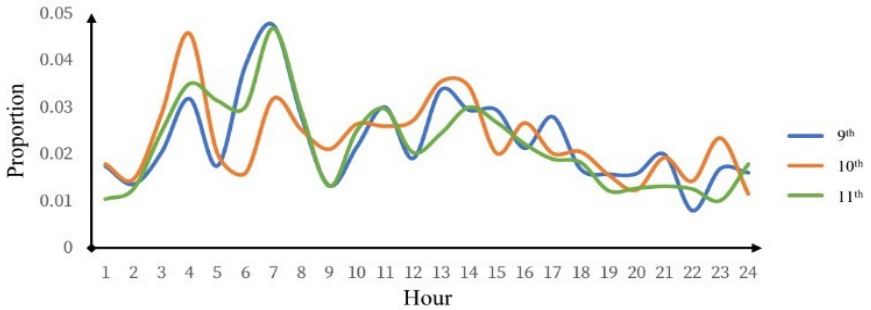

Fig. 3. Hourly proportions of vehicles that enter via Tuhua station to those that exit via Zhongluotan station (Sept. 9 to 11, 2019)

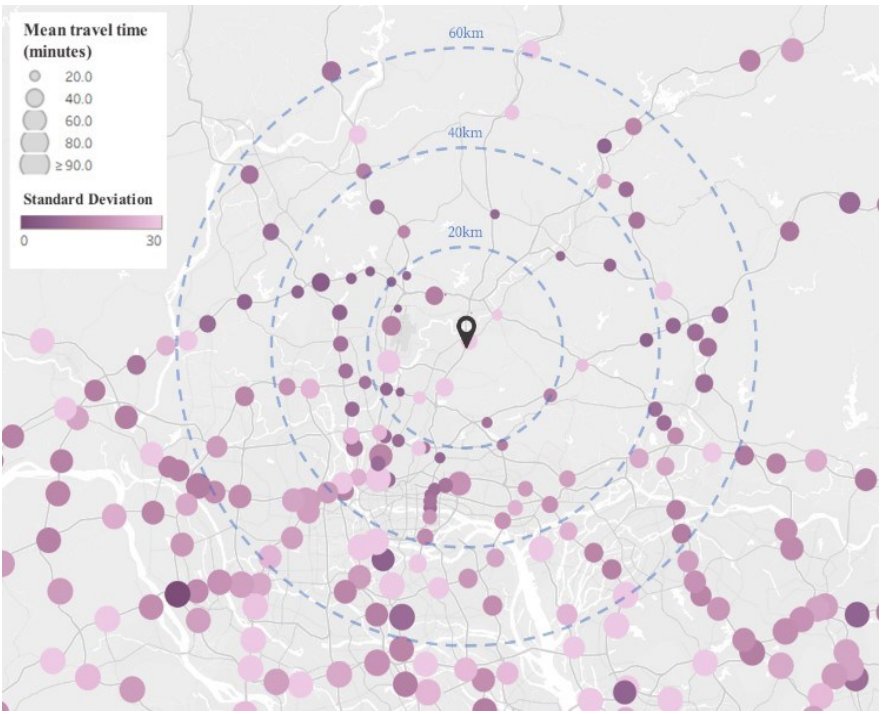

Fig. 4. Means and standard deviations of the travel times of trips with the target station as the destination and the stations within a certain range as the origins

Figure 6 presents a demonstration of how a station's entrance volume in different periods contributes differently to the target station's exit volume in a specific period. Here, the entrance and exit station are the station *S*1 and *ST* in Figure 4, respectively. On the right side of the figure is the arrival volume distribution of the traffic that enters in various time periods. The distributions are the same here, while they may be slightly different in real practice due to changing traffic conditions. To calculate the exit volume of station *ST* in period $T_5 \sim T_6$ is to sum up the areas where the yellow area overlaps each blue

area. To calculate the contribution made by each period's entrance volume is to calculate the proportion of the overlapping area to the blue area. On the left side of the figure, the degree of the contribution that the entrance volume of $S1$ in different periods makes to the exit volume of ST in period $T_5 \sim T_6$ is shown, the redder the circle means the bigger the contribution.

2.4. Global spatial correlation

The local spatial correlation based on the evolution of the traffic flow has been analysed in section 2.1. Meanwhile, another type of correlation may exist if the volume sequence is regarded as a whole feature, representing the recent evolutionary pattern. Here, the Pearson Correlation Coefficient is applied to evaluate the correlation between the hourly entrance and exit volume sequence on the selected day. As shown in Figure 7, almost all exit stations have some highly correlated entrance stations in terms of the similarity of the evolutionary patterns of their volume sequence. Therefore, sequential correlation can also be considered as an important metric for quantifying the spatial correlation.

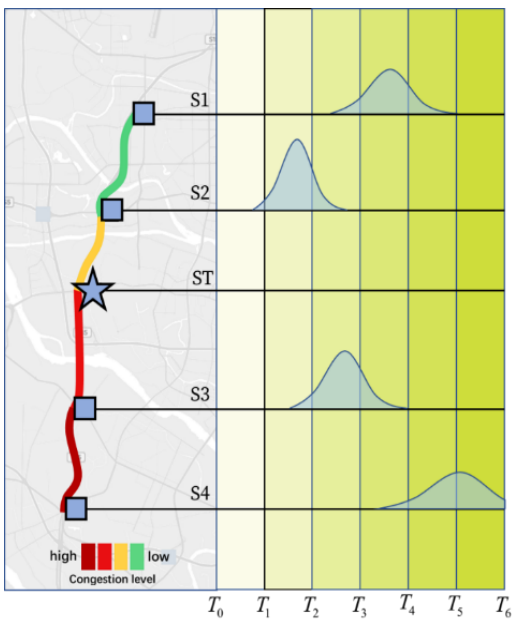

Fig. 5. Distribution of the arrival volume over time (on the map, the star represents the target station, and the square represents entrance station, vehicles enter the highway from T_{-1} to T_0 and exit the highway via the target station ST, the travel time distribution of the entrance volume is affected by factors like path distance and congestion level along the path)

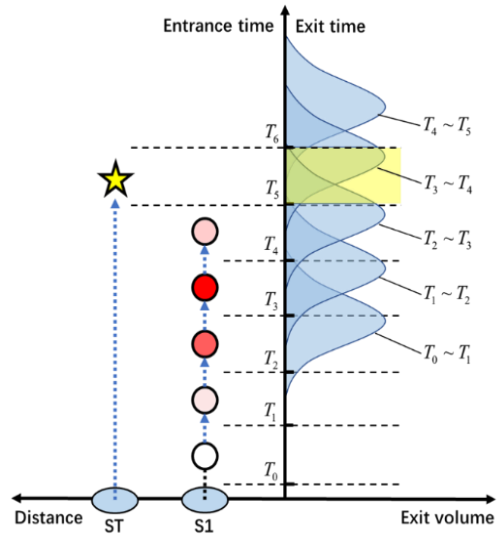

Fig. 6. Illustration that the entrance volumes in different historical time periods have different degrees of contribution to the exit volume

