An Adaptive Prediction Model for Randomly Distributed Traffic Data in Urban Road Networks

Ruiyuan Jiang, Shangbo Wang, *Member, IEEE*, Dongyao Jia, *Member, IEEE*, Guoqiang Mao, *Fellow, IEEE*, and Eng Gee Lim, *Senior Member, IEEE*

***Abstract*—Effective and efficient traffic prediction can provide a reliable data basis for traffic management in Intelligent Transportation Systems (ITS). While various machine learning methods have been proposed to enhance prediction accuracy in recent decades, there remain potential issues to be further addressed. Firstly, the inherent randomness of traffic dynamics usually leads to some outliers in historical observations, which may deviate the model parameter estimation when utilizing deep learning-based models to learn data distribution. Secondly, the spatial correlation among the road sections may dynamically change over time, posing challenges for modeling. In addition, due to the complexity of urban traffic networks, capturing such non-linear spatial dependencies based on the global road structure may consume huge computational resources. To address these issues, this paper proposes an adaptive temporal graph attention network (ATGAN), which is implemented in two steps: 1) An outlier time series filter (OTSF) technique is introduced to mitigate the adverse impact of outlier points and to adaptively learn the distribution of fluctuations of traffic data; 2) We design a group attention temporal graph convolutional network (GA-TGCN) to model the spatiotemporal features among neighboring road sections, which is achieved by adjusting the spatial correlation matrix dynamically in each training epoch with attention mechanism. We evaluate the prediction performance of ATGAN on two real-world datasets and the results show that our model can achieve higher prediction accuracy in less computational time compared with baseline methods.**

***Index Terms*—Attention mechanism, Intelligent Transportation Systems (ITS), randomly distributed data, spatiotemporal correlations, traffic prediction.**

Received 21 February 2024; revised 30 June 2024 and 25 October 2024; accepted 21 December 2024. Date of publication 2 January 2025; date of current version 20 May 2025. This work was supported in part by the National Natural Science Foundation of China under Grant 62372384, in part by the Suzhou Science and Technology Development Planning Programme under Grant ZXL2024342, and in part by the Suzhou Municipal Key Laboratory for Intelligent Virtual Engineering under Grant SZS2022004. The review of this article was coordinated by Dr. Zeeshan Kaleem. *(Corresponding author: Shangbo Wang.)*

Ruiyuan Jiang, Dongyao Jia, and Eng Gee Lim are with the School of Advanced Technology, Xi'an Jiaotong-Liverpool University, Suzhou 215000, China (e-mail: ruiyuan.jiang20@student.xjtlu.edu.cn; dongyao.jia@xjtlu.edu.cn; enggee.lim@xjtlu.edu.cn).

Shangbo Wang is with the Department of Engineering and Design, University of Sussex, BN1 9RH Brighton, U.K., and also with the Department of Civil Engineering, The University of Hong Kong, Hong Kong (e-mail: wangshangbo.best@gmail.com).

Guoqiang Mao is with the Communication and Information Systems, Xidian University, Xi'an 710126, China (e-mail: g.mao@ieee.org).

Digital Object Identifier 10.1109/TVT.2024.3525023

I. Introduction

THERE has been an exponential increase in the usage of autonomous vehicles in the past few years. This is due to a sharp increase in the popularity and improvement of advanced communication technologies and artificial intelligence techniques. In urban networks, traffic data can be viewed as time-series data collected from multiple sensors. Predicting these traffic conditions is vital for improving autonomous vehicle technologies and is a foundation for practical applications [1]. For instance, real-time traffic prediction provides a valuable data foundation for various applications, such as vehicle navigation systems [2] and predictive bus control frameworks [3]. Also, it provides the data basis for autonomous vehicles to make adaptive decisions according to road conditions. Consequently, achieving accurate and reliable traffic prediction is significant for ITS in urban networks.

In recent decades, various model-driven and data-driven approaches have been applied to traffic prediction. Model-driven techniques such as autoregressive integrated moving average (ARIMA) [4], ST-ARIMA [5], Kohonen ARIMA [6], seasonal ARIMA (SARIMA) [7], and Kalman Filtering [8] have been widely used due to their simplicity and interpretability. Among data-driven approaches, time-series prediction models such as Long Short-term Memory (LSTM) [9], Gated Recurrent Unit (GRU) [10] have been proposed to capture the temporal correlation of traffic information. Also, graph-based models such as Graph Convolutional Networks (GCNs) [11] have been employed to model the spatial features of traffic information. In addition, hybrid models that consider both temporal and spatial correlations have been developed, such as Temporal Graph Convolutional Network (T-GCN) [12] and spatiotemporal attention mechanisms [13]. These studies have made great progress in traffic prediction. However, there are still some potential fields that have value to be studied.

- First, real-world traffic data is always randomly distributed in the time domain which contains many large fluctuations and outlier data points. This phenomenon is especially evident in urban transportation networks caused by the huge traffic flow and complex traffic environment [14]. When the deep learning-based models try to estimate the model parameters from the historical traffic data, these fluctuations and outlier points may disturb the process by which the models learn the distributions from the historical observations, thus having a negative impact on the accuracy of model parameters estimation [15].

- Second, traffic data of one certain road section is correlated with that of its neighboring sections, and capturing such spatial correlation is an important element in traffic prediction tasks. Nonetheless, the spatial correlation is a dynamic feature that may change over different time steps, and how to construct a structure to fully consider the dynamics of this feature is still challenging [16]. In addition, with the development of urban transportation, the structure of road networks tends to be complicated, so extracting the spatial features based on the global network will dilute the relatively strong correlation among neighboring road sections and increase the computational complexity, thus affecting the model's efficiency.

According to the listed fields, we propose an adaptive temporal graph attention network (ATGAN) that can improve the prediction performance of traffic prediction, which contains two stages. First, an outlier time series filter (OTSF) technique is proposed to reduce the negative impact of the randomness of traffic data in the model parameter estimation process. Second, we apply a group attention temporal graph convolutional network (GA-TGCN) to fully model the spatiotemporal features of traffic data.

Specifically, the main contributions are summarized as follows:

1) We propose an OTSF technique, which can adaptively learn the distribution of outlier series and smooth non-stationary traffic data by adding a learnable bias based on the spatial correlation among each target road section. Through the OTSF technique, we can achieve more stationary historical traffic data which serves as the input of the traffic prediction network, which can improve the prediction performance when the data has some outlier time series.
2) We design the GA-TGCN module for accurately modeling the spatiotemporal features of traffic data among different sensors. GA-TGCN incorporates a group attention mechanism, which constructs a learnable spatial correlation matrix based on the topological graph information to fully extract the spatial features among the neighboring road sections, thus the dynamics of spatial correlation can be further described and the error propagation of incorrect spatial feature acquisition can be reduced during the model training process. In addition, GA-TGCN applies a GRU model to extract the temporal features from traffic data.
3) We evaluate the performance of ATGAN with two real-world datasets: the Los-loop and Seattle-loop datasets. We compare ATGAN with some state-of-the-art methods and demonstrate the advantages of our model in terms of prediction performance and computational complexity. Furthermore, we test the robustness and validate its efficiency under different types of noise.

II. Related Work

Various studies related to traffic prediction models have been proposed in the past few years. Alghamdi et al. [17] explored the use of the ARIMA model for predicting traffic congestion which showed that the ARIMA-based models such as SARIMA and spatiotemporal ARIMA have limitations in describing variations of traffic data. Run et al. [18] compared the performance of ARIMA with that of prediction models for subway traffic forecasting. Benefiting from the development of artificial intelligence (AI), deep-learning models have raised great concern. Recurrent Neural Networks (RNNs) [19] are typical models which show advantages in modeling time series data. Azad and Islam [20] introduced the prediction model using Google Maps which took into account real-time traffic information to make accurate predictions. Shu et al. [21] presented an improved GRU for short-term traffic flow prediction which used an attention mechanism to capture the temporal features of traffic data. Also, some researchers deployed CNN-based models to consider the spatial correlation [22]. GNN-based models have been widely applied by some researchers in recent years. Buapang and Muangsin [23] proposed a spectral GNN prediction model that utilized the graph theory to capture the spatial correlation of traffic networks. However, the above-mentioned prediction techniques only consider the temporal or spatial correlation, without simultaneously capturing the spatiotemporal dependencies from historical traffic data.

To fully capture the spatiotemporal correlations of data, some researchers proposed hybrid models for traffic prediction. Zhu et al. [24] proposed a knowledge-driven spatiotemporal GCNN model (KST-GCN) for traffic prediction. The proposed model uses spatial and temporal information to construct a traffic network graph, which is then fed into the GCNN. Also, Chen et al. [25] proposed a deep learning-based model for predicting the traffic flow in the Internet of Vehicles (IoV). The feature extractor is based on the CNN and LSTM, while the prediction module is based on the LSTM only. Moreover, spatiotemporal prediction models such as LSTM-GL-REMF [26], Traffic Graph Convolutional Recurrent Neural Network (TGC-LSTM) [11], Graph Convolutional Neural Network with Data-Driven Graph Filter (GCN-DDGF) [27] have shown superiorities in traffic state prediction. It has been proved by the literature that hybrid models can model spatiotemporal features. However, the graph information remains static in the model training process which cannot accurately capture the spatiotemporal correlations of traffic conditions.

