

# AST-GCN: Attribute-Augmented Spatiotemporal Graph Convolutional Network for Traffic Forecasting

Jiawei Zhu, Chao Tao, Hanhan Deng, Ling Zhao, Pu Wang, *Member, IEEE*, Tao Lin, Haifeng Li\*, *Member, IEEE*,

**Abstract**—Traffic forecasting is a fundamental and challenging task in the field of intelligent transportation. Accurate forecasting not only depends on the historical traffic flow information but also needs to consider the influence of a variety of external factors, such as weather conditions and surrounding POI distribution. Recently, spatiotemporal models integrating graph convolutional networks and recurrent neural networks have become traffic forecasting research hotspots and have made significant progress. However, few works integrate external factors. Therefore, based on the assumption that introducing external factors can enhance the spatiotemporal accuracy in predicting traffic and improving interpretability, we propose an attribute-augmented spatiotemporal graph convolutional network (AST-GCN). We model the external factors as dynamic attributes and static attributes and design an attribute-augmented unit to encode and integrate those factors into the spatiotemporal graph convolution model. Experiments on real datasets show the effectiveness of considering external information on traffic forecasting tasks when compared to traditional traffic prediction methods. Moreover, under different attribute-augmented schemes and prediction horizon settings, the forecasting accuracy of the AST-GCN is higher than that of the baselines.

**Index Terms**—traffic forecasting, graph convolutional network, external factors, spatiotemporal models.

## 1 INTRODUCTION

As one of the essential components in intelligent transportation systems (ITSs), traffic forecasting can provide a scientific basis for the management and planning of urban transportation systems[1–3]. According to predicted traffic states, transportation departments can deploy and guide traffic flows in advance, thereby improving the operating efficiency of road networks and alleviating traffic jams[4, 5].

The starting point of traditional traffic forecasting methods is generally to learn historical traffic characteristics to predict traffic at future moments[6]. However, it is difficult to achieve accurate traffic forecasting because future traffic states not only depend on historical states but can also be affected by a variety of static and dynamic external factors. Among them, the static factors will not change continuously over time, but the traffic state at a certain time will be affected. For example, on a road section with a large number of restaurants, the traffic state during the dining period will be significantly different from other periods; the latter will change over time and lead to changes in traffic conditions. Taking weather conditions as an example, as the

weather shifts from sunny to heavy rain, the traffic speed will generally decrease. These factors create randomness in traffic states and make it challenging for accurate traffic forecasting.

Aiming at the problem that traditional models cannot comprehensively consider factors affecting traffic conditions, we propose an attribute-augmented spatiotemporal graph convolutional model (AST-GCN). We consider the external factors as attributes of road sections in the road network and model the attributes and traffic features of road sections simultaneously to obtain the augmented feature vectors. By this means, the model's perception of the external information is enhanced, thereby improving forecasting accuracy.

The main contributions of this article are as follows:

(1) This paper proposes a novel traffic prediction model AST-GCN that can capture external information in combination with the spatiotemporal graph convolution model.

(2) The proposed AST-GCN model can integrate both dynamic and static external information related to the road, such as weather and surrounding POIs.;

(3) We evaluate the model using real data, and the experimental results show that the prediction results of the AST-GCN model outperform the baselines, indicating the effectiveness of modeling external information.

The remaining sections of this article are organized as follows. The second section reviews related works and development trends of traffic forecasting. Section 3 introduces the details of our method. In section 4, experiments are conducted on real-world data to evaluate the performance of the proposed method compared with baselines, and a perturbation analysis is carried out to test the robustness of

- H. Li, J. Zhu, C. Tao and L. Zhao are with School of Geosciences and Info-Physics, Central South University, Changsha 410083, Hunan, China. (email: lihaifeng@csu.edu.cn)
- H. Deng is with Hisense TransTech Co., Ltd., No.17 Donghai West Road, Qingdao, China
- P. Wang is with School of Traffic Transportation Engineering, Central South University, Changsha 410083, China.
- T. Lin is with College of Biosystems Engineering and Food Science, Zhejiang University, Hangzhou, China.

Corresponding author: Haifeng Li.  
©2020 IEEE

our model. The summary and conclusion of future works are given in section 5.

## 2 RELATED WORKS

Traffic forecasting is an essential part of ITS and plays an important role in urban traffic control and development. The traffic forecasting methods have undergone different stages of evolution. The traditional analysis is mainly based on mathematical statistics to predict traffic states at the beginning. Among them, the principle of the historical average model (HAM) uses historical average data as the prediction result, which is simple to calculate but has low prediction accuracy[7]. Time series models such as ARMA[8] and its variants[9, 10] utilize the relationship between current data and historical data for forecasting and perform modeling and analysis considering the periodicity and trend of data. However, they are based on the time series stability assumption and are thus unable to capture traffic flow mutations. Later, traffic forecasting models based on machine learning emerged. The k-nearest neighbor algorithm (KNN) was first applied to the prediction of traffic flow [11], followed by studies using the Bayesian inference method[12], and support vector machine (SVM)[13].

These algorithms can model more complex characteristics of traffic flows but have limited ability to capture nonlinear patterns. In recent years, deep learning has attracted the attention of researchers due to its advantages in capturing nonlinear and complex patterns. Many deep learning methods have been applied to traffic forecasting, such as deep confidence networks (DBNs)[14, 15] and stacked autoencoding neural networks (SAEs)[16]. One of the disadvantages of these methods is that they independently process the traffic flow information at each time and do not directly model the dependencies in traffic flows in the time series. Therefore, using recurrent neural networks (RNNs) based on sequence prediction to predict traffic flow was studied. However, recurrent neural networks suffer from short-term memory due to the vanishing gradient during backpropagation. Therefore, researchers turned to long short-term memory networks (LSTMs)[17, 18] and gated recurrent units (GRUs)[19, 20], which were created as solutions to short-term memory to extract the temporal dependencies of traffic flow data.

Although these models can capture temporal dependencies in traffic flows, researchers have gradually recognized the importance of spatial dependencies and have made improvements by introducing convolutional neural networks (CNNs) to extract spatial information and combining them with LSTMs[21, 22], which has improved prediction accuracy. Since CNNs were designed for Euclidean space, such as images and grids, they have limitations in transportation networks with non-Euclidean topology and thus cannot essentially characterize the spatial dependence of traffic flows. Emerging graph convolutional neural networks (GCNs) are dedicated to processing network structures[23, 24], which can better model the spatial dependence of road segments on traffic networks[25, 26].

However, the traffic prediction task not only relies on historical traffic information and spatial relationships but is also affected by a variety of external factors, such as weather

conditions and the distribution of surrounding POIs. How to integrate the information about external influence factors in the model is the main problem of the current traffic task. In previous studies, there have been considerations for multisource data. For example, Liao et al.[27] integrated an encoder based on LSTM [28] to encode external information and model multimodal data as a sequence input. The model proposed by Zhang et al.[20] is mainly based on the GRU model[19] and realizes the traffic forecasting task with external weather information based on the feature fusion of the input features and weather information.