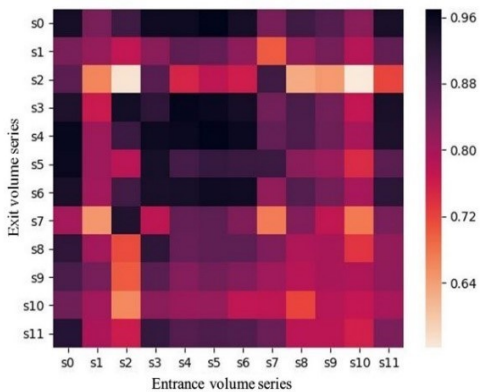

Fig. 7. Means and standard deviations of the travel times of trips with the target station as the destination and the stations within a certain range as the origins

2.5. External Features

Traffic time sequence usually exhibits periodicity to some extent, the volume to be predicted in this paper is not an exception. Figure 8 plots the heatmap of the 15-minute exit volume of Zhongluotan station in Sept. 2019. Each day shares a similar evolutionary pattern, and the crest is less intense on weekends than that on weekdays. Therefore, taking account of the periodicity may help to improve the performance of the model. Because the traffic volume to be predicted is an aggregate measurement resulting from social activities, it is less likely to experience intense fluctuations unless incidents such as infrastructure maintenance or traffic accidents occur. The ups and downs of the volume happen gradually most of the time. As is shown in Figure 8, the recent trend often continues to be exhibited in the following time steps; thus, adding the recent historical exit volume of the target station to the model is also a reasonable approach.

3. Methodology

3.1. Notation and problem statement

Suppose N stations are selected to provide their historical entrance volumes. Given a time window of length T, we use $\mathbf{X} = (\mathbf{x}^1, \mathbf{x}^2, \ldots, \mathbf{x}^N)^\top = (\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_T) \in \mathbb{R}^{N \times T}$ to denote the input series during the past T time steps with a 15-min interval, where $\mathbf{x}^k = (x_1^k, x_2^k, \ldots, x_T^k) \in \mathbb{R}^T$ is the entrance volume series of station k, and $\mathbf{x}_t = (x_t^1, x_t^2, \ldots, x_t^N) \in \mathbb{R}^N$ is the entrance volume vector of all entrance stations at time t. We use $\mathbf{y} = (y_1, y_2, \ldots, y_T) \in \mathbb{R}^T$ to denote the historical exit volume series. Moreover, $\boldsymbol{\omega} = \omega_1, \omega_2, \ldots, \omega_T \in \mathbb{R}^T$ represents the series of 24-hour timestamp, and $\boldsymbol{\psi} = \psi_1, \psi_2, \ldots, \psi_T \in \mathbb{R}^T$ represents the series of day of the week. The proposed model aims to learn a nonlinear mapping to the exit volume of the target stations in the upcoming time window of length U, which we denote as $\hat{\mathbf{y}} = (\hat{y}_1, \hat{y}_2, \ldots, \hat{y}_U) \in \mathbb{R}^U$.

3.2. Model

The framework of the proposed model is shown in Figure 9. To output a fixed-length series at once, an encoder-decoder architecture is used. This architecture was initially widely used in natural language processing studies and was recently proposed for multistep time series prediction by Qin et al. (2017) and Du et al. (2019). The spatial attention is embedded in the encoder to convert the raw input into a vector in which the features from the highly correlated stations are amplified. The output of the encoder will then be sent to the decoder, where temporal attention is embedded to amplify the features from the highly correlated historical time period, forming what is called a context vector. Then, a concatenation of the context vector and the external features are put into an LSTM network, whose final hidden state vector combined with the final context vector will then be linearly transferred to eventually obtain the predicted multi-timestep exit volume of the target station.

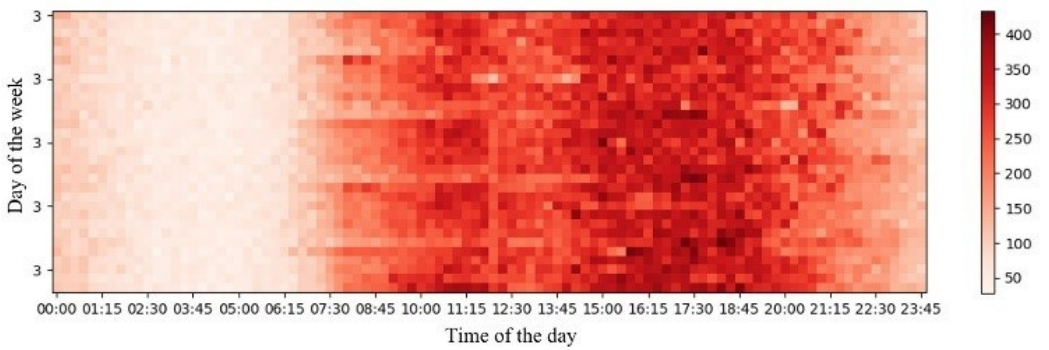

Fig. 8. Heatmap of the 15 min. exit volumes of a station in Sept. 2019

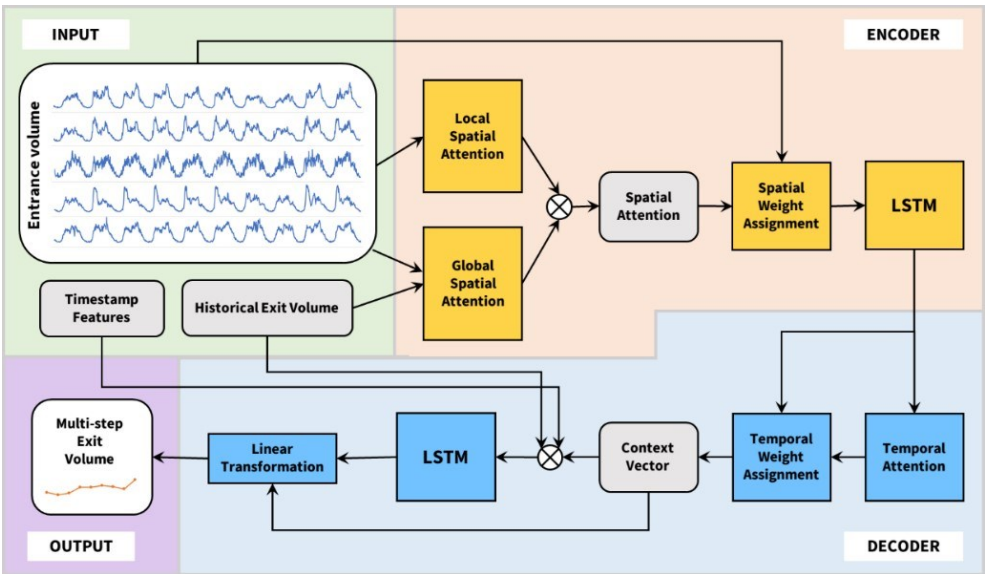

Fig. 9. Structure of the proposed model

The results from section 2.1 and 2.2 show that the local spatial correlation between x^k and y is complex and changing over time with a certain degree of periodicity. To equip the model with the ability to find the highly related entrance station from a large number of input stations, we employ an attention mechanism in the encoder as follows:

$$p_t^k = \mathbf{v}_p^\top \tanh\begin{pmatrix}\mathbf{W}_p\left[\mathbf{h}_{\text{encoder},t-1}; \mathbf{s}_{\text{encoder},t-1}\right] + \\ \mathbf{U}_p[\mathbf{x}^k; \boldsymbol{\omega}_t; \boldsymbol{\psi}_t] + \mathbf{b}_p\end{pmatrix} \quad (1)$$

$$\boldsymbol{\alpha}_t^k = \frac{exp(p_t^k)}{\sum_{i=1}^N exp(p_t^i)} \quad (2)$$