Further, the attention mechanism has raised great concern in recent years due to its advantages in fully capturing the spatiotemporal correlations. Xie et al. [28] proposed a dynamic spatiotemporal relation graph to predict the subway flow, and apply a long-term prediction module based on the transformer. Zhang et al. [29] developed a STRGAT model that applied the deep residual attention module to capture the dynamic spatial characteristics. Moreover, Abdelraouf et al. [30] proposed an attention-based network for freeway traffic speed prediction

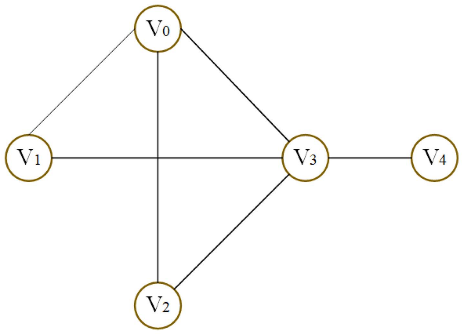


Fig. 1. An example of road network.

which uses multiple encoders and a decoder to extract spatiotemporal correlations from historical traffic data and generate the predicted traffic speed respectively.

Attention-based deep learning models can accurately consider the complex relationship of traffic data. For example, a hierarchical attention GCN fusing multi-sensor signals [31] was proposed for remaining useful life prediction. The model used a hierarchical attention mechanism to extract important features from multi-sensor signals, which is further fused into a GNN. A novel spatiotemporal self-attention network [32] was proposed for video saliency prediction. Duan et al. [33] proposed a fully dynamic self-attention spatiotemporal graph network that uses a fully dynamic self-attention mechanism to capture the spatiotemporal correlations of traffic flow data. Similarly, self-attention bi-LSTM network [34], dual attention-based federated learning approach [35] and GraphSanet [36] use a self-attention mechanism to capture the spatiotemporal features in observed data. However, these methods can only capture spatial or temporal dimensions of traffic information because the input and output of the attention mechanism are two sequences of factors.

III. Preliminaries

The road network can be considered as a directed graph $\mathcal{G} = (\mathcal{V}, \mathcal{E}, \mathcal{A})$, where $\mathcal{V}$ is a collection of vertices which represents the road sections in the network, $\mathcal{E}$ represents the connection states among road sections and $N = |\mathcal{V}|$ denotes the number of road sections; $A \in \mathbb{R}^{N\times N}$ is the binary adjacency matrix with zero entries on the diagonal, where $A_{i,j}$ denotes the physical connection between road section i and road section j ($A_{i,j} = 1$ represents there is a connection, while $A_{i,j} = 0$ denotes that there is no connection, and $A_{i,i} = 0$). For example, Fig. 1 shows the connectivity of the five vertices where the solid line indicates that there is a connection between vertices, and the corresponding adjacency matrix can be described by:

$$A = \begin{bmatrix} 0 & 1 & 1 & 1 & 0 \\ 1 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 & 0 \\ 1 & 1 & 1 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 \end{bmatrix} \tag{1}$$

The traffic conditions over T time steps can be considered as a two-dimensional matrix $X \in \mathbb{R}^{N\times T}$ on graph $\mathcal{G}$ and the traffic condition at time t can be represented as $X_t \in \mathbb{R}^{N\times 1}$.

Problem Studied: As shown in Fig. 2, given the observed data of the historical T time steps $\mathcal{X}^{history} \in \mathbb{R}^{N\times T}$, we aim to build a deep-learning model being able to predict the traffic data of the next future P time steps by fully extracting the spatiotemporal features from the historical data, which can be shown as:

$$\mathcal{Y} = \mathcal{F}(X^{history}) \tag{2}$$

where $\mathcal{F}$ denotes the function of our proposed model.

Spatiotemporal correlations of traffic information: For a better understanding of our paper, here we give an explanation of the spatiotemporal correlations of traffic information among different road sections. That is, the traffic data of one road section at a certain time step is correlated with the historical observations of this road section and the traffic data of other road sections in this road network [12].

IV. Adaptive Temporal Graph Attention Network

Fig. 3 illustrates the structure of the ATGAN model, where the outlier time series filter (OTSF) is applied to smooth the outlier data points in the historical observations, and the group attention GCN model and GRU model are applied to fully model the spatiotemporal features from the filtered data and achieve the prediction results. In addition, we apply the structure of the Residual Network (ResNet) to ensure the stability of the model training process [37].

A. Outlier Time Series Filter

The real-world traffic data is usually non-stationary and has some outlier points over the time domain, which may negatively affect the model parameters estimation in the model training process [38]. To overcome this limitation, we propose the Outlier Time Series Filter (OTSF) technique by adding a learnable bias to the historical observations which can adaptively smooth the outlier time series with the attention mechanism.

Fig. 4 presents the framework of the proposed OTSF technique. Firstly, we take the adjacency matrix $A \in \mathbb{R}^{N\times N}$ to describe the topological structure of the road network. Based on the adjacency matrix, we employ the attention score to calculate the spatial correlation coefficient between each road section:

$$G = \frac{\langle \mathrm{ReLU}(W_q A), \mathrm{ReLU}(W_k A)\rangle}{\sqrt{D}} \tag{3}$$

$$\alpha_{i,j} = \frac{\exp(G_{i,j})}{\sum_{i=1}^{N}\sum_{j=1}^{N}\exp(G_{i,j})} \tag{4}$$

where $\langle \bullet, \bullet \rangle$ represents the inner product operator, $G \in \mathbb{R}^{N\times N}$ is the attention score of the adjacency matrix, ReLU is the activation function which is used to maintain positive correlations between sensors, W_q and W_k are two trainable model parameters, $G_{i,j}$ represents the *i*-th value of the *j*-th column of attention score matrix G, D represents the dimension of A, α is the spatial correlation matrix.

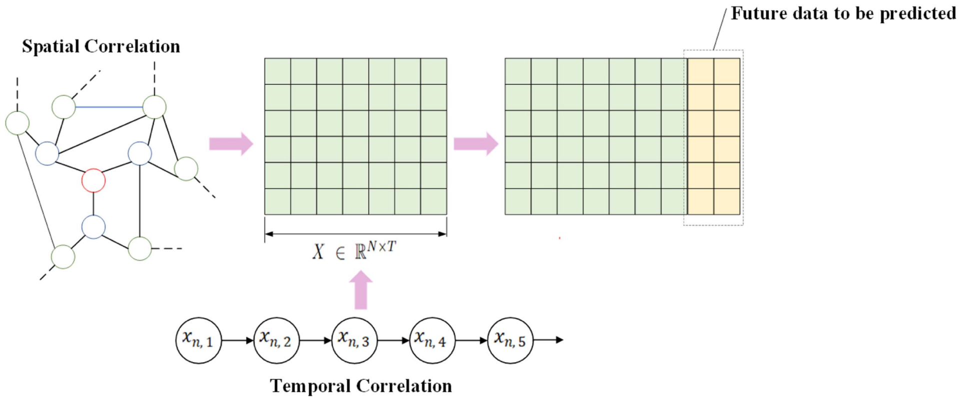


Fig. 2. Problem formulation.

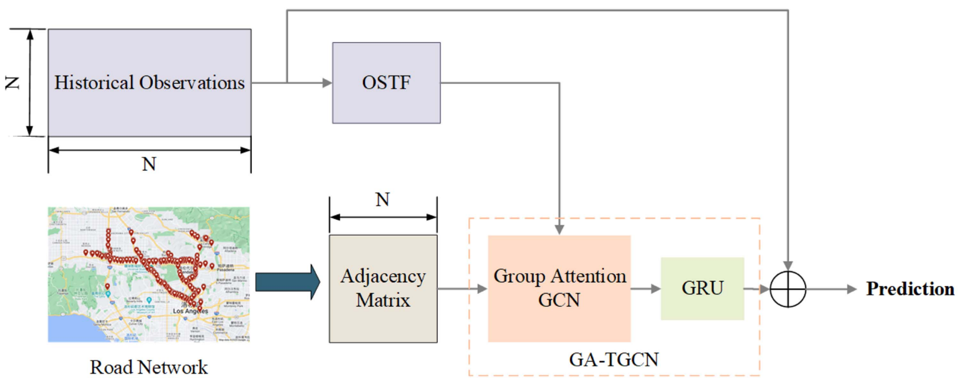


Fig. 3. Structure of ATGAN.

As the spatial correlation between a sensor and itself is always 1, the spatial correlation coefficient matrix S can be obtained by:

$$S = \alpha + I_N \tag{5}$$

where I_N is a N-dimensional identity matrix.

In the spatial correlation coefficient matrix S, we consider each value $S_{m,n}$ as the spatial correlation coefficient between road section m and road section n. Based on S, we construct an adaptive bias matrix based on the historical information and adjacency matrix. The procedure of bias construction is shown below.

First, we extract the traffic information for each time step and separate the correlation coefficient matrix S to N vectors:

$$X = [X_1, X_2, \ldots, X_T], X_t \in \mathbb{R}^{1\times N}(t = 1, 2, \ldots, T) \tag{6}$$

$$S = [S_1, S_2, \ldots, S_n], S_i \in \mathbb{R}^{1\times N}(i = 1, 2, \ldots, N) \tag{7}$$

Then, we construct a learnable bias B that fully utilizes the spatial correlation of traffic data. That is, according to the historical observations X and the correlation coefficient matrix S, we construct spatial correlation-based bias B using the weighted sum of traffic data of each road section and the spatial correlation coefficient vector of the corresponding road section:

$$B_{i,j} = (X_{j,1}, X_{j,2}, \ldots, X_{j,N}) \cdot (S_{i,1}, S_{i,2}, \ldots, S_{i,N})^T \tag{8}$$

where $B_{i,j}$ represents the *i*-th value the *j*-th column of B.

In matrix S, we use S_i to represent the spatial correlation between road section i and the other target road sections. Hence, we refine (8) as:

$$B_{i,j} = \sum_{m=1}^{N} X_{j,m} S_{i,m} \tag{9}$$

After the construction of the trainable bias, we use the sum of historical data X and bias B as the training data I in the following model training process:

$$I_{i,j} = X_{i,j} + B_{i,j} \tag{10}$$

By applying OSTF, the variance of historical data can be reduced by equally sharing the fluctuations caused by the outliers. Specifically, the procedure of OSTF can be found in Algorithm 1.