In summary, the existing methods do not fully consider the impact of internal and external factors simultaneously. How to combine multisource data to realize the task of traffic prediction is an urgent problem to be solved. Therefore, this paper proposes an attribute-augmented spatiotemporal graph convolutional model (AST-GCN) for traffic forecasting, which regards external factors as attribute information of the road segments in the road network. The attribute information and traffic characteristics are then integrated to enhance the model's perception of external information, thereby improving traffic forecasting accuracy.

## 3 METHOD

### 3.1 Problem Definition

The goal of traffic forecasting is to predict future traffic states based on historical states and auxiliary information. While the traffic states of road sections are mainly described by the average traffic volume, speed, and occupancy rate, we take the average traffic speed as an example to illustrate our work. Therefore, the traffic forecasting task of this paper is mainly based on the traffic speeds in the past period of time and external factors that affect the traffic to predict the traffic speed in the future.

**Definition 1:** Road network  $G$ . We use a road network  $G = (V, E)$  to represent the connection relationship between road sections.  $V = v_1, v_2, \dots, v_n$  represents the collection of road sections,  $n$  is the number of road sections,  $E = e_1, e_2, \dots, e_m$  is the collection of edges that indicate the connectivity between two road segments, and  $m$  represents the number of edges. Without loss of generality, the adjacency matrix  $A$  is used to illustrate the connectivity of the road network.  $A$  is a matrix composed of 0 and 1 when  $G$  is an unweighted network, where 1 suggests that corresponding road segments are connected, and 0 otherwise.

**Definition 2:** Traffic feature matrix  $X$ . In this paper, the traffic speed is regarded as an inherent attribute of each node on the urban road network, represented by a matrix  $X$ , while  $x_i^t$  represents the traffic speed on the  $i$ -th road section at time  $t$ .

**Definition 3:** Attribute matrix  $K$ . In this paper, the external factors that affect traffic conditions are regarded as the auxiliary attributes of the road segments on the urban road network. These factors can form an attribute matrix  $K = \{K_1, K_2, \dots, K_l\}$ , where  $l$  is the category number of auxiliary information. The set of auxiliary information of type  $j$  is represented as  $K_j = \{j^1, j^2, \dots, j^t\}$ , and  $j_i^t$  is the  $j$ -th auxiliary information of the  $i$ th road section at time  $t$ .

In summary, the traffic prediction problem can be understood as learning the function  $f$  on the basis of the basic

topology  $G$ , feature matrix  $X$  and attribute matrix  $K$  of the road network to obtain traffic information in the future period  $T$ , as shown in Eq.1:

$$[x_{t+1}, x_{t+2}, \dots, x_{t+T}] = f(G, X | K) \quad (1)$$

### 3.2 External factors

Without loss of generality, this article analyzes the impact of external factors on traffic states from static and dynamic perspectives:

(1) Static factors, which mainly refer to static geographic information that does not change with time but still exerts influence on traffic states. For example, the distribution of POIs around a road section can determine the visiting pattern of people and the attractiveness of the road section, which is reflected by its traffic states in return. Fig.1-(a) shows the distribution of three different kinds of POIs (catering, transportation, and accommodation) around two road sections. Fig. 1-b shows the variation in traffic speed on two road sections on the same day. Road 1, with more restaurants and fewer traffic facilities, has a higher traffic speed than road 2 between 12:00-20:00, which validates that the difference in the distribution of POIs will lead to a difference in traffic states.

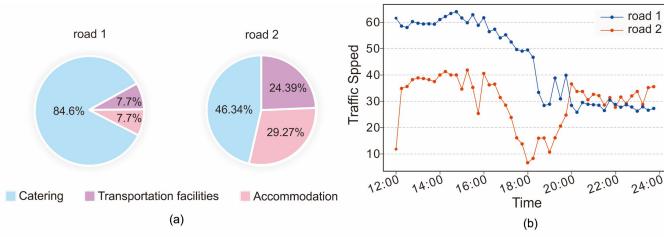


Fig. 1. Illustration of the influence of POIs on traffic states.

(2) Dynamic factors, such as the weather condition, which is a time-varying determinant of road conditions and visions while driving, can directly affect the traffic states. As an illustration of the influence of weather conditions, Fig.2 shows the change in traffic speed over time of a certain road section on separate days with different weather conditions. Specifically, one day has heavy rain from 16 o'clock to 18 o'clock and then turns into light rain; the other day is sunny all day. It can be seen that the traffic speed decreased greatly during the rainy hours compared to the counterpart on a sunny day, which indicates the great impact of weather on traffic states.

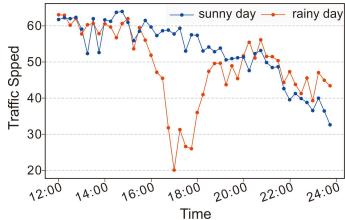


Fig. 2. Illustration of the influence of weather conditions on traffic states.

### 3.3 Attribute augmentation unit

To comprehensively consider factors that affect the traffic states, this study models external factors as the dynamic ( $D$ ) and static ( $S$ ) attributes of the road segments in the road network. Then, the traffic feature matrix  $X$  and the attribute matrix  $K = \{S, D\}$  are fed into the attribute augmentation unit (A-Cell) to derive augmented matrices.

(1) Incorporating static attributes  $S$ .

$S \in R^{n \times p}$  is a collection of  $p$  different static attributes  $\{\vec{s}_1, \vec{s}_2, \dots, \vec{s}_p\}$ . Since the attribute values do not vary with time, the whole matrix  $S$  is constantly used, while only the corresponding column of feature matrix  $X$  is extracted in the process of generating the augmented matrix at each timestamp. Specifically, the matrix augmented by static attributes at time  $t$  is formed as:

$$E_s^t = [X^t, S], E_s^t \in R^{n \times (p+l)} \quad (2)$$

(2) Incorporating dynamic attributes  $D$ .

Different from  $S$ ,  $D \in R^{n \times (w*t)}$  is a collection of  $w$  different dynamic attributes  $\{D_1, D_2, \dots, D_w\}$ . Notably, considering that the traffic states are subject to the cumulative effects of dynamic factors over a period, instead of selecting attribute values corresponding to time  $t$ , we expand the selecting window size to  $m + 1$  when forming  $E^t$ , that is, selecting  $D_w^{t-m,t} = [D_w^{t-m}, D_w^{t-m-1}, \dots, D_w^t]$  for each dynamic attribute submatrix  $D_w$ . Finally, through the attribute augmentation unit (A-Cell), the augmented matrix containing both static and dynamic external attributes as well as traffic characteristic information at time  $t$  is formed as:

$$E^t = [X^t, S, D_1^{t-m,t}, D_2^{t-m,t}, \dots, D_w^{t-m,t}] \quad (3)$$

where  $E^t \in R^{n \times (p+l+w*(m+1))}$ .