Where $[\cdot;\cdot]$ represents the concatenation operation. The time features are appended to cope with the time-varying characteristics mentioned in section 2.2. $\mathbf{v}_p, \mathbf{b}_p \in \mathbb{R}^T, \mathbf{W}_p \in \mathbb{R}^{T\times 2m}$, and $\mathbf{U}_p \in \mathbb{R}^{3T\times T}$ are the parameters to be learned, where m is the size of the encoder hidden state. $\mathbf{h}_{\text{encoder},\,t-1}$ and $\mathbf{s}_{\text{encoder},\,t-1}$ are the hidden state and cell state of the LSTM network in the previous time step, respectively. The tanh activation function is applied to enable nonlinear transformation. Vector $\mathbf{p}_t$ with N scalars is then normalized with the Softmax function, getting the attention weight vector $\boldsymbol{\alpha}_t = (\alpha_t^1, \alpha_t^2, \ldots, \alpha_t^N)$.

The global correlation between the series of station k's entrance volume and the target station's exit volume is measured by the Pearson Correlation Coefficient (PCC) as follows:

$$\rho_{\mathbf{x}^k,\mathbf{y}} = \frac{\sum_{t=1}^T (x_t^k - \overline{x}^k)(y_t^k - \overline{y}^k)}{\sqrt{\sum_{t=1}^T (x_t^k - \overline{x}^k)^2 \sum_{t=1}^T (y_t^k - \overline{y}^k)^2}} \quad (3)$$

Here, PCC can map the linear correlation of two series to the range of $[-1,1]$, where -1 indicates a totally negative linear correlation, and 1 indicates a totally positive linear correlation.
Both the local and global spatial correlation can affect the exit volume; thus, to jointly consider both of them, a confusion procedure is applied to calculate the spatial attention score:

$$\boldsymbol{\sigma}_t = \mathbf{W}_\sigma[\boldsymbol{\alpha}_t; \boldsymbol{\rho}] + \mathbf{b}_\sigma \quad (4)$$

Where $\boldsymbol{\rho} = (\rho_{\mathbf{x}^1,\mathbf{y}}, \rho_{\mathbf{x}^2,\mathbf{y}}, \ldots, \rho_{\mathbf{x}^N,\mathbf{y}})^\top$, $\boldsymbol{\alpha}_t = (\alpha_t^1, \alpha_t^2, \ldots, \alpha_t^N)^\top$. $\mathbf{W}_\sigma \in \mathbb{R}^{N\times 2N}$ and $\mathbf{b}_\sigma \in \mathbb{R}^N$ are the parameters to be learned. Then the raw input features of time t are adjusted by being multiplied by its corresponding spatial attention score, which is shown as follows:

$$\tilde{\mathbf{p}}_t = (\sigma_t^1 x_t^1, \sigma_t^2 x_t^2, \ldots, \sigma_t^N x_t^N)^\top \quad (5)$$

Where the features from the strongly correlated entrance stations are amplified, while features from the weakly correlated ones are attenuated. Then, the spatially adjusted input features $\tilde{\mathbf{p}} = (\tilde{\mathbf{p}}_1, \tilde{\mathbf{p}}_2, \ldots, \tilde{\mathbf{p}}_t)$ are fed directly into LSTM cell as follows:

$$\mathbf{h}_{\text{encoder},t} = f_e(\mathbf{h}_{\text{encoder},t-1}, \tilde{\mathbf{p}}_{t-1}) \tag{6}$$

Where f_e is the LSTM cell function for the encoder, and $\mathbf{h}_{\text{encoder},t}$ is the hidden state of time t. By concatenating all the hidden state vectors, we obtain $\mathbf{H}_{\text{encoder}} = (\mathbf{h}_{\text{encoder},1}, \mathbf{h}_{\text{encoder},2}, \ldots, \mathbf{h}_{\text{encoder},T}) \in \mathbb{R}^{T\times m}$, which will be the input vectors for the decoder.

To learn the local temporal correlation analysed in section 2.3., we employ an attention mechanism for the time dimension in the decoder so that features from highly correlated historical periods can be given higher attention weights, therefore contributing more to the output $\hat{\mathbf{y}}$. Specifically, the attention weight for each decoder hidden state at time t' is computed by the following:

$$q_{t'}^{t} = \mathbf{v}_q^{\top}\tanh(\mathbf{W}_q'[\mathbf{h}_{\text{decoder},t'-1}; \mathbf{s}_{\text{decoder},t'-1}] + \mathbf{W}_q\mathbf{H}_{\text{encoder},t} + \mathbf{b}_q) \tag{7}$$

$$\beta_{t'}^{t} = \frac{exp(q_{t'}^{t})}{\sum_{j=1}^{T} exp(q_{t'}^{j})} \tag{8}$$

Where $\beta_{t'}^{t}$ is the temporal attention score at time t' for encoder hidden state t. $\mathbf{v}_q$, $\mathbf{b}_q \in \mathbb{R}^m$, $\mathbf{W}_q' \in \mathbb{R}^{m\times m}$, and $\mathbf{W}_q \in \mathbb{R}^{m\times 2n}$ are learnable parameters; n is the size of the decoder hidden state; $\mathbf{h}_{\text{decoder},t'-1}$ and $\mathbf{s}_{\text{decoder},t'-1}$ are the hidden state and cell state of the LSTM network in the previous time step, respectively.

The method used to generate temporal attention weight is similar to that for generating the spatial attention weight, as explained above. Then, each temporal weight is assigned to the corresponding encoder hidden state to calculate the weighted summed vector:

$$\mathbf{c}_{t'} = \sum_{t=1}^{T} \beta_{t'}^{t}\mathbf{h}_{\text{encoder},t} \tag{9}$$

Where $\mathbf{c}_{t'}$ is what we call the context vector at time t' of the decoder, which allows the learned features from stations or time with high correlations to more substantial contributions.