B. Group Attention Temporal Graph Convolutional Network

In this paper, we design the GA-TGCN model to extract the spatiotemporal features of traffic information. Fig. 5 shows the structure of the GA-TGCN model, which applies a group attention GCN (GA-GCN) to capture the spatial dependence and a GRU model to capture the temporal dependence of the traffic information.

Group attention GCN: To fully model the dynamic spatial features among neighboring road sections, we propose a Group

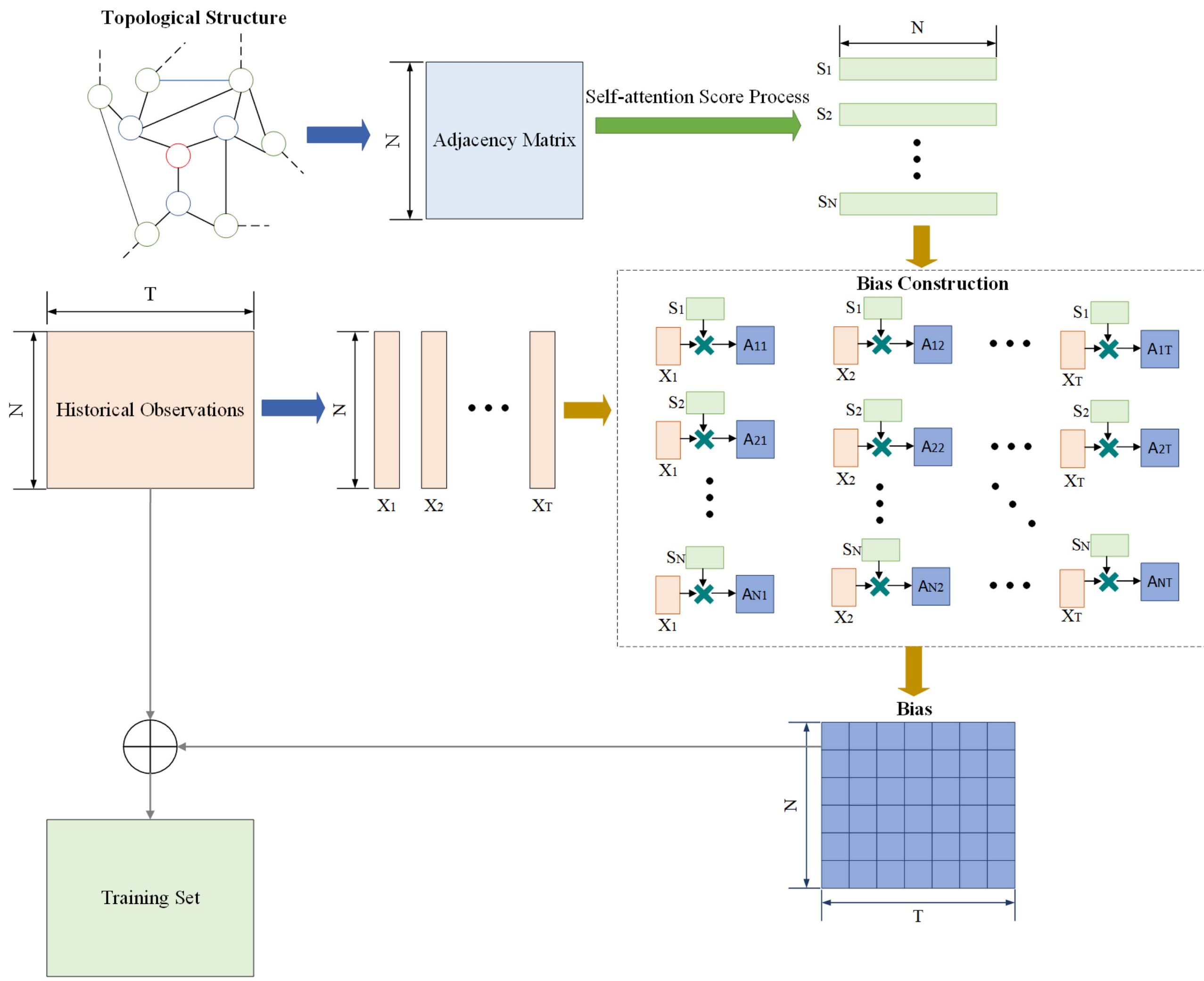


Fig. 4. The framework of OTSF technique.

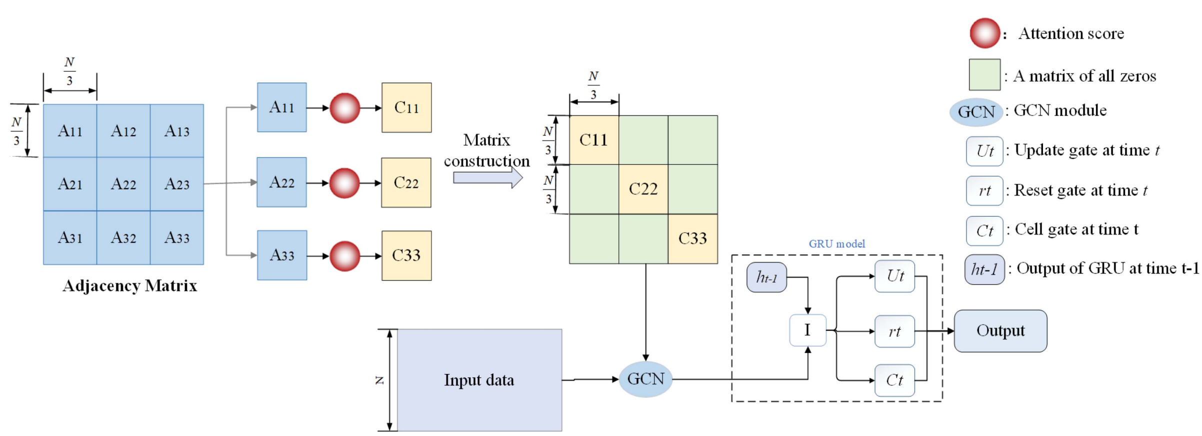


Fig. 5. Structure of GA-TGCN (when k=3).

Algorithm 1: Procedure of Outlier Time Series Filter.

Input: Original traffic data matrix $X \in \mathbb{R}^{N\times T}$; Adjacency matrix $A \in \mathbb{R}^{N\times N}$;
Output: Filtered training data I;
Initialize two parameters W_q, W_k for adjacency matrix.
Extract data of each time step from X.
$G \leftarrow \frac{\langle \mathrm{ReLU}(W_q \times A), \mathrm{ReLU}(W_k \times A)\rangle}{\sqrt{D}}$.
$\alpha \leftarrow \frac{\exp(G_{i,j})}{\sum_{i=1}^{N}\sum_{j=1}^{N}\exp(G_{i,j})}$.
$S \leftarrow \alpha + I_N$.
for $i = 0$ to T **do**
  for $j = 0$ to N **do**
    $B_{i,j} \leftarrow \sum_{m=1}^{N} X_{j,m} S_{i,m}$.
    $I_{i,j} \leftarrow X_{i,j} + B_{i,j}$.
  end for
end for
Get training data I based on X, A.
return I;

Attention Graph Convolutional Network (GA-GCN) which considers the spatial correlation of traffic data as a learnable parameter during model training. By applying the group attention mechanism, the spatial correlation coefficient among multi-located sensors can be updated in each training epoch, which can further extract the spatial correlation from historical data and reduce the error propagation caused by incorrect spatial correlation representation during a deep learning model training process. The process of GA-GCN is shown in the following contexts.

When the number of road sections N becomes large, the computational complexity goes high when we directly model the global road network because we need to compute N^2 attention scores. In addition, when modeling the global road network, the correlation among road sections that are geographically far apart is also calculated, and the focus of attention may be shifted which will dilute the spatial correlation among the adjacent road sections, thus the prediction performance can be affected. To address these issues, we model the spatial features of traffic data which focuses on the correlation among the neighboring road sections.

As shown in Fig. 5, we first partition N road sections into K groups (K=3 in Fig. 5 as an example) based on the adjacency matrix A, where each group contains N/K sensors and it should be noted that N/K should be a positive integer, shown as:

$$A = \begin{bmatrix} A_{1,1} & A_{2,1} & \cdots & A_{k,1} \\ A_{1,2} & A_{2,2} & \cdots & A_{k,2} \\ \vdots & \vdots & \ddots & \vdots \\ A_{1,k} & A_{2,k} & \cdots & A_{k,k} \end{bmatrix} \tag{11}$$

To enhance the correlation of road sections, we ignore the impact of those low-correlation road sections and calculate the correlation of the road segments within each group, so that we only calculate the spatial correlation coefficient grouped on the diagonal of the adjacency matrix (for example, in Fig. 5, such groups are $A_{1,1}$, $A_{2,2}$ and $A_{3,3}$). According to these groups, we employed the scaled dot-product attention approach to compute the spatial correlation of the traffic data among the corresponding road sections:

$$\beta_{i,i} = \frac{ReLU(W_m^i A_{i,i}) \cdot ReLU(W_n^i A_{i,i})}{|W_m^i A_{i,i}||W_n^i A_{i,i}|} \tag{12}$$

$$G_{i,i} = \frac{\exp(\beta_{i,i})}{\sum_{i=1}^{\frac{N}{k}}\sum_{j=1}^{\frac{N}{k}}\exp(\beta_{i,i})} \tag{13}$$

where $A_{i,i} \in \mathbb{R}^{\frac{N}{K}\times\frac{N}{K}}$ ($i = 1, 2, \ldots, k$) represents the subgroups on the diagonal of the adjacency matrix A, W_m^i and W_n^i are learnable parameters allocated for $A_{i,i}$ to achieve attention scores, and $G_{i,i}$ are spatial correlation coefficient matrix of $A_{i,i}$.