### 3.4 Spatiotemporal graph convolutional network

Road networks are naturally graph structures. Since traffic flows on connected road sections, modeling spatial dependencies on the network is essential in the traffic forecasting task. Graph convolutional networks (GCNs)[23] have recently been favored by many researchers because of their ability to model complex relationships and interdependency in non-Euclidean domains and have made great progress in recent years. The main purpose of the graph convolutional neural network is to obtain the representation of each node in the graph while considering the influence of neighboring nodes. To capture the dependency created by topological structures, GCN takes the adjacency matrix and the feature matrix as inputs, and the modeling process can be expressed as follows.

$$y_{l+1} = \sigma \left( \tilde{D}^{-\frac{1}{2}} \tilde{A} \tilde{D}^{-\frac{1}{2}} y_l W_l \right) \quad (4)$$

where  $\sigma$  is the activation function,  $\tilde{A} = A + I$  represents the adjacency matrix with self-loops,  $\tilde{D}$  is the corresponding degree matrix,  $W_l$  is the weight matrix of the  $l$ -th convolutional layer,  $y_l$  is the output representation, and  $y_0 = X$ .

Moreover, traffic states can constantly change over time in the real world. To capture the temporal dependencies in traffic time series, recurrent neural networks[29] (RNNs) are

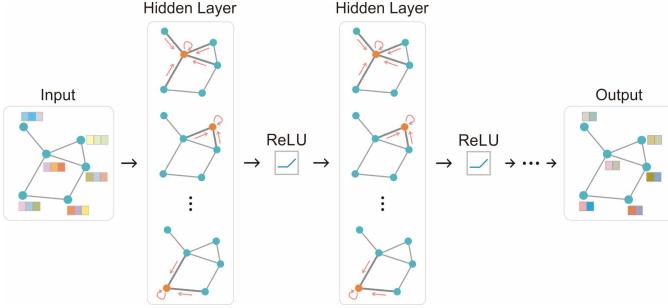


Fig. 3. The architecture of the GCN model.

introduced. However, traditional RNNs have defects such as the inability to retain long-term memory and the existence of exploding gradient. LSTM[30], a variant of RNN, provides a solution to such problems. It designs the input gate and the forget gate unit to retain and forget information of the previous state and current input and derive the current state with the output gate unit. Based on LSTM, GRU[31] replaces the forget gate and input gate with one update gate, which both reduces the training time and improves the training efficiency while having equivalent calculation accuracy. Specifically, the GRU model can be regarded as compositions of reset gates and update gates, as shown in Fig. 4, where  $x_{t-1}$  is the input feature of one node at time  $t-1$  and  $h_{t-k}, \dots, h_{t-1}, h_t$  represent the hidden states at time  $t-k, \dots, t-1, t$ .  $\sigma$  and  $\tanh$  refer to the sigmoid and tanh activation, indicating gate signals. Consider gates at  $t$  for example,  $r_t$  is the reset gate used to combine previous state  $h_{t-1}$  with current information  $x_t$  to derive the candidate hidden state  $c_t$ .  $u_t$  refers to the update gate that can be used to determine how much of  $h_{t-1}$  to discard and what new information of  $c_t$  to incorporate to derive the final hidden state  $h_t$ .

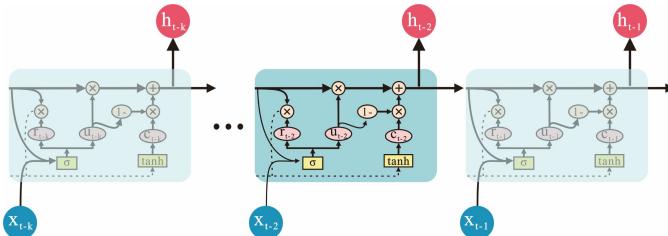


Fig. 4. The architecture of the GRU model.

By using the combination of the GCN and GRU model, the purpose of capturing the spatiotemporal dependencies in real-world traffic data can be achieved. The GCN model is used to generate representations of road sections that capture the spatial dependencies on the road network at each timestamp. Then, these time-varying representations are fed into the GRU model to capture the temporal dependencies.

### 3.4.1 Attribute-enhanced spatiotemporal graph convolution model (AST-GCN)

Based on the spatiotemporal graph convolutional network, we propose a traffic forecasting model AST-GCN that integrates the information of external factors with the attribute augmentation units.

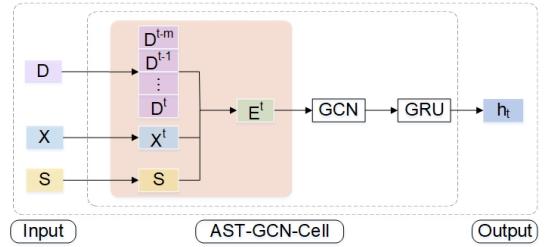


Fig. 5. The architecture of an AST-GCN cell.

As shown in Fig. 5, the attribute enhancement unit expands the dimension of the original feature matrix by incorporating static and dynamic external attributes. At time  $t$ ,  $X^t$  is extracted from the original traffic feature matrix  $X$ ,  $\{D^{t-m}, \dots, D^t\}$  is the collection of dynamic information from time  $t-m$  to  $t$ ,  $S$  is the static attribute, which is constantly the same for different timestamps, and  $E^t$  is the enhanced matrix after the fusion of the external attributes and original traffic features.

The enhancement matrix  $E$  is then used as the input of the spatiotemporal model  $f$  to obtain the final prediction result  $y$ .

$$\hat{y} = f(A, X, E) \quad (5)$$

Specifically, the enhanced matrices are fed into a series of GCNs to generate time-varying features of road sections that encode the spatial characteristics of the traffic states. Then, the feature series are used as the input of GRUs to model the temporal dependencies and derive hidden traffic states.

$$u_t = \sigma(W_u \cdot [gc(E^t, A), h_{t-1}] + b_u) \quad (6)$$

$$r_t = \sigma(W_r \cdot [gc(E^t, A), h_{t-1}] + b_r) \quad (7)$$

$$c_t = \tanh(W_c \cdot [gc(E^t, A), (r_t, h_{t-1})] + b_c) \quad (8)$$

$$h_t = u_t * h_{t-1} + (1 - u_t) * c_t \quad (9)$$

where  $gc(\cdot)$  represents the graph convolution operation.

### 3.4.2 Loss Function

In the process of model training, the goal of traffic forecasting is to make the prediction result approximate the real traffic states as much as possible. Therefore, the objective of the loss function is to minimize the prediction error.

$$Loss = \|y_t - \hat{y}_t\| + \lambda L_{reg} \quad (10)$$

where  $y_t$  and  $\hat{y}_t$  are the ground truth and prediction,  $L_{reg}$  represents the regular term to avoid overfitting, and  $\lambda$  is a hyperparameter.