As analysed in section 2.4., time features and the recent historical exit volume of the target station may also be correlated with the output series based on the periodicity characteristics and the continuity of the recent trend. Here, we concatenate these features with the context vector as follows:

$$\tilde{\mathbf{y}}_{t'} = \tilde{\mathbf{w}}^{\top}[\mathbf{y}_{t'}, \boldsymbol{\omega}_{t'}, \boldsymbol{\psi}_{t'}, \mathbf{c}_{t'}] + \tilde{\mathbf{b}} \tag{10}$$

Where $\tilde{\mathbf{w}} \in \mathbb{R}^{U\times(m+3)}$ and $\tilde{\mathbf{b}} \in \mathbb{R}^U$ are the parameters to be learned.

Then, the decoder hidden state of the current time t' can be updated as follows:

$$\mathbf{h}_{\text{decoder},t'} = f_d(\mathbf{h}_{\text{decoder},t'-1}, \tilde{\mathbf{y}}_{t'-1}) \tag{11}$$

Where f_d is an LSTM cell function for the decoder, and $\mathbf{h}_{\text{decoder},t'}$ is the hidden state of time t'.

Considering that the context vector is a key intermediate vector that has been modified through optimization iterations to reveal the features that have been determined to be strongly correlated with the target series, which indicates that it may already be linearly related to $\hat{\mathbf{y}}$. Therefore, in the last layer of the model, we include not only the hidden state of the decoder but also the context vector to generate the prediction output as follows:

$$\hat{y}_{t'} = \mathbf{v}_y^{\top}(\mathbf{W}_m[\mathbf{c}_{t'}; \mathbf{h}_{\text{decoder},t'}] + \mathbf{b}_m) + b_y \tag{12}$$

Where $\hat{y}_{t'}$ is the predicted exit volume of the target station for time t'. $\mathbf{W}_m \in \mathbb{R}^{n\times(m+n)}$, $\mathbf{b}_m \in \mathbb{R}^n$, $\mathbf{v}_y \in \mathbb{R}^{U\times n}$, and $b_y \in \mathbb{R}^U$ are the parameters to be learned. The final output is generated through this linear transformation.

Since the distance between the real series and the predicted series is to be narrowed through optimization iterations, the loss function we use to define the objective of learning is the mean squared error (MSE):

$$L(\boldsymbol{\theta}) = \|\hat{\mathbf{y}} - \mathbf{y}\|_2^2 \tag{13}$$

Where $\boldsymbol{\theta}$ represents the set of learnable parameters in the model, and $\mathbf{y}$ is the vector for the real exit volume.

4. Experiments

4.1. Datasets

The proposed model was applied to the data collected by the highway toll collection system of Guangdong Province as a numerical example. Detailed description of the dataset has been presented in Section 2. We choose 15 minutes as the time interval considering that larger intervals, such as 30 minutes, have been greatly smoothened and would be too easy to predict, while smaller intervals, such as 5 minutes, would fluctuate too much to acquire precise predictions. The time range of the training set and test set are shown in Table 1.

Table 1. The time range of the datasets

Data set	From time	To time	Num. of data points
training set	2019.6.1 00:00:00	2019.8.31 24:00:00	8832
test set	2019.9.1 0:00:00	2019.9.20 24:00:00	1920

Z.-W. Wang (2019) in 2019 shows that 80% of trips take less than 50 minutes on the highway network of Guangdong, which means that in most cases, most exit flow of a station in the next 15 minutes comes from vehicles that have entered during the last 4 periods. However, the travel time will be lengthened during peak periods such as rush hours and toll-free times, when an accurate prediction is more required. Moreover, increasing the length of input series can help better represent the recent evolutionary patterns when calculating the global spatial attention. Therefore, we choose 8 to be the length of the entrance volume series input, which covers the time window of the last 2 hours.

In this experiment, the Zhongluotan station (ID:52-25) is chosen as the target station. Through pre-analysis, we found that the scale of the exit volume of this station is at a medium to high level, and the sources of the exit volume are relatively evenly distributed, making it more difficult for the model to find the most correlated entrance stations. The all 112 stations within a travel distance of 60 km from the target station are selected to provide the recent entrance volume input. For a given target station, the entrance stations that have impact on its exit volume are selected based on travel demand analysis obtained from both historical data and prior knowledge of inter-regional transport connections. In this example, the range is chosen for the following reasons: (1) the average highway travel distance in the core area of the Pearl River Delta is about 45km based on historical data; and (2) considering the close economic relationship and strongly connected highway network between Guangzhou and its adjacent cities, we expand the range to 60km to cover not only the urban and suburban areas of Guangzhou but also a part of Foshan and Dongguan city.

4.2. Spatial attention validation

The evolution of the spatial attention weights during the first 1000 epochs is shown in Figure 10. Due to space limitations, the stations with the weight less than 5e-4 are omitted. The apparent differentiation of the attention weights occurred at about the 50th epoch. From then on, only a few stations were assigned high scores, and others were assigned such low scores that their contributions could be ignored. The weights kept slightly adjusting afterwards.

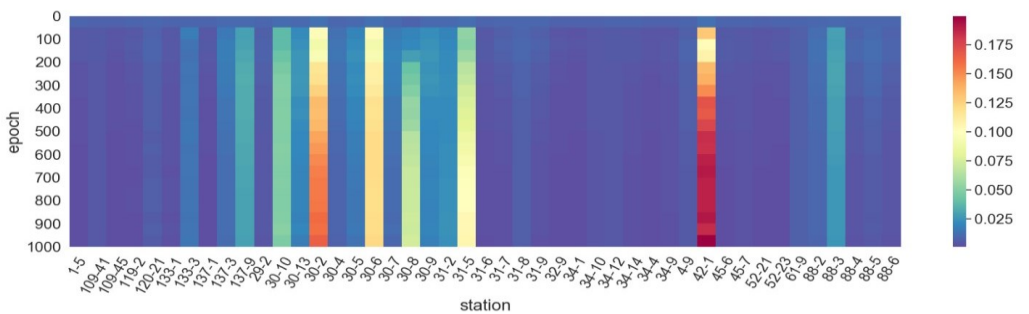

Fig. 10. Evolution of the spatial attention weights during the first 1000 epochs

The spatial attention weights converged at around the 2500th epoch. The training result of the spatial attention weights of all entrance stations are plotted on a map as shown in Figure 11, where the larger the blue "X" mark is the bigger the weight. The target station is marked by a yellow "X", whose size has no meaning.

From Figure 10, we can see that many highly weighted stations are on the road 30, while the most heavily weighted station is 42-1. We find that these stations are all located within the area marked by the red dashed line, which highly overlaps the area we marked in Figure 2 as the potentially biggest contributing area regarding both the scale and the proportion. This consistency proves that the trained result of the spatial attention is highly explainable, and the proposed mechanism can determine the highly correlated stations from many candidates.