Then, we construct the coefficient correlation matrix of the global road network C by splicing $G_{i,i}$ into the correlation coefficient matrix:

$$C = \begin{bmatrix} G_{1,1} & O & \cdots & O \\ O & G_{2,2} & \cdots & O \\ \vdots & \vdots & \ddots & \vdots \\ O & O & \cdots & G_{k,k} \end{bmatrix} \tag{14}$$

where $O \in \mathbb{R}^{\frac{N}{K}\times\frac{N}{K}}$ is an all-zeros matrix.

Hence, the learnable correlation coefficient matrix of the road can be constructed, wherein each road section focuses on the spatial correlation between itself and that of the neighboring road sections.

Based on the coefficient matrix C and input data I from the OTSF filter (in Section IV-A), we use a GCN model to extract the spatial features. Here, we first estimate the matrix with added self-connections $\widetilde{C}$ and the degree matrix $\widetilde{D}$:

$$\widetilde{C} = C + I_N \tag{15}$$

$$\widetilde{D} = \sum_j \widetilde{C_{i,j}} \tag{16}$$

where $\widetilde{C_{i,j}}$ denotes the value at the *j*-th column and the *i*-th row in $\widetilde{C}$.

Then, the GCN network can be built by stacking multiple convolutional layers. In this paper, we use two layers of GCN structure to improve the prediction performance:

$$O = \sigma(\widetilde{S}\mathrm{ReLU}(\widetilde{S} IW_1)W_2) \tag{17}$$

where σ is the sigmoid function, $\widetilde{S} = \widetilde{D}^{-\frac{1}{2}}\widetilde{C}\widetilde{D}^{-\frac{1}{2}}$ represents the pre-processing step, W_1 and W_2 are the weighting matrices allocated for the two layers, and O denotes the output of the GA-GCN model.

Gated Recurrent Unit: In GA-TGCN, we apply a GRU model to extract the temporal features of traffic data. GRU network is an extension of the Recurrent Neural Network (RNN), which has advantages in modeling the temporal features of spatiotemporal data. From Fig. 5, it can be seen that in GRU, the cell gate is used to store the data of the current time step, and the update gate is applied to control the extent to which state information from the previous moment is transmitted to the current state, and the reset gate is applied to capture short-term correlation in time series.

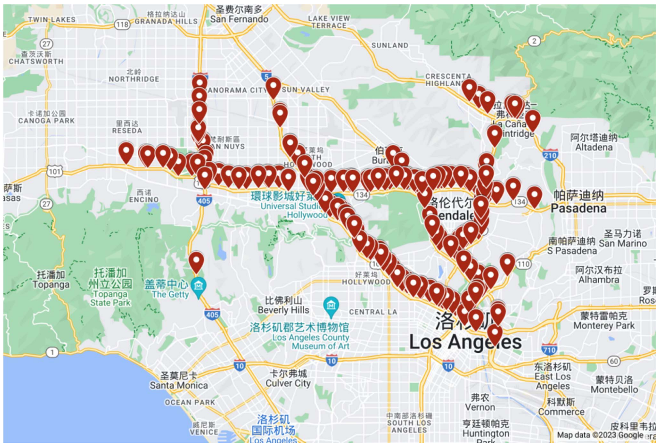


Fig. 6. Locations of sensors in Los-loop datatset.

Given the input data I_t at time t, the specific process of GRU is shown as follows, where u_t, r_t, c_t are the state of update gate, cell gate and reset gate, W_u, W_r, W_r and b_u, b_r, b_r are the weights and bias for GRU, respectively:

$$u_t = \sigma(W_u[I_t, h_{t-1}] + b_u) \tag{18}$$

$$r_t = \sigma(W_r[I_t, h_{t-1}] + b_r) \tag{19}$$

$$c_t = tanh(W_c[I_t, (r_t \cdot h_{t-1})] + b_c) \tag{20}$$

$$h_t = u_t * h_{t-1} + (1 - u_t) * c_t \tag{21}$$

C. Model Training Process

As shown in Fig. 3, ATGAN first utilizes an OTSF filter to smooth the randomly distributed traffic data and reduce the negative impact of outlier data points in the model training process, then applies a GA-TGCN model to extract the spatiotemporal features of traffic data. During training, we update all the model parameters in the ATGAN model through back-propagation by minimizing the loss between the prediction results and real data:

$$L = ||Y_t - \widehat{Y_t}|| + \lambda L_{reg} \tag{22}$$

where Y_t represents the real traffic data, $\widehat{Y_t}$ is the predicted value, λ is the hyper-parameter, and L_{reg} is the L2 regularization term.

V. Experiments and Discussion

A. Data Description

We evaluate the prediction performance of the ATGAN model with two real-world datasets: the Los-loop dataset and the Seattle-loop dataset.

1) Los-loop dataset: The dataset provides valuable information about traffic speed in Los Angeles between March 1–7, 2012. The data was gathered from 207 sensors located throughout the county (locations of sensors can be seen in Fig. 6), and the sampling time interval is 5 minutes. The dataset consists of two parts. The first part is a two-dimensional matrix that denotes the speed data over all time steps, wherein each row represents the traffic data of each sensor at all time instants, and each column represents the speed data of all sensors at a single time step.

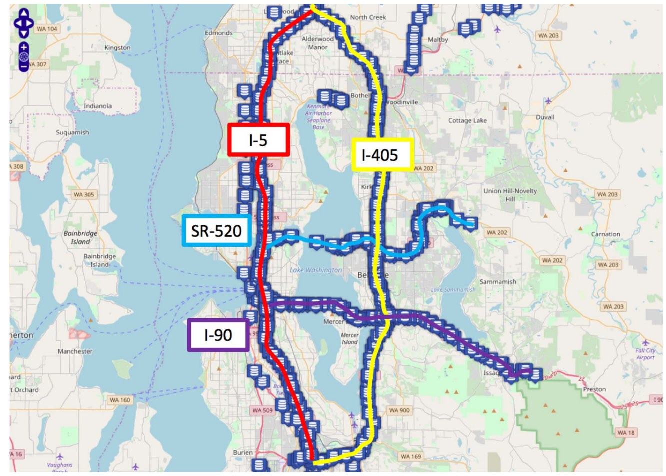


Fig. 7. Locations of sensors in Seattle-loop datatset.

2) Seattle-loop dataset: The dataset describes the traffic data collected from the loop sensors in the Seattle area in 2015 (the locations of sensors can be seen in Fig. 7), where the sampling time interval is 5 minutes. The dataset contains two parts. The first part is a two-dimensional matrix that represents the spatiotemporal information of traffic speed data from multi-located sensors on the main stems at different time steps. The second part is an adjacency matrix which describes the physical connections among sensors in the road network.

B. Computing Environment

The detailed setting of the computing environment of this paper is shown as follows. The experiments are conducted using a server with 16 CPU cores (Intel i7 13700k) and one GPU (RTX 3060). The version of Python is 3.7, we use Scikit-learn and Tensorflow for the network construction.

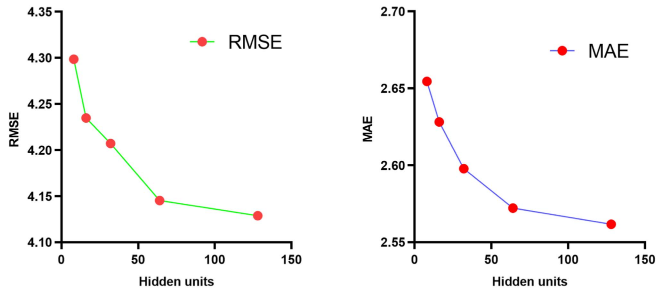
C. Metrics

In this paper, to validate the superiority of our model, we apply the same as the metrics in [12], i,e., *Root Mean Square Error* (RMSE), *Mean Square Error* (MAE), *Accuracy* (ACC), *Coefficient of Determination* (R^2) and *Explained Variance Score* (VAR). The metrics are computed by the following equations, where $y_{i,j}$ and $y_{i,j}^{pred}$ represent the real traffic data and the prediction value of the *j*-th time step in the *i*-th road respectively, where f_{var} refers to the variance function:

$$RMSE = \sqrt{\frac{1}{MN} \cdot \sum_{j=1}^{M} \sum_{i=1}^{N} (y_{i,j} - y_{i,j}^{pred})^2} \tag{23}$$

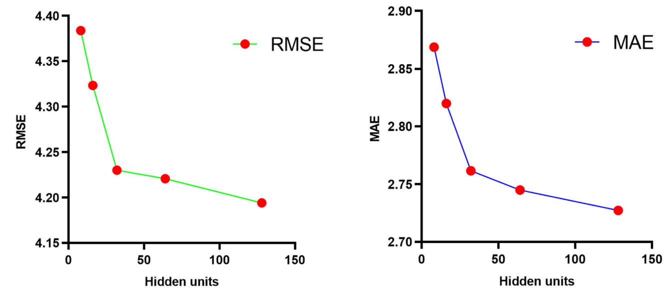
$$MAE = \frac{1}{MN} \cdot \sum_{j=1}^{M} \sum_{i=1}^{N} |y_{i,j} - y_{i,j}^{pred}| \tag{24}$$

$$ACC = 1 - \frac{||y_{i,j} - y_{i,j}^{pred}||_F}{||y_{i,j}||_F} \tag{25}$$



(a) (b)

Fig. 8. Changes of prediction accuracy with Los-loop dataset. (a) Changes in RMSE. (b) Changes in MAE.