### 3.5 Framework

By integrating the spatiotemporal graph convolution network and the attribute augmentation unit, we propose a traffic prediction model (AST-GCN) that can integrate the external influence information (weather condition and surrounding POIs in experiments) to facilitate traffic forecasting. The framework of our work is shown in Fig. 6, which is mainly divided into four parts: data preprocessing, attribute augmentation, spatial-temporal dependency modeling and prediction.

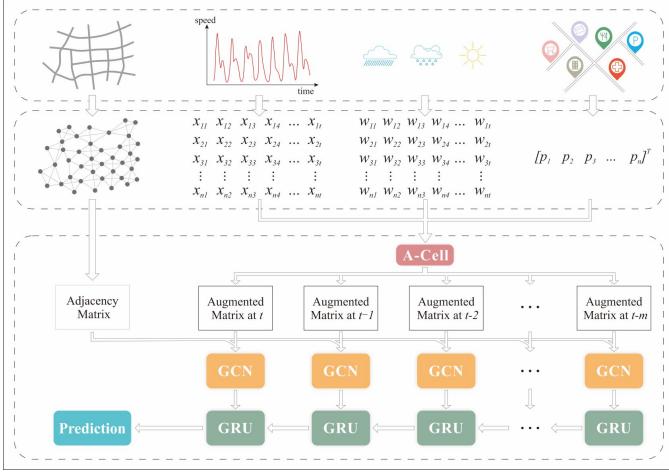


Fig. 6. Framework.

## 4 EXPERIMENTS

### 4.1 Datasets

The following datasets are used in the experiment:

(1) SZ\_taxi: this dataset contains Shenzhen taxi trajectory data collected from January 1 to January 31, 2015. A total of 156 major road sections in the Luohu District are selected, and their connectivity is modeled by a  $156 \times 156$  adjacency matrix. The traffic speed time series of selected sections are calculated and form a feature matrix, where rows are indexed by road sections and columns are indexed by the timestamps.

(2) SZ\_POI: this dataset provides information about POIs surrounding selected road sections. The POI categories can be divided into 9 types: catering services, enterprises, shopping services, transportation facilities, education services, living services, medical services, accommodations, and others. After calculating the distribution of POIs on each road section, the type of POI with the largest proportion is used as the feature of the road section. Therefore, the obtained static attribute matrix is of size  $156 \times 1$ .

(3) SZ\_Weather: this auxiliary information contains the weather conditions about the study area recorded every 15 minutes in January 2015. The weather conditions are divided into five categories: sunny, cloudy, fog, light rain and heavy rain. With the information of time-varying weather conditions, we construct the dynamic attribute matrix with size  $156 \times 2,976$ .

### 4.2 Evaluation metrics

To evaluate the prediction performance of the proposed model, we use the following metrics to evaluate the prediction results.

#### (1) Root mean square error (RMSE)

The smaller the RMSE value, the smaller the prediction error, and the better the performance of the model.

$$RMSE = \left[ \frac{1}{n} \sum_{t=1}^n (y_t - \hat{y}_t)^2 \right]^{\frac{1}{2}} \quad (11)$$

#### (2) Mean absolute error (MAE)

MAE describes the average of the sum of the absolute difference between the predicted result and the ground truth. It is mainly used to evaluate the prediction error.

$$MAE = \frac{\sum_{i=1}^n |y_i - \hat{y}_i|}{n} \quad (12)$$

#### (3) Accuracy

$$Accuracy = 1 - \frac{\|y - \hat{y}\|_F}{\|y\|_F} \quad (13)$$

where  $\|\cdot\|_F$  represents the Frobenius norm.

#### (4) $R^2$

$R^2$  is mainly used to measure the predictive ability of the model, and the larger the  $R^2$ , the better the prediction results.

$$R^2 = 1 - \frac{\sum_{t=1}^n (y_t - \hat{y}_t)^2}{\sum_{t=1}^n (y_t - \bar{y}_t)^2} \quad (14)$$

#### (5) Explained variation (VAR)

VAR measures the proportion to which the proposed model accounts for the variation in real traffic states. It is mainly used to measure the predictive ability of the model.

$$Var = 1 - \frac{Var(y - \hat{y})}{Var(y)} \quad (15)$$

### 4.3 Parameter Settings

The hyperparameters of the AST-GCN model mainly include the learning rate, training epoch, hidden units, batch size, and proportion of the training set. While the learning rate, batch size and proportion of the training set are manually set to 0.001, 64 and 0.8, the other parameters are searched through experiments.

First, training epochs in set [500, 1,000, 1,500, 2,000, 3,000, 3,500] are tested to analyze the change in model performance. Fig. 7 shows the evaluation results under varying training epoch settings. As the value of the training epoch increases, the changes in the evaluation metrics become stable, and the turning point is 3,000. Then, when the training epoch is fixed to 3,000, the number of hidden units is selected from the candidate set [8, 16, 32, 64, 100, 128]. As shown in Fig. 8, when the number of units reaches 100, the model becomes stable. Therefore, the training epoch is determined to be 3,000, while the number of hidden units is 100.

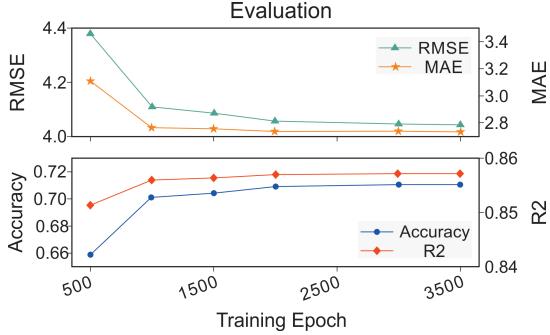


Fig. 7. The influence of the selection of epochs on the prediction performance.

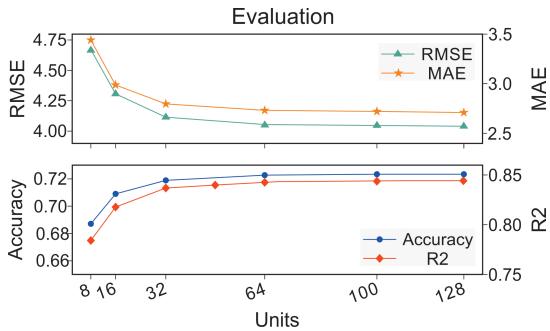


Fig. 8. The influence of the selection of units on the prediction performance.

#### 4.4 Baselines

We compare the proposed AST-GCN with the following baselines: (1) autoregressive integral moving average model (ARIMA)[32], (2) support vector regression (SVR)[33], (3) graph convolution model (GCN)[23], (4) gated recurrent unit model (GRU)[34], (5) spatiotemporal graph convolution model (TGCN)[25] and (6) diffusion convolutional recurrent neural network (DCRNN)[35]. The hyperparameters of these baseline methods are the same as those in the original paper or released codes.