4.3. Temporal attention validation

The evolution of the average temporal attention weight over the epochs is shown in Figure 12 and Figure 13 with a heatmap and a line chart, respectively. Unlike the evolution of the spatial attention weight, which showed a clear tendency towards several stations early in the training, the differentiation of the temporal attention weight happened slowly, and the convergence remained unnoticeable until around the 2000th epoch. At the end of the training, the temporal attention weight increased as the historical time step got closer to the current time, which

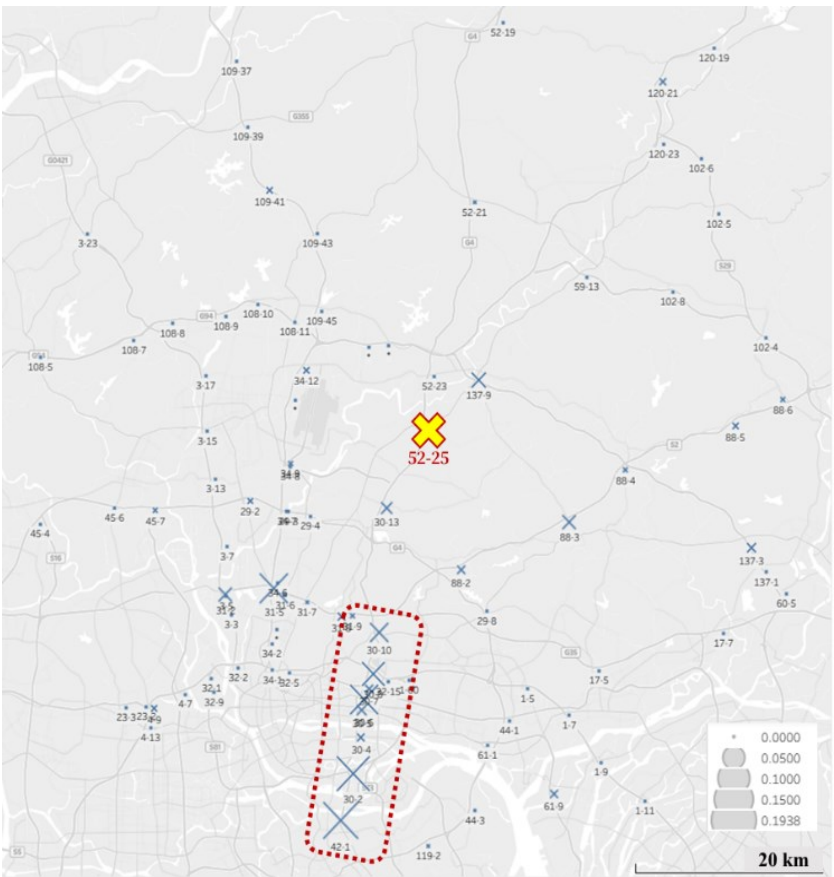

Fig. 11. Training result of the spatial attention weights

might be explained as follows: (1) the historical entrance traffic of the most recent time step were most likely to arrive completely within the forecast period while the entrance traffic of earlier periods from some very close stations had already arrived before the prediction started; (2) the travel time distributions for short-range trips were less dispersed, making it easier to fit a model; and (3) the recent evolutionary trend of the traffic was more likely to continue in the future.

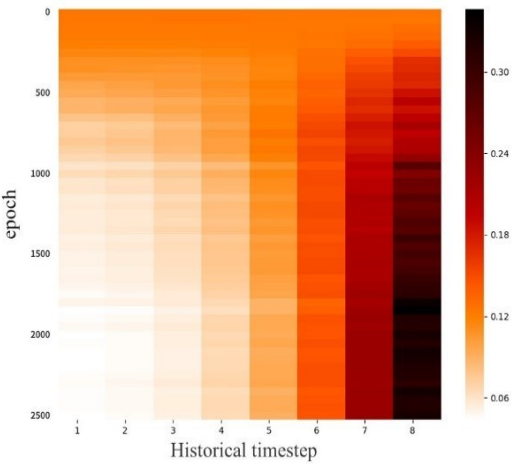

Fig. 12. Heatmap of the evolutionary process of the average temporal attention weight

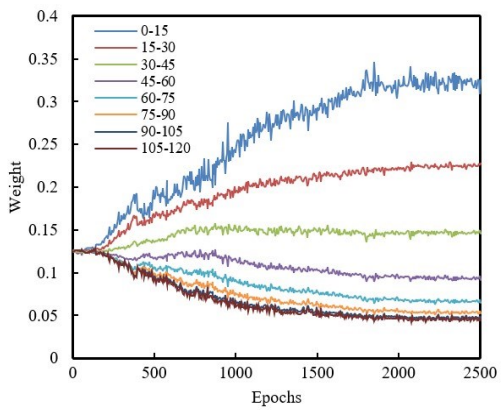

Fig. 13. Line chart for evolving process of average temporal attention weight

4.4. Model comparison

To test the effectiveness of each model component, we compare the proposed model with the following baselines:

(1) AR-LSTM: Only the historical exit volume feature is used as an input to perform an auto regression with an LSTM network.

(2) ED-LSTM: The historical entrance volume feature is included with the input features, and an LSTMembedded Encoder-Decoder without attention mechanism is applied.

(3) ED-Spatial: The spatial attention mechanism is added on the basis of ED-LSTM.

(4) ED-Temporal: The temporal attention mechanism is added on the basis of ED-LSTM.

(5) ST-GCN(Cui et al., 2020): A spatial-temporal graph convolution model based on the spatial method. In the computation of the weighted adjacency matrix, we set σ and ε to 10 and 0.4, respectively. Since we do not have road distance data between stations, we simply set d as the straight-line distance.

(6) ST-ResNet(J. Zhang et al., 2018): The model is widely used in grid-based flow prediction. It learns the spatio-temporal correlations by residual unit. Since our dataset is node-based, we simply reshaped the entrance series of each timestep into a 11 × 11 matrix and fill the last 9 values with 0, then fed the matrix into both the inflow and outflow features. We set p, q, l_c, l_p, l_q to 2, 4, 4, 4, 4, respectively.

To measure the prediction accuracy, we consider 3 evaluation metrics, the mean squared error (MSE), the mean absolute error (MAE) and the mean absolute percentage error (MAPE). The definitions of them are:

$$\mathrm{MSE} = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 \tag{14}$$

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

$$\mathrm{MAPE} = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{y_i - \hat{y}_i}{y_i}\right| \tag{16}$$

Where $\hat{y}_i$ is the predicted value at time i and y_i is the real value at time i. The test results of various models are listed in Table 2.

In terms of the overall 2-hour performance, the accuracy of AR-LSTM is significantly lower than that

of the other models. The MSE of ED-LSTM is less than that of AR-LSTM by 22%, indicating that the historical entrance volume can provide valuable information for predictions. The single additions of spatial attention (ED-Spatial) and temporal attention (ED-Temporal) can help reduce the MSE by 9.55% and 23.86%, respectively, for the EDLSTM. This result demonstrates that by focusing on either the highly correlated spatial or temporal features, the performance can be improved. And temporal attention seems to be more effective. Lastly, the proposed model outperforms all the other models, with a 53.85% and 29.51% lower MSE, a 20.04% and 13.93% lower MAE, and a 14.55% and 5.75% lower MAPE compared with the ED-Spatial and ED-Temporal models, respectively, indicating that the combination of the two levels of attention can help learn the complex correlation mechanism between the input and output more thoroughly. Lastly, the proposed model is also superior to the 2 state-of-the-art models, ST-GCN and ST-ResNet, in terms of all the metrics.