(a) (b)

Fig. 9. Change of prediction accuracy with Seattle-loop dataset. (a) Changes in RMSE. (b) Changes in MAE.

$$R^2 = 1 - \frac{\sum_{j=1}^{M}\sum_{i=1}^{N}(y_{i,j} - y_{i,j}^{pred})^2}{\sum_{j=1}^{M}\sum_{i=1}^{N}(y_{i,j} - \bar{Y})^2} \tag{26}$$

$$VAR = 1 - \frac{f_{var}(Y - Y^{pred})}{f_{var}(Y)} \tag{27}$$

D. Baselines

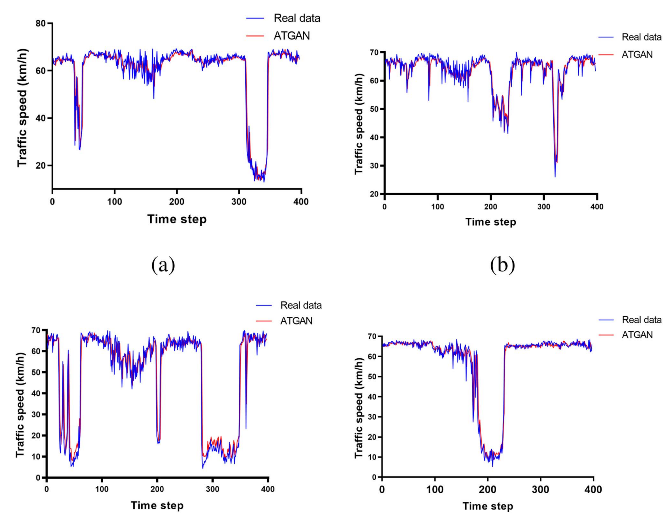
In this paper, we compare the prediction performance of ATGAN with the following baseline methods: (1) Auto-regressive integrated moving average (**ARIMA**) [39]; (2) Historical average (**HA**) [40] (3) Support vector regression (**SVR**) [41]; (4) Long short-term memory (**LSTM**) [42]; (5) Temporal graph convolutional network (**T-GCN**), which is a hybrid model that combined with GCN and GRU [43]; (6) Attention temporal graph convolutional network (**A3T-GCN**), which adds a self-attention layer on the output of T-GCN network to capture the global trends of temporal correlation of traffic data [13]; (7) Traffic graph convolutional long short-term memory neural network (**TGC-LSTM**) [11], which combines GCN with LSTM and uses an arrival matrix to replace the adjacency matrix.

E. Prediction Performance

The number of hidden units is an important hyper-parameter in deep-learning models. Hence, to evaluate the universality of ATGAN, we set the number of hidden units to **8, 16, 32, 64, 128**, respectively and test the prediction accuracy of ATGAN given by RMSE and MAE with the two real-world datasets.

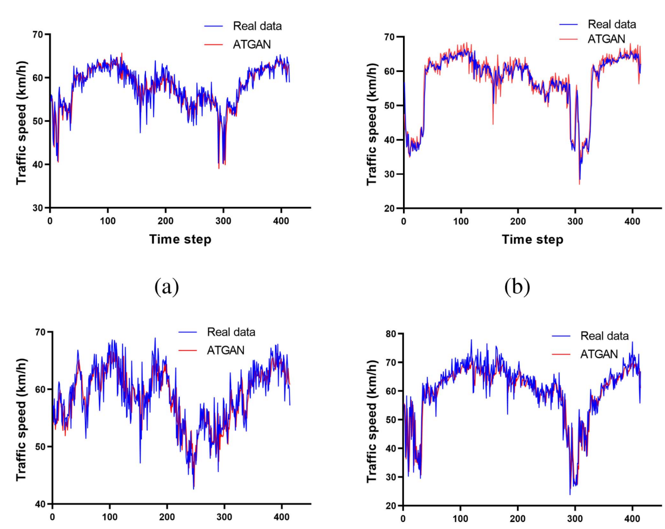
As shown in Figs. 8 and 9, the prediction accuracy of ATGAN increases with the larger number of hidden units, but the increasing rate becomes smaller when the number reaches 64. In addition, the prediction accuracy remains stable under different settings of hidden units, which validates the effectiveness and stability of our model.

In real-world datasets, various road sections may exhibit distinct traffic conditions at varying time steps. Hence, we randomly selected data from four road sections for testing to adequately verify the effectiveness of our proposed model. In particular, upon configuring the number of hidden units to 64, the prediction outcomes of ATGAN using the two datasets are illustrated in Figs. 10 and 11, respectively.



(a) (b) (c) (d)

Fig. 10. Prediction results tested with the Los-loop dataset. (a) Road section 1. (b) Road section 2. (c) Road section 3. (d) Road section 4.



(a) (b) (c) (d)

Fig. 11. Prediction results tested with the Seattle-loop dataset. (a) Road Section 1. (b) Road section 2. (c) Road section 3. (d) Road section 4.

TABLE I
PREDICTION PERFORMANCE BASED ON THE LOS-LOOP DATASET

Method	ATGAN	TGC-LSTM	A3T-GCN	T-GCN	LSTM	SVR	HA	ARIMA
$RMSE$	**4.1454**	4.8700	4.8360	5.0200	9.6028	11.1333	7.3067	10.0625
MAE	**2.5722**	3.1656	3.2037	3.3667	6.7561	7.1441	3.8782	7.7306
ACC	**0.9294**	0.9171	0.9077	0.9146	0.8026	0.8106	0.8756	0.8272
R^2	**0.9098**	0.8755	0.9077	0.8677	0.4935	0.3495	0.7225	*
VAR	**0.9098**	0.8756	0.8773	0.8702	0.5078	0.3514	0.7225	0.0015

TABLE II
PREDICTION PERFORMANCE BASED ON THE SEATTLE-LOOP DATASET

Method	ATGAN	TGC-LSTM	A3T-GCN	T-GCN	LSTM	SVR	HA	ARIMA
$RMSE$	**4.2207**	4.3812	5.0562	4.6907	8.3794	10.7602	7.3067	10.8712
MAE	**2.7449**	2.9530	3.4987	3.2031	5.3749	7.0126	6.8031	8.5355
ACC	**0.9267**	0.9240	0.9123	0.9186	0.8373	0.8134	0.8821	0.8097
R^2	**0.8832**	0.8736	0.8317	0.8551	0.5208	0.2377	0.6966	*
VAR	**0.8833**	0.8736	0.8317	0.8562	0.5210	0.2532	0.6967	*

Both the two figures illustrate the prediction results obtained by ATGAN are in good line with the real data, which indicates the effectiveness of ATGAN. Specifically, when the real data are stationary, the prediction results are more precise. When the data has some large fluctuations, the results will deviate from the real data to some extent, i.e., the results from time step 270 to time step 330 in Fig. 11. Moreover, the curve of predicted results is smoother than that of real data. This is mainly because the GCN-based models use a smooth filter to capture the topological structure by moving the filter, leading to a smaller change compared with the real data.

F. Comparison

Prediction performance comparison: To demonstrate the superiority of our model, we compare the prediction performance of ATGAN with the baseline methods. The setting of hyperparameters is listed as follows: the number of hidden units is 64, the learning rate is set as 0.001, the batch size is 33, and the number of epochs for model training is 300.

By the evaluation metrics given in Section V-B, we first compare the prediction accuracy based on the Los-loop dataset.

The results of the accuracy comparison are shown in Table I, it can be seen that ATGAN achieves the highest prediction accuracy among all the baseline methods. In addition, within all tested methods, hybrid deep learning-based models including ATGAN, TGC-LSTM, A3T-GCN, and T-GCN perform far better than other baseline methods. This is mainly because the hybrid models extract both spatial and temporal features of traffic data while the single prediction models such as LSTM and SVR only capture one scale of feature (temporal feature or spatial feature) so that the hybrid models can better couple the spatiotemporal correlations and achieve higher prediction accuracy. Specifically, among all hybrid prediction models, RMSE of ATGAN are improved by approximately 14.9%, 14.3%, and 17.4% compared with TGC-LSTM, A3T-GCN, and T-GCN; MAE of ATGAN are improved by approximately 18.7%, 19.7% and 23.6% compared with TGC-LSTM, A3T-GCN, and T-GCN.

Then, we compare the prediction accuracy of ATGAN with that of baseline methods based on the Seattle-loop dataset, and the results are shown in Table II. Similarly, ATGAN obtains the highest prediction accuracy among baseline methods and the prediction accuracy of hybrid deep learning-based models is higher than the single prediction models. Also, we mainly focus on the comparison of prediction accuracy between ATGAN and hybrid models. We find that RMSE of ATGAN is improved by approximately 3.6%, 16.5%, and 10.0% compared with TGC-LSTM, A3T-GCN, and T-GCN; MAE of ATGAN are improved by approximately 7.0%, 21.5% and 14.3% compared with TGC-LSTM, A3T-GCN, and T-GCN.

Furthermore, to demonstrate the superiority of ATGAN in predicting traffic data with outlier time series, we extract the periods of data with large fluctuations and compare the prediction performance of ATGAN with the hybrid prediction models including TGC-LSTM, A3T-GCN, and T-GCN.

As shown in Fig. 12, we find that compared with the baseline methods, prediction results obtained from ATGAN maintain a higher degree of agreement with real data with the four periods of data with several large fluctuations selected from the two datasets. Further, all methods predict poorly at the local minimum/maximum, but the results obtained by ATGAN have the highest approximation with the real data, which evaluates the effectiveness of ATGAN in predicting non-stationary traffic data.

Computational complexity comparison: Efficient traffic prediction can greatly improve the efficiency of ITS, so the computational complexity is an important factor to be measured. In this paper, we test the computational complexity of ATGAN in terms of the computational time and space complexity for model training.