#### 4.5 Experimental Results

Limited by the data sources, this paper simply uses the POI distributions and weather conditions to exemplify the importance of enhancing the model's ability to perceive external factors. The experiments are designed from five perspectives: the comparison of forecasting accuracy with baselines, the influence of introducing different kinds of external information, the influence of different predicting lengths, the interpretability of the proposed model, and the model's robustness.

##### 4.5.1 Forecasting Comparison with Baselines

To verify the effectiveness of the AST-GCN model in the traffic forecasting task, the prediction accuracy is compared with baselines, as shown in Table. 1.

From the experimental results, it can be found that the prediction accuracy of the methods based on deep learning methods (AST-GCN, GCN, GRU, and TGCN) is higher than that of other methods. Compared with the SVR and ARIMA models, the RMSE of the attribute-aware AST-GCN model is

reduced by approximately 44.19% and 40.49%, respectively. From the spatiotemporal perspective, compared with the GCN and GRU, which only focus on spatial or temporal relationships, the RMSE of AST-GCN, which considers both, is reduced by approximately 28.58% and 20.44%, respectively, and the overall evaluation given by other metrics is also significantly improved. From the perspective of attribute enhancement, the AST-GCN model, which considers external factors, outperforms spatiotemporal models such as TGCN and DCRNN, and the RMSE is reduced by approximately 0.98% and 10.46%, respectively. The comparison results verify the effectiveness of the proposed AST-GCN model.

#### 4.5.2 Ablation Experiments

We conduct ablation experiments to prove that dynamic and static auxiliary information can play a role in the task of traffic forecasting. The experimental setting is divided into adding only static external information, adding only dynamic external information, adding dynamic and static external information, and not adding any external information. The results are shown in Table. 2, and the fourth column is the result of adding dynamic weather conditions, the fifth column is the result of adding static POI distribution information, and the sixth column is the result of adding both.

From the perspective of simply adding dynamic external information, the RMSE of AST-GCN (weather) is reduced by approximately 0.78% and 10.27% compared with the TGCN and DCRNN models. From the perspective of simply adding static external information, the RMSE of AST-GCN (POI) is 0.68% and 10.18% lower than that of the TGCN and DCRNN models. After adding both dynamic and static external information, the RMSE of AST-GCN (weather+POI) is 0.98% and 10.46% lower than those of the TGCN and DCRNN models.

When considering the type of information to introduce, the model enhanced with dynamic external information is better than the model with static external information, indicating that the impact of weather conditions on traffic states is greater than that of the surrounding POIs. In addition, adding both dynamic and static external information, the model performs better than the model that adds a single type of information. The prediction error is lower by 0.21% and 0.31%, indicating the complementarity of dynamic and static factors. In general, enhancing the model with external information can facilitate the traffic forecasting task.

##### 4.5.3 The Forecasting Performance of AST-GCN for Different Predicting Horizons

This experiment tests the performance of AST-GCN for different prediction horizons (15 min, 30 min, 45 min, and 60 min), and the performance comparison between the AST-GCN model and the baseline models under different prediction horizons is shown in Table. 3. For the prediction horizon of 15 minutes, the RMSE of the AST-GCN model is approximately 0.98% and 10.46% lower than that of TGCN and DCRNN. For the prediction horizon of 30 minutes, the RMSE of the AST-GCN model's prediction error is approximately 0.59% and 11.77% lower than those of TGCN and DCRNN. For the prediction horizon of 45 min, the RMSE of the AST-GCN model is approximately

TABLE 1  
Performance comparison of different models for traffic forecasting on the Shenzhen dataset.

Evaluation metrics	SVR	ARIMA	GCN	GRU	DCRNN	TGCN	AST-GCN
RMSE	7.2203	6.7708	5.6419	5.0649	4.5000	4.0696	4.0294
MAE	4.7762	4.6656	4.2265	2.5988	3.1700	2.7460	2.7035
Accuracy	0.7060	0.3852	0.6119	0.7243	0.2913	0.7165	0.7193
R2	0.8367	*	0.6678	0.8322	0.8391	0.8388	0.8512
var	0.8375	0.0111	0.6679	0.8322	0.8391	0.8388	0.8512

\* represents a negative value.

TABLE 2  
Ablation Experiments under different experimental settings.

Evaluation metrics	TGCN	DCRNN	AST-GCN		
			Weather	POI	Weather+POI
RMSE	4.0696	4.5000	4.0378	4.0418	4.0294
MAE	2.7460	3.1700	2.7059	2.6985	2.7035
Accuracy	0.7165	0.2913	0.7187	0.7180	0.7193
R2	0.8388	0.8391	0.8506	0.8498	0.8512
Var	0.8388	0.8391	0.8506	0.8498	0.8512

TABLE 3

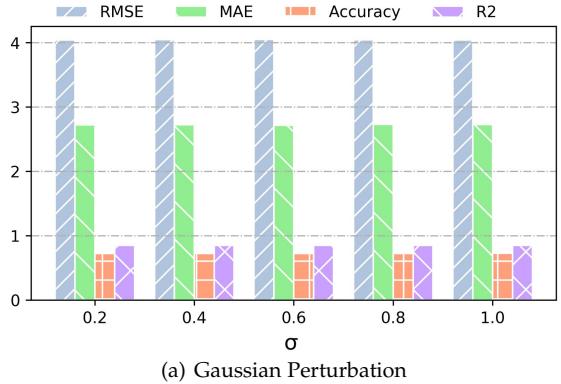
Performance comparison between the AST-GCN model and the baseline models for different prediction horizons.

Time	Metric	TGCN	DCRNN	AST-GCN
15 min	RMSE	4.0696	4.5000	4.0294
	MAE	2.7460	3.1700	2.7035
	Accuracy	0.7165	0.2913	0.7193
	R2	0.8388	0.8391	0.8512
	Var	0.8388	0.8391	0.8512
	RMSE	4.0770	4.5600	4.0529
30 min	MAE	2.7470	3.2300	2.7265
	Accuracy	0.7159	0.2970	0.7176
	R2	0.8377	0.8332	0.8494
	Var	0.8377	0.8360	0.8495
45 min	RMSE	4.1035	4.6000	4.0822
	MAE	2.7788	3.2700	2.7611
	Accuracy	0.7141	0.3021	0.7156
	R2	0.8357	0.8275	0.8473
60 min	Var	0.8357	0.8314	0.8474
	RMSE	4.1266	4.6400	4.1001
	MAE	2.7911	3.3100	2.7744
	Accuracy	0.7125	0.3069	0.7143
	R2	0.8339	0.8219	0.8459
	Var	0.8340	0.8267	0.8460

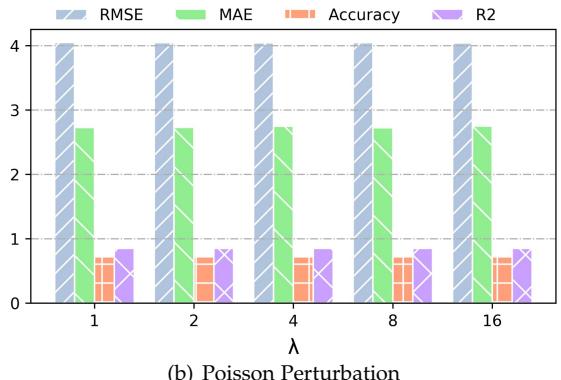
0.52% and 11.26% lower than that of TGCN and DCRNN, respectively. When the prediction horizon is set to 60 min, the RMSE of the AST-GCN model is approximately 0.64% and 11.64% lower than TGCN and DCRNN, respectively. It can be concluded that the AST-GCN model can maintain good performance for different horizons and the capability for long-term forecasting.