Such improvement in prediction accuracy can help the demand-responsive traffic control approaches to react to the upcoming traffic volume more precisely and effectively. Take adaptive traffic light control as an example, a slight difference of the predicted volume of the incoming lanes will result in a difference of seconds that a traffic light phase will be assigned to.

It can be seen in Table 2 that all metrics at each time step are generally consistent with the overall results described above in terms of their mutual relationship. To further investigate the performance differences at each time step, we visualize the stepwise MSE of each model in Figure 14. Initially, we expect to see the MSE to increase over time for all models, but this only exhibits in the case of AR-LSTM, while the curves of the other models are unexpectedly flat, indicating that the error differences between time steps are small. We interpret this as follows: (1) Causality-based mapping can overcome the problem of lacking information to capture in the long-term for traditional multi-timestep prediction models; (2) The input features from the selected entrance stations and periods can deliver long-term impacts on the exit traffic over the next 2 hours.

In conclusion, the proposed model can learn the complicated and dynamic traffic correlation more precisely and effectively, with a more explainable internal mechanism. And it can achieve a more accurate predicted result of the exit volume in the long term.

Finally, we compare the computation time for different models to test their ability for real-time prediction. Table 3 gives the average computation time of 1000 predictions for various models. The computation time increases as the model gets more complex in structure. For the proposed model, which requires the most computation power, the predicted exit volume of the upcoming 2 hours can be computed in around 140ms, which is fast enough for even extremely real-time prediction tasks.

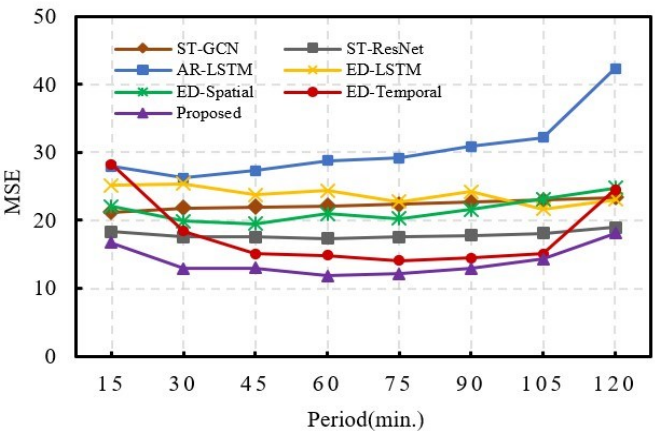

Fig. 14. Stepwise MSE errors of various models

Table 2. The time range of the datasets

Error	Period (min.)	ST-GCN	ST-Res-Net	AR-LSTM	ED-LSTM	ED-Spatial	ED-Temporal	Proposed
MSE	0-120	22.3	17.923	30.596	23.797	21.525	18.12	13.991
	0-15	21.223	18.375	27.892	25.136	22.11	28.341	16.726
	15-30	21.784	17.624	26.206	25.411	19.881	18.404	12.946
	30-45	21.959	17.524	27.293	23.749	19.527	15.158	13.019
	45-60	22.157	17.36	28.802	24.456	20.935	14.827	11.924
	60-75	22.333	17.609	29.152	22.727	20.234	14.084	12.128
	75-90	22.66	17.793	30.895	24.176	21.568	14.469	12.903
	90-105	22.962	18.117	32.151	21.648	23.166	15.093	14.268
	105-120	23.321	18.985	42.374	23.073	24.782	24.583	18.014
MAE	0-120	3.459	3.112	3.755	3.738	3.563	3.31	2.849
	0-15	3.393	3.142	3.579	3.773	3.619	3.96	3.113
	15-30	3.428	3.077	3.516	3.856	3.415	3.189	2.744
	30-45	3.436	3.081	3.568	3.758	3.397	2.989	2.718
	45-60	3.446	3.062	3.664	3.795	3.535	3.016	2.628
	60-75	3.463	3.096	3.606	3.679	3.453	3.073	2.677
	75-90	3.484	3.113	3.83	3.78	3.543	3.141	2.751
	90-105	3.498	3.131	3.885	3.603	3.71	3.141	2.906
	105-120	3.526	3.193	4.396	3.658	3.829	3.975	3.254
MAPE	0-120	6.33%	5.59%	7.85%	6.43%	6.40%	5.80%	5.47%
	0-15	6.06%	5.36%	8.15%	6.29%	6.42%	6.20%	5.75%
	15-30	6.46%	5.63%	7.96%	6.39%	6.30%	5.96%	5.42%
	30-45	6.56%	5.78%	7.78%	6.58%	6.21%	5.76%	5.50%
	45-60	6.38%	5.60%	7.78%	6.54%	6.55%	5.29%	5.28%
	60-75	6.44%	5.66%	7.66%	6.40%	6.40%	5.42%	5.28%
	75-90	6.26%	5.55%	7.65%	6.51%	6.35%	5.35%	5.31%
	90-105	6.34%	5.59%	7.72%	6.25%	6.40%	5.74%	5.43%
	105-120	6.18%	5.55%	8.09%	6.52%	6.56%	6.69%	5.78%

Table 3. Computation time of various models

	AR LSTM	ED LSTM	ED Spatial	ED Temporal	Proposed
average computation time (ms)	97.05	98.44	121.57	115.07	142.74

5. Conclusions

In this paper, we propose a novel spatiotemporal attention mechanism-based deep learning model for forecasting the short-term multistep exit volume of highway toll stations. In the first level, the local and global spatial attention mechanisms considering the micro traffic evolution and macro pattern similarity, respectively, are applied to capture and amplify the features from the highly correlated entrance stations. In the second level, a temporal attention mechanism is employed to amplify the features from the time steps captured as being more influential to the future exit volume. Moreover, with a fusion module, the model considers the effects of the external features, including the timestamp and historical exit volume series of the target station. We extract and visualize the learnt spatial and temporal attention scores, and found that both can be reasonably explained by characteristics acquired from historical data. We evalu-

ate our model on a dataset from the highway toll collection system of Guangdong Province, China. The experiment results show that our model outperforms all the other models in terms of the overall and stepwise metrics (MSE, MAE, and MAPE), and prove that all the proposed components are effective for improving prediction accuracy. Moreover, we test the computation time and prove that the proposed model is fast enough for real-time prediction.

Acknowledgements
This work was supported in part by the National Natural Science Foundation of China under Grant 52072130 and U1811463, in part by the Science and Technology Program Project of Guangdong under Grant 2020A1515010349, and in part by the Science and Technology Program Project of Guangzhou under Grant 201807010008.