First, we test the computational time as well as the prediction accuracy under different numbers of hidden units. As shown in Table III, it can be seen that for both two datasets, a larger number of hidden units can lead to better prediction performance with some sacrifices in computational efficiency. Hence, it is necessary to strike a balance between accuracy and computational time to ensure both effectiveness and efficiency.

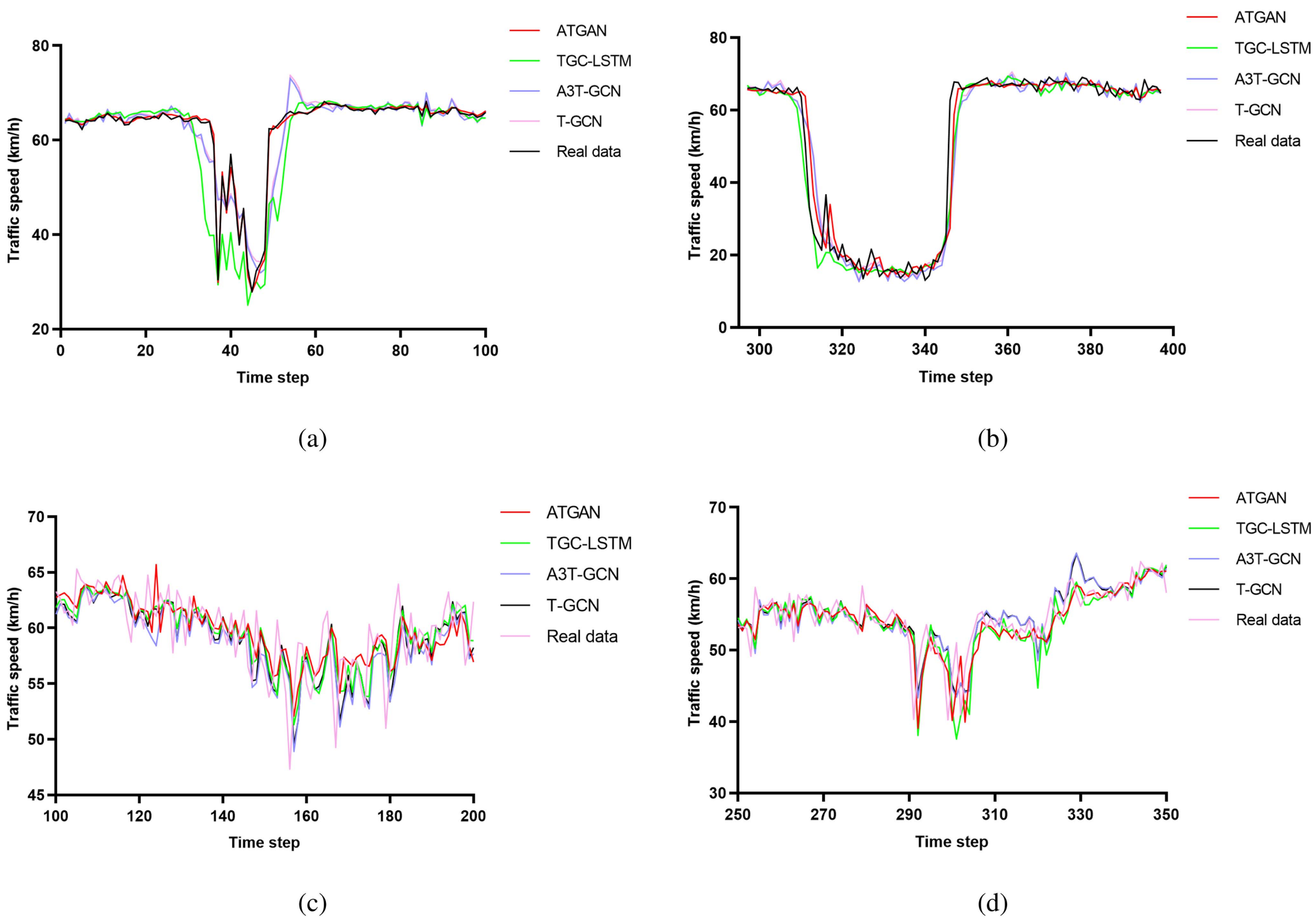


Fig. 12. Prediction performance comparison of traffic data with outlier time series. (a) Los-loop dataset from time step 1 to time step 100. (b) Los-loop dataset from time step 297 to 397. (c) Seattle-loop dataset from time step 100 to time step 200. (d) Seattle-loop dataset from time step 250 to 350.

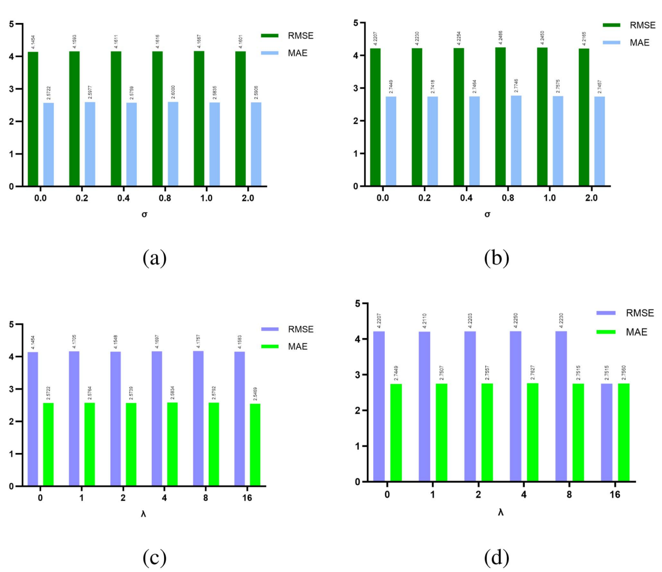


Fig. 13. Adding noise. (a) Adding Gaussian noise to the Los-loop dataset. (b) Adding Gaussian noise to the Seattle-loop dataset. (c) Adding Poisson noise to the Los-loop dataset. (d) Adding poisson noise to the Seattle-loop dataset.

Then, we compare the computational time of ATGAN with that of baseline methods, it should be mentioned that the numbers of hidden units in ATGAN are set to 8, 16, 32, and 64, respectively and the number of hidden units in baseline methods is set to 64, as shown in Table IV. From Tables I and II, it is easy to find that hybrid prediction models including ATGAN, TGC-LSTM, A3T-GCN, and T-GCN performs far better than the other baseline methods. Hence, although it shows that the computational time of single prediction models including LSTM, SVR, ARIMA, and HA is much lower than that of hybrid prediction models, we mainly conduct a comparison of computational complexity among these four hybrid models.

TABLE III
COMPUTATIONAL TIME AND PREDICTION ACCURACY OF ATGAN UNDER DIFFERENT SETTINGS OF HIDDEN UNITS

Hidden units	Los-loop dataset			Seattle-loop dataset		
	Time/s	RMSE	MAE	Time/s	RMSE	MAE
8	797	4.2987	2.6546	1558	4.3838	2.8687
16	1050	4.2349	2.6282	2204	4.3235	2.8200
32	1839	4.2987	2.5978	3867	4.2300	2.7616
64	4353	4.1454	2.5722	7812	4.2207	2.7449
128	7440	4.1289	2.5617	14139	4.1941	2.7273

TABLE IV
COMPUTATIONAL TIME COMPARISON

Methods	Los-loop dataset (s)	Seattle-loop dataset (s)
$ATGAN(8units)$	797	1558
$ATGAN(16units)$	1050	2204
$ATGAN(32units)$	1839	3867
$ATGAN(64units)$	4353	7812
$TGC-LSTM$	2249	3116
$A3T-GCN$	2671	3784
$T-GCN$	3131	3270
$LSTM$	1851	1851
SVR	17	24
HA	0.7	1.1
$ARIMA$	52	86

TABLE V
SPACE COMPLEXITY OF ATGAN UNDER DIFFERENT SETTINGS OF HIDDEN UNITS

Hidden units	Los-loop dataset		Seattle-loop dataset	
	$FLOPs$/G	$n_{parameters}$	$FLOPs$/G	$n_{parameters}$
8	0.0061	9.77k	0.0224	23.15k
16	0.0061	10.40k	0.0224	23.78k
32	0.0061	12.82k	0.0224	26.20k
64	0.0062	22.26k	0.0225	35.63k

When we set the number of hidden units to the same value, the computational time of ATGAN is higher than the other hybrid prediction models. However, combining Tables I, II, and IV, it can be seen that when the number of hidden units in ATGAN is small (i.e., 8, 16, 32), the prediction accuracy is still higher than the other baselines (the number of hidden units in baseline methods is 64) but the computational time is lower, which validates the efficiency of ATGAN as well.

Space complexity is also one of the important factors to evaluate computational complexity. Referred to [44], we utilize floating point operations ($FLOPs$) and numbers of parameters ($n_{parameters}$), which are applied in many studies, as metrics for calculating space complexity.

Initially, we assess the space complexity of ATGAN across various configurations of hidden units. As demonstrated in Table V, it is evident that an increase in the number of hidden units correlates with a rise in both the number of floating point operations per second ($FLOPs$) and the number of parameters. Notably, when the unit is G, the increase in $FLOPs$ remains marginal. This observation suggests that, despite the growing number of hidden units, our model does not impose a significant additional burden on computational resources.