#### 4.5.4 Perturbation Analysis

Gaussian noises and Poisson noises are added to the data to test the robustness of the model. The two types of noise obey Gaussian distribution  $N \in (0, \sigma^2)$  ( $\sigma \in 0.2, 0.4, 0.6, 0.8, 1, 2$ ) and Poisson distribution  $P(\lambda)$  ( $\lambda \in 1, 2, 4, 8, 16$ ), respectively. The experimental results are shown in Fig. 9, and the robustness of the AST-GCN model is verified by the fact that the changes in evaluation metrics across different noise settings are negligible.



(a) Gaussian Perturbation



(b) Poisson Perturbation

#### 4.5.5 Interpretation of the AST-GCN

To explain the predictive ability of the model more clearly, this experiment visually compares and analyze the true speed value of the test set and the prediction result of the attribute enhancement graph convolution model (AST-GCN) and explain the model from the following two perspectives:

##### (1) Long-term forecasting

Based on the historical one-hour data, the results of predicting traffic speeds for the next 15 minutes, 30 minutes, 45 minutes, and 60 minutes are visualized in Fig. 10-13.

The upper subfigure in each figure is the forecasting result from January 26, 2015, to January 31, 2015, and the lower image is the forecasting result for January 27, 2015. From the visualization results, we can conclude the following:

- For different forecasting horizons, the model can well predict the traffic speed value. The prediction result of the model is similar to the changing trend of the real speed;
- The performance of short-term forecasting is better than that of long-term forecasting. Comparing the prediction results for 15 min and 60 min, it can be found that the curve of short-term prediction is closer to the real curve, which indicates that the model can better capture the short-range dependencies while losing some information in long-term prediction;
- For the capturing of the turning points of the speed changing trend, the AST-GCN model has considerable deviations at high and low peaks. The reason may be that the sudden change in traffic states is affected not only by the weather and POI factors used in this study but also by a combination of other factors.

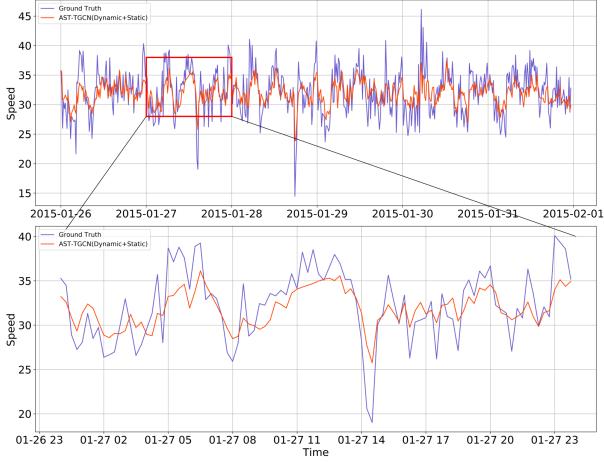


Fig. 10. The visualization results for the 15 minute prediction horizon.

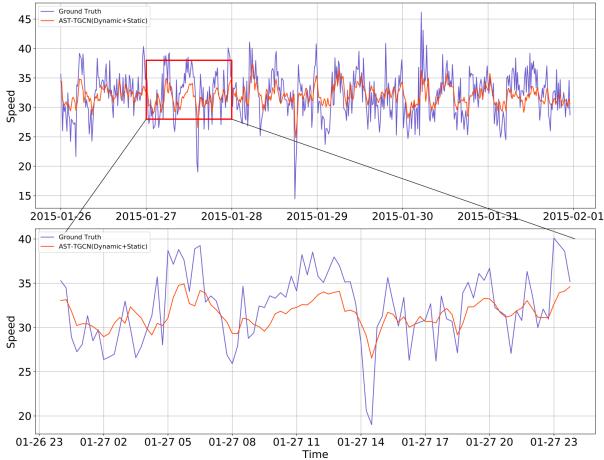


Fig. 11. The visualization results for the 30 minute prediction horizon .

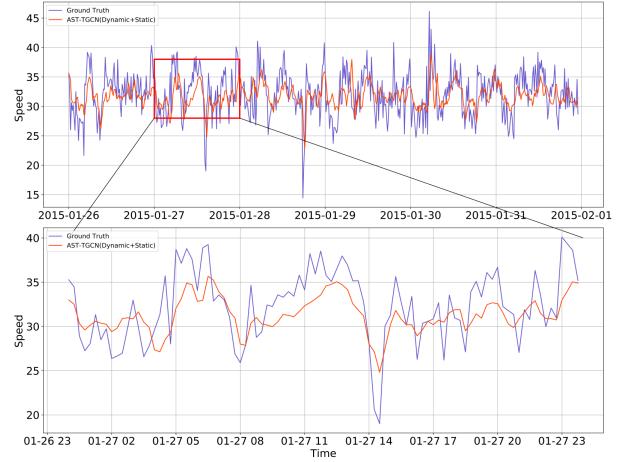


Fig. 12. The visualization results for the 45 minute prediction horizon.

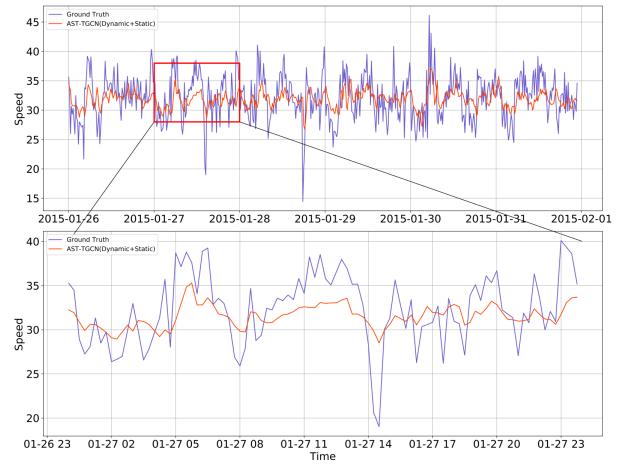


Fig. 13. The visualization results for the 60 minute prediction horizon .