References

[1] Connor, J. T., Martin, R. D., & Atlas, L. E. (1994). Recurrent neural networks and robust time series prediction. *IEEE Transactions on Neural Networks*, *5*(2), 240-253. DOI: 10.1109/72.279188

[2] Cui, Z., Henrickson, K., Ke, R., & Wang, Y. (2020). Traffic graph convolutional recurrent neural network: A deep learning framework for network-scale traffic learning and forecasting. *IEEE Transactions on Intelligent Transportation Systems*, *21*(11), 4883-4894. DOI: 10.1109/TITS.2019.2950416

[3] Du, S., Li, T., Gong, X., & Horng, S.-J. (2020). A hybrid method for traffic flow forecasting using multimodal deep learning. *International Journal of Computational Intelligence Systems*, *13*(1), 85-97. DOI: 10.2991/ijcis.d.200120.001

[4] Du, S., Li, T., Yang, Y., Gong, X., & Horng, S.-J. (2019). An lstm based encoder-decoder model for multistep traffic flow prediction. In *2019 international joint conference on neural networks, ijcnn 2019, july 14, 2019 - july 19, 2019* (Vol. 2019-July). Institute of Electrical and Electronics Engineers Inc. DOI: 10.1109/IJCNN.2019.8851928

[5] Feng, X., Ling, X., Zheng, H., Chen, Z., & Xu, Y. (2019). Adaptive multi-kernel svm with spatial-temporal correlation for short-term traffic flow prediction. *IEEE Transactions on Intelligent Transportation Systems*, *20*(6), 2001-2013. DOI: 10.1109/TITS.2018.2854913

[6] Ge, W., Ding, Z., Cao, Y., & Guo, L. (2019). Forecasting model of traffic flow prediction model based on multiresolution svr. In *3rd international conference on innovation in artificial intelligence, iciai 2019, march 15, 2019 - march 18, 2019* (Vol. Part F148152, p. 1-5). Association for Computing Machinery. DOI: 10.1145/3319921.3319923

[7] Ghosh, B., Basu, B., & O'Mahony, M. (2007). Bayesian time-series model for short-term traffic flow forecasting. *Journal of Transportation Engineering*, *133*(3), 180-189. DOI: 10.1061/(ASCE)0733947X(2007)133:3(180)

[8] Giraka, O., & Selvaraj, V. K. (2020). Short-term prediction of intersection turning volume using seasonal arima model. *Transportation Letters*, *12*(7), 483-490. DOI: 10.1080/19427867.2019.1645476

[9] Guo, J., Huang, W., & Williams, B. M. (2014). Adaptive kalman filter approach for stochastic short-term traffic flow rate prediction and uncertainty quantification. *Transportation Research Part C: Emerging Technologies*, *43*, 50-64. DOI: 10.1016/j.trc.2014.02.006

[10] Hong, W.-C. (2011). Traffic flow forecasting by seasonal svr with chaotic simulated annealing algorithm. *Neurocomputing*, *74*(12-13), 2096-2107. DOI: 10.1016/j.neucom.2010.12.032

[11] Hu, X., Wang, W., & Lu, J. (2010). Urban short-term traffic flow forecasting based on the semi-variable cell transmission model. In *Traffic and transportation studies 2010 - proceedings of the 7th international conference on traffic and transportation studies* (Vol. 383, p. 861-871). American Society of Civil Engineers (ASCE). DOI: 10.1061/41123(383)81

[12] Hu, X., Wang, W., & Sheng, H. (2010). Urban traffic flow prediction with variable cell transmission model. *Journal of Transportation Systems Engineering and Information Technology*, *10*(4), 73-78. DOI: 10.1016/S1570-6672(09)60055-6

[13] Huang, W., Shen, F., & Yang, X. (2008). Research on the characteristic and applicability of traffic flow simulation based on ctm. In *8th international conference of chinese logistics and transportation professionals - logistics: The emerging frontiers of transportation and development in china, july 31, 2008 - august 3, 2008*

(p. 1837-1842). ASCE - American Society of Civil Engineers. DOI: 10.1061/40996(330)269

[14]Huang, W., Song, G., Hong, H., & Xie, K. (2014). Deep architecture for traffic flow prediction: Deep belief networks with multitask learning. *IEEE Transactions on Intelligent Transportation Systems*, *15*(5), 2191-2201. DOI: 10.1109/TITS.2014.2311123

[15]Ji, Y., Daamen, W., Zhang, X., & Sun, L. (2009). Traffic incident recovery time prediction model based on cell transmission model. In *2009 12th international ieee conference on intelligent transportation systems, itsc '09, october 3, 2009 - october 7, 2009* (p. 809-812). Institute of Electrical and Electronics Engineers Inc. DOI: 10.1109/ITSC.2009.5309829

[16]Karlaftis, M. G., & Vlahogianni, E. (2011). Statistical methods versus neural networks in transportation research: Differences, similarities and some insights. *Transportation Research Part C: Emerging Technologies*, *19*(3), 387-399. DOI: 10.1016/j.trc.2010.10.004

[17]Kyriakou, K., Lakakis, K., Savvaidis, P., & Basbas, S. (2019). Analysis of spatiotemporal data to predict traffic conditions aiming at a smart navigation system for sustainable urban mobility. *Archives of Transport*, *52*(4), 27 - 46. DOI: 10.5604/01.3001.0014.0206

[18]Li, S., Shen, Z., & Xiong, G. (2012). A k-nearest neighbor locally weighted regression method for short-term traffic flow forecasting. In *2012 15th international ieee conference on intelligent transportation systems, itsc 2012, september 16, 2012 - september 19, 2012* (p. 1596-1601). Institute of Electrical and Electronics Engineers Inc. DOI: 10.1109/ITSC.2012.6338648

[19]Li, T., Yang, Y., Wang, Y., Chen, C., & Yao, J. (2016). Traffic fatalities prediction based on support vector machine. *Archives of Transport*, *39*(3), 21 - 30. DOI: 10.5604/08669546.1225447

[20]Lin, F., Xu, Y., Yang, Y., & Ma, H. (2019). A spatial-temporal hybrid model for short-term traffic prediction. *Mathematical Problems in Engineering*, *2019*, 1V. DOI: 10.1155/2019/4858546

[21]Lu, Z., Lv, W., Cao, Y., Xie, Z., Peng, H., & Du, B. (2020). Lstm variants meet graph neural networks for road speed prediction. *Neurocomputing*, *400*, 34-45. DOI: 10.1016/j.neucom.2020.03.031

[22]Luo, X., Li, D., Yang, Y., & Zhang, S. (2019). Spatiotemporal traffic flow prediction with knn and lstm. *Journal of Advanced Transportation*, *2019*. DOI: 10.1155/2019/4145353

[23]Lv, Y., Duan, Y., Kang, W., Li, Z., & Wang, F.-Y. (2015). Traffic flow prediction with big data: A deep learning approach. *IEEE Transactions on Intelligent Transportation Systems*, *16*(2), 865-873. DOI: 10.1109/TITS.2014.2345663