TABLE VI
SPACE COMPLEXITY COMPARISON BETWEEN ATGAN AND HYBRID BASELINES (64 HIDDEN UNITS)

Methods	Los-loop dataset		Seattle-loop dataset	
	$FLOPs$/G	$n_{parameters}$	$FLOPs$/G	$n_{parameters}$
ATGAN	0.0062	22.26k	0.0225	35.63k
TGC-LSTM	0.003	15.17k	0.013	19.20k
A3T-GCN	0.005	16.51k	0.023	23.37k
T-GCN	0.002	12.74k	0.008	15.56k

Typically, the space complexity of hybrid deep-learning models tends to exceed that of single prediction models. Meanwhile, in light of the imperative to maintain high prediction accuracy, we conduct a comparison of the space complexity of ATGAN against several hybrid prediction methods including TGC-LSTM, A3T-GCN, and T-GCN. As illustrated in Table VI, we find that when the hidden units are set as 64, the space complexity of ATGAN is higher than that of the baseline methods without a significant disparity in magnitude. However, similar to the comparison performance in computational time, it becomes evident that for hidden unit settings of 8, 16, and 32, the predictive performance of ATGAN consistently outperforms that of the hybrid baseline methods (the hidden units of baseline methods is set as 64), but the number of hidden units of ATGAN is lower in most cases, underscoring the exceptional efficiency of our proposed model.

In conclusion, despite the small number of hidden units, the prediction accuracy of ATGAN can remain at a high level, which is higher than the recent hybrid prediction models such as TGC-LSTM and T-GCN. That is, at similar prediction accuracy, the computational time is lower than the baseline methods, and the space complexity and the space complexity of our model will not significantly increase, which also indicates the efficiency of ATGAN.

G. Robustness

The real-world datasets usually contain noise which is generated in the data-collecting process. Hence, robustness is one of the important measurements to evaluate the effectiveness of the prediction model. In this paper, we add two types of common noise to the historical observations and test the robustness of ATGAN, respectively. First, we add a Gaussian-distributed noise which obeys $N \in (0, \sigma^2)$, where $\sigma \in (0.2, 0.4, 0.8, 1, 2)$. Then, we add the Poisson-distributed noise $P(\lambda)$, where $\lambda \in (1, 2, 4, 8, 16)$. It should be mentioned that the value of the noise matrices is normalized to be between 0 and 1.

The prediction accuracy of the robustness test given by RMSE and MAE for ATGAN is shown in Fig. 13. From Fig. 13, it can be seen that the change in prediction accuracy is small when we add such two different distributions of noise to the historical data, which proves that the ATGAN model is robust and can solve the issues caused by different types of noise.

VI. Conclusion

In this paper, we introduce a novel approach called Adaptive Temporal Graph Attention Network (ATGAN) to predict future traffic data with consideration of outlier data points. Specifically, ATGAN is implemented in two stages: (1) an outlier time series filter (OTSF) is proposed, which mitigates the negative impact of outlier data points in historical observations by constructing an adaptive bias to learn the distribution of traffic data; during model training; (2) group Attention T-GCN (GA-TGCN) network is used to accurately capture the spatiotemporal correlations in traffic information. GA-TGCN utilizes a group attention GCN network to model the spatial factors based on road section connections, and a GRU structure to capture the temporal correlation in traffic data. We evaluate the prediction performance of ATGAN with two real-world datasets: the Los-loop dataset and the Seattle-loop dataset and compare the performance of ATGAN with recent baseline methods. The experimental results highlight the superiority of ATGAN over the baseline methods, showing that ATGAN consistently achieves the best prediction performance across various evaluation metrics.

In future research, we plan to extend the concept of the OTSF technique to reconstruct historical data, allowing the spatiotemporal correlations of traffic data to adapt to dynamic changes in traffic conditions which aims to further enhance the prediction performance. Moreover, multi-source data-based traffic prediction has garnered considerable interest recently. Therefore, the integration of multi-source traffic information into the group attention map in ATGAN, leveraging the multi-scale graph attention mechanism to comprehensively extract spatiotemporal features, represents a promising avenue for research. Meanwhile, developing a parallel model training method is also an effective approach to adapt our model to tailor our model for multi-source traffic prediction tasks.

References

[1] C. Faloutsos, J. Gasthaus, T. Januschowski, and Y. Wang, "Forecasting big time series: Old and new," *Proc. VLDB Endowment*, vol. 11, no. 12, pp. 2102–2105, 2018.

[2] S. Lee, Y. I. Lee, and B. Cho, "Short-term travel speed prediction models in car navigation systems," *J. Adv. Transp.*, vol. 40, no. 2, pp. 122–139, 2006.

[3] M. Andres and R. Nair, "A predictive-control framework to address bus bunching," *Transp. Res. Part B. Methodological*, vol. 104, pp. 123–148, 2017.

[4] E. Stellwagen and L. Tashman, "ARIMA: The models of box and Jenkins," *Foresight, Int. J. Appl. Forecasting*, vol. 30, pp. 28–33, 2013.

[5] P. Duan, G. Mao, C. Zhang, and J. Kang, "A trade-off between accuracy and complexity: Short-term traffic flow prediction with spatio-temporal correlations," in *Proc. 21st Int. Conf. Intell. Transp. Syst.*, 2018, pp. 1658–1663.

[6] M. Voort, M. Dougherty, and S. Watson, "Combining Kohonen maps with ARIMA time series models to forecast traffic flow," *Transp. Res. Part C. Emerg. Technol.*, vol. 4, no. 5, pp. 307–318, 1996.

[7] B. M. Williams and L. A. Hoel, "Modeling and forecasting vehicular traffic flow as a seasonal ARIMA process: Theoretical basis and empirical results," *J. Transp. Eng.*, vol. 129, no. 6, pp. 664–672, 2003.

[8] N. Barimani, A. R. Kian, and B. Moshiri, "Real time adaptive non-linear estimator/predictor design for traffic systems with inadequate detectors," *IET Intell. Transport Syst.*, vol. 8, no. 3, pp. 308–321, 2014.

[9] X. Song, H. Kanasugi, and R. Shibasaki, "DeepTransport: Prediction and simulation of human mobility and transportation mode at a citywide level," in *Proc. 25th Int. Joint Conf. Artif. Intell.*, 2016, pp. 2618–2624.

[10] J. Chung, C. Gulcehre, K. H. Cho, and Y. Bengio, "Empirical evaluation of gated recurrent neural networks on sequence modeling," 2014, *arXiv:1412.3555*.

[11] Z. Cui, K. Henrickson, R. Ke, Z. Pu, and Y. Wang, "Traffic graph convolutional recurrent neural network: A deep learning framework for network-scale traffic learning and forecasting," *IEEE Trans. Intell. Transp. Syst.*, vol. 21, no. 11, pp. 4883–4894, Nov. 2020.

[12] L. Zhao et al., "T-GCN: A temporal graph convolutional network for traffic prediction," *IEEE Trans. Intell. Transp. Syst.*, vol. 21, no. 9, pp. 3848–3858, Sep. 2020.

[13] J. Bai et al., "A3T-GCN: Attention temporal graph convolutional network for traffic forecasting," *ISPRS Int. J. Geo- Inf.*, vol. 10, no. 7, 2021, Art. no. 485. [Online]. Available: https://www.mdpi.com/2220-9964/10/7/485

[14] S. Saha, A. Haque, and G. Sidebottom, "Analyzing the impact of outlier data points on multi-step internet traffic prediction using deep sequence models," *IEEE Trans. Netw. Serv. Manage.*, vol. 20, no. 2, pp. 1345–1362, Jun. 2023.

[15] C. Ma, G. Dai, and J. Zhou, "Short-term traffic flow prediction for urban road sections based on time series analysis and LSTM_BILSTM method," *IEEE Trans. Intell. Transp. Syst.*, vol. 23, no. 6, pp. 5615–5624, Jun. 2022.

[16] C. Zheng, X. Fan, C. Wang, and J. Qi, "GMAN: A graph multi-attention network for traffic prediction," in *Proc. AAAI Conf. Artif. Intell.*, 2020, pp. 1234–1241.

[17] T. Alghamdi, K. Elgazzar, M. Bayoumi, T. Sharaf, and S. Shah, "Forecasting traffic congestion using ARIMA modeling," in *Proc. IEEE 15th Int. Wireless Commun. Mobile Comput. Conf.*, 2019, pp. 1227–1232.

[18] L. Run, L. X. Min, and Z. X. Lu, "Research and comparison of ARIMA and grey prediction models for subway traffic forecasting," in *Proc. Int. Conf. Intell. Comput., Automat. Syst.*, 2020, pp. 63–67.

[19] H. Bohan and B. Yun, "Traffic flow prediction based on BRNN," in *Proc. IEEE 9th Int. Conf. Electron. Inf. Emerg. Commun.*, 2019, pp. 320–323.

[20] A. K. Azad and M. S. Islam, "Traffic flow prediction model using Google map and LSTM deep learning," in *Proc. IEEE Int. Conf. Telecommun. Photon.*, 2021, pp. 1–5.

[21] W. Shu, K. Cai, and N. N. Xiong, "A short-term traffic flow prediction model based on an improved gate recurrent unit neural network," *IEEE Trans. Intell. Transp. Syst.*, vol. 23, no. 9, pp. 16654–16665, Sep. 2022.

[22] D. Liang, J. Zhang, S. Jiang, X. Zhang, J. Wu, and Q. Sun, "Mobile traffic prediction based on densely connected CNN for cellular networks in highway scenarios," in *Proc. 11th Int. Conf. Wireless Commun. Signal Process.*, 2019, pp. 1–5.

[23] S. Buapang and V. Muangsin, "Traffic prediction with a spectral graph neural network," in *Proc. IEEE 7th Int. Conf. Bus. Ind. Res.*, 2022, pp. 341–346.

[24] J. Zhu et al., "KST-GCN: A knowledge-driven spatial-temporal graph convolutional network for traffic forecasting," *IEEE Trans. Intell. Transp. Syst.*, vol. 23, no. 9, pp. 15055–15065, Sep. 2022.