## (2) Attribute importance

To further study the validity of static and dynamic external information, we visualize the results of the ablation experiments. Fig. 14-16 show the visual comparison of forecasting enhanced by static POI information, dynamic weather condition information, and their combination with forecasting without external information. shows the comparison between the forecasting results enhanced by different attributes.

- External information improves the model's perception of peaks and turning points. Fig. 14-16 show that the prediction results of the model with external information are closer to the ground truth than the prediction result without attribute auxiliary information at turning points and peaks.
- From the visualization results in Fig. 17, it can be found that the deviation between the predicted results of AST-GCN (dynamic+static attribute) and the real speed value is smaller than that of AST-GCN (static attribute) and AST-GCN (dynamic attribute), which indicates that the diversity of external information can better facilitate forecasting.

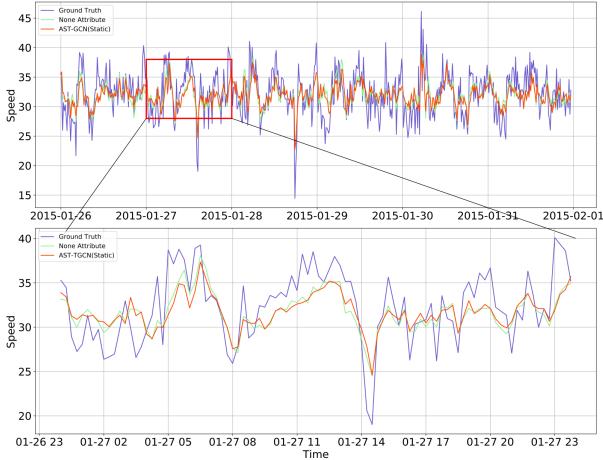


Fig. 14. Comparison between forecasting enhanced by static POI information and forecasting without external information.

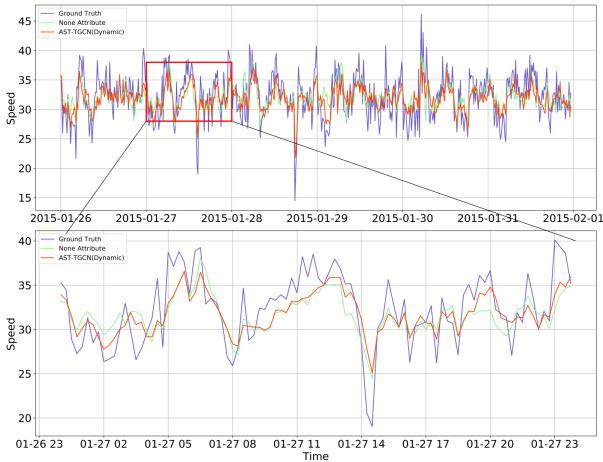


Fig. 15. Comparison between forecasting enhanced by dynamic weather condition information and forecasting without external information.

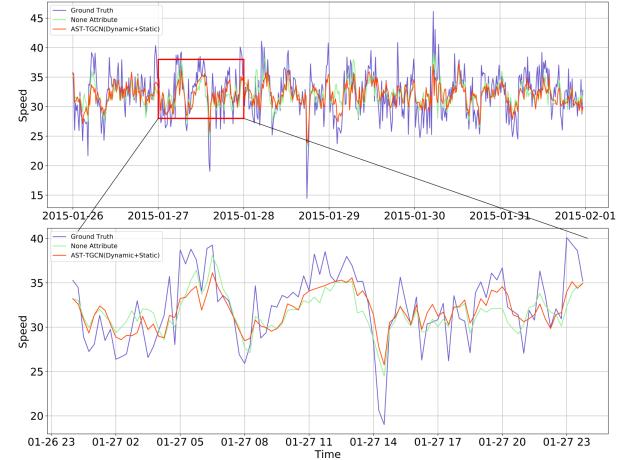


Fig. 16. Comparison between forecasting enhanced by the combination of static and dynamic information and forecasting without external information.

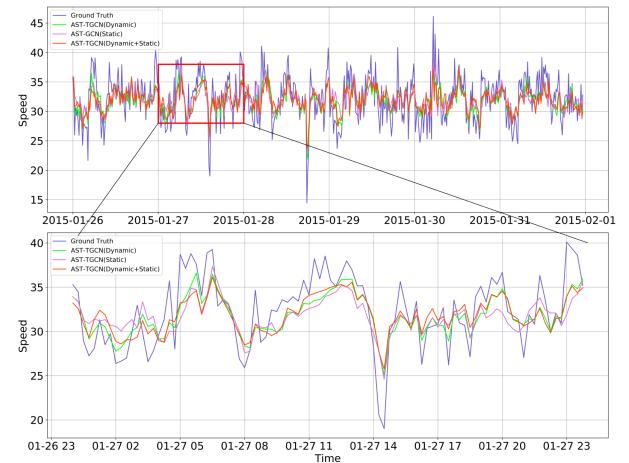


Fig. 17. Comparison among forecasting enhanced by different external information.

## 5 CONCLUSION

This paper addresses the problem that the traditional urban traffic forecasting models cannot comprehensively consider the external factors that affect the traffic states and proposes an attribute-enhanced spatiotemporal graph convolution model AST-GCN. The model can not only integrate static external information but also dynamic external data. Through comparison with the baseline methods, the prediction results verify the importance of considering external information in traffic forecasting tasks. In addition, this paper uses a perturbation analysis to test the robustness of the model.

## ACKNOWLEDGMENTS

This work was supported by the National Science Foundation of China [grant numbers ].

## REFERENCES

- [1] J. Zhang, F.-Y. Wang, K. Wang, W.-H. Lin, X. Xu, and C. Chen, "Data-driven intelligent transportation sys-

- tems: A survey," *IEEE Transactions on Intelligent Transportation Systems*, vol. 12, no. 4, pp. 1624–1639, 2011.
- [2] Z. Cui, K. Henrickson, R. Ke, and Y. Wang, "Traffic graph convolutional recurrent neural network: A deep learning framework for network-scale traffic learning and forecasting," *IEEE Transactions on Intelligent Transportation Systems*, 2019.
- [3] C. Zheng, X. Fan, C. Wen, L. Chen, C. Wang, and J. Li, "Deepstd: Mining spatio-temporal disturbances of multiple context factors for citywide traffic flow prediction," *IEEE Transactions on Intelligent Transportation Systems*, 2019.
- [4] B. Yu, H. Yin, and Z. Zhu, "Spatio-temporal graph convolutional networks: A deep learning framework for traffic forecasting," *arXiv preprint arXiv:1709.04875*, 2017.
- [5] J. Park, D. Li, Y. L. Murphrey, J. Kristinsson, R. McGee, M. Kuang, and T. Phillips, "Real time vehicle speed prediction using a neural network traffic model," in *The 2011 International Joint Conference on Neural Networks*. IEEE, 2011, pp. 2991–2996.