[24]Ma, X., Dai, Z., He, Z., Ma, J., Wang, Y., & Wang, Y. (2017). Learning traffic as images: A deep convolutional neural network for large-scale transportation network speed prediction. *Sensors (Switzerland)*, *17*(4). DOI: 10.3390/s17040818

[25]Ma, X., Tao, Z., Wang, Y., Yu, H., & Wang, Y. (2015). Long short-term memory neural network for traffic speed prediction using remote microwave sensor data. *Transportation Research Part C: Emerging Technologies*, *54*, 187-197. DOI: 10.1016/j.trc.2015.03.014

[26]Newell, G. F. (2002). A simplified car-following theory: A lower order model. *Transportation Research Part B: Methodological*, *36*(3), 195-205. DOI: 10.1016/S0191-2615(00)00044-8

[27]Park, H.-C., Kim, D.-K., & Kho, S.-Y. (2018). Bayesian network for freeway traffic state prediction. *Transportation Research Record*, *2672*(45), 124-135. DOI: 10.1177/0361198118786824

[28]Qi-ming, W., Ai-wan, F., & He-sheng, S. (2017). Network traffic prediction based on improved support vector machine. *International Journal of System Assurance Engineering and Management*, *8*(3), S1976-S1980. DOI: 10.1007/s13198-016-0412-8

[29]Qin, Y., Song, D., Cheng, H., Cheng, W., Jiang, G., & Cottrell, G. W. (2017). A dual-stage attention-based recurrent neural network for time series prediction. In *26th international joint conference on artificial intelligence, ijcai 2017, august 19, 2017 - august 25, 2017* (Vol. 0, p. 2627-2633). International Joint Conferences on Artificial Intelligence.

[30]Sun, B., Cheng, W., Goswami, P., & Bai, G. (2018). Short-term traffic forecasting using self-

adjusting k-nearest neighbours. *IET Intelligent Transport Systems*, *12*(1), 41-48. DOI: 10.1049/iet-its.2016.0263

[31] Sun, S., Zhang, C., & Yu, G. (2006). A bayesian network approach to traffic flow forecasting. *IEEE Transactions on Intelligent Transportation Systems*, *7*(1), 124-133. DOI: 10.1109/TITS.2006.869623

[32] Wang, J., Deng, W., & Guo, Y. (2014). New bayesian combination method for short-term traffic flow forecasting. *Transportation Research Part C: Emerging Technologies*, *43*, 79-94. DOI: 10.1016/j.trc.2014.02.005

[33] Wang, Y., Xiao, Y., Lai, J., & Chen, Y. (2020). An adaptive k nearest neighbour method for imputation of missing traffic data based on two similarity metrics. *Archives of Transport*, *54*(2), 59 - 73. DOI: 10.5604/01.3001.0014.2968

[34] Wang, Z., Ji, S., & Yu, B. (2019). Short-term traffic volume forecasting with asymmetric loss based on enhanced knn method. *Mathematical Problems in Engineering*, *2019*. DOI: 10.1155/2019/4589437

[35] Wang, Z.-W. (2019). The trip characteristic analysis of guangdong highway based on network toll collection. *Journal of Guangdong Communication Polytechnic*, *18*(03), 20-25.

[36] Williams, B. M., & Hoel, L. A. (2003). Modeling and forecasting vehicular traffic flow as a seasonal arima process: Theoretical basis and empirical results. *Journal of Transportation Engineering*, *129*(6), 664-672. DOI: 10.1061/(ASCE)0733947X(2003)129:6(664)

[37] Xiao, X., Duan, H., & Wen, J. (2020). A novel car-following inertia gray model and its application in forecasting short-term traffic flow. *Applied Mathematical Modelling*, *87*, 546-570. DOI: 10.1016/j.apm.2020.06.020

[38] Xie, B., Xu, M., Harri, J., & Chen, Y. (2013). A traffic light extension to cell transmission model for estimating urban traffic jam. In *2013 ieee 24th annual international symposium on personal, indoor, and mobile radio communications, pimrc 2013, september 8, 2013 - september 11, 2013* (p. 2566-2570). Institute of Electrical and Electronics Engineers Inc. DOI: 10.1109/PIMRC.2013.6666579

[39] Xu, D., Wang, Y., Peng, P., Beilun, S., Deng, Z., & Guo, H. (2020). Real-time road traffic state prediction based on kernel-knn. *Transportmetrica A: Transport Science*, *16*(1), 104-118. DOI: 10.1080/23249935.2018.1491073

[40] Xu, J., Zhang, Y., Jia, Y., & Xing, C. (2018). An efficient traffic prediction model using deep spatial-temporal network. In *14th eai international conference on collaborative computing: Networking, applications and worksharing, collaboratecom 2018, december 1, 2018 - december 3, 2018* (Vol. 268, p. 386-399). Springer Verlag. DOI: 10.1007/978-3030-12981-1 27

[41] Yang, Q., & Koutsopoulos, H. N. (1996). Microscopic traffic simulator for evaluation of dynamic traffic management systems. *Transportation Research Part C: Emerging Technologies*, *4*(3), [d]113-129. DOI: 10.1016/S0968-090X(96)00006-X

[42] Yu, Y. J., & Cho, M.-G. (2008). A short-term prediction model for forecasting traffic information using bayesian network. In *3rd international conference on convergence and hybrid information technology, iccit 2008, november 11, 2008 - november 13, 2008* (Vol. 1, p. 242-247). IEEE Computer Society. DOI: 10.1109/ICCIT.2008.355

[43] Zhang, J., Zheng, Y., Qi, D., Li, R., Yi, X., & Li, T. (2018). Predicting citywide crowd flows using deep spatio-temporal residual networks. *Artificial Intelligence*, *259*, 147-166. DOI: https://doi.org/10.1016/j.artint.2018.03.002

[44] Zhang, L., Sun, Y., & Ma, J. (2011). An adaptive kalman filter for short-term traffic flow forecasting.In *Icte 2011* (p. 97-102). DOI: 10.1061/41184(419)17

[45] Zhao, X., & Gao, Z. (2005). A new car-following model: Full velocity and acceleration difference model. *European Physical Journal B*, *47*(1), 145-150. DOI: 10.1140/epjb/e2005-00304-3

[46] Zhao, Z., Chen, W., Wu, X., Chen, P. C. Y., & Liu, J. (2017). Lstm network: A deep learning approach for short-term traffic forecast. *IET Intelligent Transport Systems*, *11*(2), 68-75. DOI: 10.1049/iet-its.2016.0208

[47] Zhou, T., Jiang, D., Lin, Z., Han, G., Xu, X., & Qin, J. (2019). Hybrid dual kalman filtering model for short-term traffic flow forecasting. *IET Intelligent Transport Systems*, *13*(6), 1023-1032. DOI: 10.1049/iet-its.2018.5385