[25] C. Chen, Z. Liu, S. Wan, J. Luan, and Q. Pei, "Traffic flow prediction based on deep learning in Internet of Vehicles," *IEEE Trans. Intell. Transp. Syst.*, vol. 22, no. 6, pp. 3776–3789, Jun. 2021.

[26] J. M. Yang, Z. R. Peng, and L. Lin, "Real-time spatiotemporal prediction and imputation of traffic status based on LSTM and graph Laplacian regularized matrix factorization," *Transp. Res., Part C. Emerg. Technol.*, vol. 129, 2021, Art. no. 103228.

[27] L. Lin, Z. He, and S. Peeta, "Predicting station-level hourly demands in a large-scale bike-sharing network: A graph convolutional neural network approach," *Transp. Res., Part C. Emerg. Technol.*, vol. 97, pp. 258–276, 2017.

[28] P. Xie et al., "Spatio-temporal dynamic graph relation learning for urban metro flow prediction," *IEEE Trans. Knowl. Data Eng.*, vol. 35, no. 10, pp. 9973–9984, Oct. 2023.

[29] Q. Zhang, C. Li, F. Su, and Y. Li, "Spatiotemporal residual graph attention network for traffic flow forecasting," *IEEE Internet Things J.*, vol. 10, no. 13, pp. 11518–11532, Jul. 2023.

[30] A. Abdelraouf, M. A. Abdel-Aty, and J. Yuan, "Utilizing attention-based multi-encoder-decoder neural networks for freeway traffic speed prediction," *IEEE Trans. Intell. Transp. Syst.*, vol. 23, no. 8, pp. 11960–11969, Aug. 2022, doi: 10.1109/TITS.2021.3108939.

[31] T. Li, Z. Zhao, C. Sun, R. Yan, and X. Chen, "Hierarchical attention graph convolutional network to fuse multi-sensor signals for remaining useful life prediction," *Rel. Eng. Syst. Saf.*, vol. 215, 2021, Art. no. 107878.

[32] Z. Wang et al., "Spatio-temporal self-attention network for video saliency prediction," *IEEE Trans. Multimedia*, vol. 25, pp. 1161–1174, 2023.

[33] Y. Duan, N. Chen, S. Shen, P. Zhang, Y. Qu, and S. Yu, "FDSA-STG: Fully dynamic self-attention spatio-temporal graph networks for intelligent traffic flow prediction," *IEEE Trans. Veh. Technol.*, vol. 71, no. 9, pp. 9250–9260, Sep. 2022.

[34] S. Wei, Q. Qu, X. Zeng, J. Liang, J. Shi, and X. Zhang, "Self-attention Bi-LSTM networks for radar signal modulation recognition," *IEEE Trans. Microw. Theory Techn.*, vol. 69, no. 11, pp. 5160–5172, Nov. 2021.

[35] C. Zhang, S. Dang, B. Shihada, and M.-S. Alouini, "Dual attention-based federated learning for wireless traffic prediction," in *Proc. IEEE Conf. Comput. Commun.*, 2021, pp. 1–10.

[36] H. Li, S. Zhang, L. Su, H. Huang, and X. Li, "GraphSANet: A graph neural network and self attention based approach for spatial temporal prediction in sensor network," in *Proc. IEEE Int. Conf. Big Data*, 2020, pp. 5756–5758.

[37] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2016, pp. 770–778.

[38] Y. Qu, Z. Li, J. Zhao, and H. Li, "Unbalanced data processing for software defect prediction," in *Proc. IEEE 4th Int. Conf. Data-Driven Optim. Complex Syst.*, 2022, pp. 1–6.

[39] S. Makridakis and M. Hibon, "ARMA models and the box-Jenkins methodology," *J. Forecasting*, vol. 16, no. 3, pp. 147–163, 1997.

[40] M. Taylor, K. Dhuyvetter, and T. Kastens, "Incorporating current information into historical-average-based forecasts to improve crop price basis forecasts," in *Proc. NCR-134 Conf. Appl. Commodity Price Anal., Forecasting, Market Risk Manage.*, Apr. 2004.

[41] C. H. Wu, C. C. Wei, D. C. Su, M. H. Chang, and J. M. Ho, "Travel time prediction with support vector regression," in *Proc. Int. Conf. Intell. Transp. Syst.*, 2003, pp. 1438–1442.

[42] S. Hochreiter and J. Schmidhuber, "LSTM can solve hard long term lag problems," in *Proc. Int. Conf. Adv. Neural Inf. Process. Syst.*, 1997, vol. 9, pp. 473–479.

[43] L. Zhao et al., "T-GCN: A temporal graph convolutional network for traffic prediction," *IEEE Trans. Intell. Transp. Syst.*, vol. 21, no. 9, pp. 3848–3858, Sep. 2020.

[44] Y. Bao et al., "Spatial–temporal complex graph convolution network for traffic flow prediction," *Eng. Appl. Artif. Intell.*, vol. 121, 2023, Art. no. 106044. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0952197623002282

Ruiyuan Jiang received the M.Sc. degree in 2020, from Xi'an Jiaotong-Liverpool University, Suzhou, China, where he is currently working toward the Ph.D. degree. His research interests include macroscopic traffic prediction, trajectory prediction and autonomous vehicle techniques.

Shangbo Wang received the Dr.-Ing degree in wireless communication from the University of Duisburg-Essen, Duisburg, Germany, in 2014, and the Ph.D. degree in intelligent transportation systems from the University of Technology Sydney, Ultimo, NSW, Australia, in 2019. He was a Visiting Scholar with the The University of Hong Kong, Hong Kong, Assistant Professor with Xi'an Jiaotong Liverpool University, Suzhou, China, Postdoctoral Research Fellow with The University of Sydney, Camperdown, NSW, Radar Engineer with Continental AG, Lindau, Germany, and Development Engineer with Siemens AG, Munich, Germany. He joined the University of Sussex, Brighton and Hove, U.K., in 2024, as a Lecturer (Assistant Professor). He has authored or coauthored more than 35 papers in leading international journal and conference papers, including IEEE TRANSACTIONS ON INTELLIGENT TRANSPORTATION SYSTEMS and IEEE TRANSACTIONS ON SIGNAL PROCESSING. His research interests mainly include intelligent transportation systems, wireless communications, radar signal processing, wireless localization techniques, machine learning, deep reinforcement learning and its applications in transportation systems. He participated in a NSFC (National Natural Science Foundation of China) general program as a core member. He was the recipient of multiple academic awards, including Jiangsu Yong Talent Program, Hubei Yong Talent Program, Texas Instruments Excellence in Signal Processing Award, and Australian Research Council full scholarship.

Dongyao Jia (Member, IEEE) received the B.E. degree in automation from Harbin Engineering University, Harbin, China, in 1998, and the Ph.D. degree in computer science from The City University of Hong Kong, Hong Kong, in 2014. He was a Research Fellow with the Institute for Transport Studies, University of Leeds, Leeds, U.K., from 2015 to 2018, and The University of Queensland, Brisbane, QLD, Australia, from 2018 to 2021. He is currently an Associate Professor with the School of Advanced Technology, Xi'an Jiaotong-Liverpool University, Suzhou, China. He also holds ten years of working experience in the telecom industry in China. His research interests include connected and automated driving, intelligent transport systems, digital twins, and the Internet of Things.

Guoqiang Mao (Fellow, IEEE) is currently a Leading Professor, Founding Director of the Research Institute of Smart Transportation, and Vice-Director of the ISN State Key Lab, Xidian University, Xi'an, China. He was with the University of Technology Sydney, Ultimo, NSW, Australia, and The University of Sydney, Camperdown, NSW. He has authored or coauthored 300 papers in international conferences and journals that have been cited more than 14 000 times. His research interest includes intelligent transport systems, Internet of Things, wireless localization techniques, wireless sensor networks, and applied graph theory and its applications in telecommunications. His H-index is 54 and is in the list of Top 2% most-cited Scientists Worldwide 2022 by Stanford University in 2022 and 2023. He is a Vice-Director of Smart Transportation Information Engineering Society, Chinese Institute of Electronics (2022-), and was a Co-Chair IEEE ITS Technical Committee on Communication Networks (2014–2017). He is an Editor of IEEE TRANSACTIONS ON INTELLIGENT TRANSPORTATION SYSTEMS (since 2018), IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS (2014–2019), IEEE TRANSACTIONS ON VEHICULAR TECHNOLOGY (2010–2020), and received "Top Editor" award for outstanding contributions to IEEE TRANSACTIONS ON VEHICULAR TECHNOLOGY in 2011, 2014 and 2015. He was a chair, co-chair and TPC member in a number of international conferences. He is a Fellow of AAIA and IET.

Eng Gee Lim (Senior Member, IEEE) received the B.Eng. (Hons.) and Ph.D. degrees in electrical and electronic engineering from Northumbria University, Newcastle, U.K., in 1998 and 2002, respectively. He was with Andrew Ltd., a leading communications systems company in U.K., from 2002 to 2007. Since August 2007, he has been with Xian Jiaotong-Liverpool University, Suzhou, China, where he was formally the Head of EEE Department and University Dean of Research and Graduate studies. He is currently School Dean of Advanced Technology, Director of AI University Research Centre and also Professor with the Department of Communications and Networking. He has authored or coauthored more than 200 refereed international journal and conference papers. His research interests include artificial intelligence, robotics, AI+ health care, future education, management in higher education, international standard (ISO/ IEC) in robotics, antennas, RF/microwave engineering, EM measurements/simulations, energy harvesting, power/energy transfer, smart-grid communication, wireless communication networks for smart and green cities. He is a Charter Engineer, and Fellow of IET and Engineers Australia. He is also a Senior Fellow of HEA.