- [6] H. Yao, X. Tang, H. Wei, G. Zheng, and Z. Li, "Revisiting spatial-temporal similarity: A deep learning framework for traffic prediction," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 33, 2019, pp. 5668–5675.
- [7] B. L. Smith and M. J. Demetsky, "Traffic flow forecasting: comparison of modeling approaches," *Journal of transportation engineering*, vol. 123, no. 4, pp. 261–266, 1997.
- [8] M. S. Ahmed and A. R. Cook, *Analysis of freeway traffic time-series data by using Box-Jenkins techniques*, 1979, no. 722.
- [9] B. M. Williams, P. K. Durvasula, and D. E. Brown, "Urban freeway traffic flow prediction: application of seasonal autoregressive integrated moving average and exponential smoothing models," *Transportation Research Record*, vol. 1644, no. 1, pp. 132–141, 1998.
- [10] S. Lee and D. B. Fambro, "Application of subset autoregressive integrated moving average model for short-term freeway traffic volume forecasting," *Transportation Research Record*, vol. 1678, no. 1, pp. 179–188, 1999.
- [11] G. A. Davis and N. L. Nihan, "Nonparametric regression and short-term freeway traffic forecasting," *Journal of Transportation Engineering*, vol. 117, no. 2, pp. 178–188, 1991.
- [12] C. Tebaldi and M. West, "Bayesian inference on network traffic using link count data," *Journal of the American Statistical Association*, vol. 93, no. 442, pp. 557–573, 1998.
- [13] W. Hu, L. Yan, K. Liu, and H. Wang, "A short-term traffic flow forecasting method based on the hybrid pso-svr," *Neural Processing Letters*, vol. 43, no. 1, pp. 155–172, 2016.
- [14] W. Huang, G. Song, H. Hong, and K. Xie, "Deep architecture for traffic flow prediction: deep belief networks with multitask learning," *IEEE Transactions on Intelligent Transportation Systems*, vol. 15, no. 5, pp. 2191–2201, 2014.
- [15] Y. Jia, J. Wu, and Y. Du, "Traffic speed prediction using deep learning method," in *2016 IEEE 19th International Conference on Intelligent Transportation Systems (ITSC)*. IEEE, 2016, pp. 1217–1222.
- [16] Y. Lv, Y. Duan, W. Kang, Z. Li, and F.-Y. Wang, "Traffic flow prediction with big data: a deep learning approach," *IEEE Transactions on Intelligent Transportation Systems*, vol. 16, no. 2, pp. 865–873, 2014.
- [17] Y. Tian and L. Pan, "Predicting short-term traffic flow by long short-term memory recurrent neural network," in *2015 IEEE international conference on smart city/SocialCom/SustainCom (SmartCity)*. IEEE, 2015, pp. 153–158.
- [18] X. Ma, Z. Tao, Y. Wang, H. Yu, and Y. Wang, "Long short-term memory neural network for traffic speed prediction using remote microwave sensor data," *Transportation Research Part C: Emerging Technologies*, vol. 54, pp. 187–197, 2015.
- [19] R. Fu, Z. Zhang, and L. Li, "Using lstm and gru neural network methods for traffic flow prediction," in *2016 31st Youth Academic Annual Conference of Chinese Association of Automation (YAC)*. IEEE, 2016, pp. 324–328.
- [20] D. Zhang and M. R. Kabuka, "Combining weather condition data to predict traffic flow: a gru-based deep learning approach," *IET Intelligent Transport Systems*, vol. 12, no. 7, pp. 578–585, 2018.
- [21] Y. Liu, H. Zheng, X. Feng, and Z. Chen, "Short-term traffic flow prediction with conv-lstm," in *2017 9th International Conference on Wireless Communications and Signal Processing (WCSP)*. IEEE, 2017, pp. 1–6.
- [22] Y. Wu and H. Tan, "Short-term traffic flow forecasting with spatial-temporal correlation in a hybrid deep learning framework," *arXiv preprint arXiv:1612.01022*, 2016.
- [23] T. N. Kipf and M. Welling, "Semi-supervised classification with graph convolutional networks," *arXiv preprint arXiv:1609.02907*, 2016.
- [24] J. Zhou, G. Cui, Z. Zhang, C. Yang, Z. Liu, L. Wang, C. Li, and M. Sun, "Graph neural networks: A review of methods and applications," *arXiv preprint arXiv:1812.08434*, 2018.
- [25] L. Zhao, Y. Song, C. Zhang, Y. Liu, P. Wang, T. Lin, M. Deng, and H. Li, "T-gcn: A temporal graph convolutional network for traffic prediction," *IEEE Transactions on Intelligent Transportation Systems*, 2019.
- [26] Z. Li, G. Xiong, Y. Chen, Y. Lv, B. Hu, F. Zhu, and F.-Y. Wang, "A hybrid deep learning approach with gcn and lstm for traffic flow prediction," in *2019 IEEE Intelligent Transportation Systems Conference (ITSC)*. IEEE, 2019, pp. 1929–1933.
- [27] B. Liao, J. Zhang, C. Wu, D. McIlwraith, T. Chen, S. Yang, Y. Guo, and F. Wu, "Deep sequence learning with auxiliary information for traffic prediction," in *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, 2018, pp. 537–546.
- [28] J. Van Lint, S. Hoogendoorn, and H. J. van Zuylen, "Freeway travel time prediction with state-space neural networks: modeling state-space dynamics with recurrent neural networks," *Transportation Research Record*, vol. 1811, no. 1, pp. 30–39, 2002.
- [29] J. T. Connor, R. D. Martin, and L. E. Atlas, "Recurrent neural networks and robust time series prediction," *IEEE transactions on neural networks*, vol. 5, no. 2, pp. 240–254, 1994.
- [30] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural computation*, vol. 9, no. 8, pp. 1735–1780, 1997.
- [31] K. Cho, B. Van Merriënboer, D. Bahdanau, and Y. Bengio, "On the properties of neural machine translation: Encoder-decoder approaches," *arXiv preprint arXiv:1409.1259*, 2014.
- [32] M. M. Hamed, H. R. Al-Masaeid, and Z. M. B. Said, "Short-term prediction of traffic volume in urban arterials," *Journal of Transportation Engineering*, vol. 121, no. 3, pp. 249–254, 1995.
- [33] H. Su, L. Zhang, and S. Yu, "Short-term traffic flow prediction based on incremental support vector regression," in *Third International Conference on Natural Computation (ICNC 2007)*, vol. 1. IEEE, 2007, pp. 640–645.
- [34] J. Chung, C. Gulcehre, K. Cho, and Y. Bengio, "Empirical evaluation of gated recurrent neural networks on sequence modeling," *arXiv preprint arXiv:1412.3555*,

- 2014.
- [35] Y. Li, R. Yu, C. Shahabi, and Y. Liu, "Diffusion convolutional recurrent neural network: Data-driven traffic forecasting," *arXiv preprint arXiv:1707.01926*, 2017.