

# FDSA-STG: Fully Dynamic Self-Attention Spatio-Temporal Graph Networks for Intelligent Traffic Flow Prediction

Youxiang Duan<sup>1</sup>, Ning Chen<sup>1</sup>, Graduate Student Member, IEEE, Shigen Shen<sup>1</sup>, Member, IEEE, Peiying Zhang<sup>1</sup>, Member, IEEE, Youyang Qu, Member, IEEE, and Shui Yu<sup>2</sup>, Senior Member, IEEE

**Abstract**—With the development of transportation and the ever-improving of vehicular technology, Artificial Intelligence (AI) has been popularized in Intelligent Transportation Systems (ITS), especially in Traffic Flow Prediction (TFP). TFP plays an increasingly important role in alleviating traffic pressure caused by regional emergencies and coordinating resource allocation in advance to deployment decisions. However, existing research can hardly model the original intricate structural relationships of the transportation network (TN) due to the lack of in-depth consideration of the dynamic relevance of spatial, temporal, and periodic characteristics. Motivated by this and combined with deep learning (DL), we propose a novel framework entitled *Fully Dynamic Self-Attention Spatio-Temporal Graph Networks* (FDSA-STG) by improving the attention mechanism using Graph Attention Networks (GATs). In particular, to dynamically integrate the correlations of spatial dimension, time dimension, and periodic characteristics for highly-accurate prediction, we correspondingly devised three components including the spatial graph attention component (SGAT), the temporal graph attention component (TGAT), and the fusion layer. In this case, three groups of similar structures are designed to extract the flow characteristics of recent periodicity, daily periodicity, and weekly periodicity. Extensive evaluation results show the superiority of FDSA-STG from perspectives of prediction accuracy and prediction stability improvements, which also testifies high model adaptability to the dynamic characteristics of the actual observed traffic flow (TF).

Manuscript received 31 December 2021; revised 24 March 2022; accepted 17 May 2022. Date of publication 26 May 2022; date of current version 19 September 2022. This work was supported in part by the Shandong Provincial Natural Science Foundation, China under Grant ZR2020MF006, in part by the Industry-University Research Innovation Foundation of Ministry of Education of China under Grant 2021FNA01001, in part by the Major Scientific and Technological Projects of CNPC under Grant ZD2019-183-006, in part by the Open Foundation of State Key Laboratory of Integrated Services Networks (Xidian University) under Grant ISN23-09, and in part by the Zhejiang Provincial Natural Science Foundation of China under Grant LZ22F020002. The review of this article was coordinated by Prof. Yu Cheng. (*Corresponding authors:* Shigen Shen; Peiying Zhang.)

Youxiang Duan, Ning Chen, and Peiying Zhang are with the College of Computer Science and Technology, China University of Petroleum (East China), Qingdao 266580, China, and also with the State Key Laboratory of Integrated Services Networks, Xidian University, Xi'an 710071, China (e-mail: yxduan@upc.edu.cn; nchen@s.upc.edu.cn; zhangpeiying@upc.edu.cn).

Shigen Shen is with the Department of Computer Science and Engineering, Shaoxing University, Shaoxing 312000, China (e-mail: shigens@usx.edu.cn).

Youyang Qu is with the Data61, Australia Commonwealth Scientific and Industrial Research Organization, Canberra, ACT 2601, Australia (e-mail: youyang.qu@data61.csiro.au).

Shui Yu is with the School of Computer Science, University of Technology Sydney, Ultimo, NSW 2007, Australia (e-mail: shui.yu@uts.edu.au).

Digital Object Identifier 10.1109/TVT.2022.3178094

**Index Terms**—Deep learning (DL), graph attention networks (GATs), intelligent transportation systems (ITS), self-attention, traffic flow prediction (TFP).

## I. INTRODUCTION

IN RECENT years, the vigorous development of the transportation industry has generated massive amounts of Traffic Big Data (TBD) [1], [2]. As shown in Fig. 1, there are intricate relationships among the TBD of vehicles considering intersections, environment, vehicle speed, etc. A considerable number of recent studies analyze and extract hidden patterns and knowledge of the generated data to mine valuable information [3], which is exploited to provide scientific and reasonable support for urban transportation-related decision-making, so as to further improve the level of management. To enhance the performances, technologies based on Artificial Intelligence (AI) [4] have been developed and widely applied to ITS [5].

Intuitively, accurate and real-time road section traffic information is the key to the successful application of ITS-related technologies [6], [7]. In particular, vehicle congestion problems frequently occur in the traffic systems, especially during peak periods and holidays [8]. This significantly affects the normal operation of the transportation network (TN) by increasing the time and travel costs and potentially raising the probability of traffic accidents [9]. To alleviate the above-mentioned issues, Traffic Flow Prediction (TFP) has been advanced to the forefront of academic research [10], for purposes of making adjustments and decisions in advance to relieve traffic pressure, and coordinating resource allocation in response to regional emergencies [11].

Traffic Flow (TF) data generated from TNs can be regarded as a kind of spatio-temporal graph data with time and spatial dimensions [12], [13]. Different types of traffic data (vehicle speed, number of lanes, etc.) is embedded in the spatial dimension and changes dynamically over time [14]. Consequently, the crucial target of TFP is to capture temporal and spatial correlation [15] considering periodic characteristics [16], [17]. Therefore, the TF at the current time is not only affected by neighboring nodes, but also by the periodic characteristics of the previous time slots. To sum up, the purpose of TFP is to predict the TFs in the coming time interval based on the characteristics of temporal and spatial correlation and periodicity in the historical TFs [18].

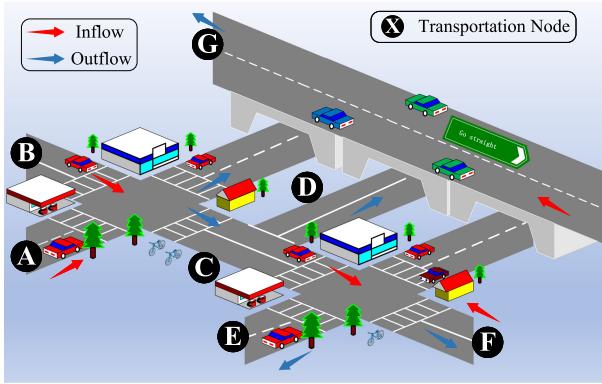


Fig. 1. Transportation network diagram with in-and-out traffic flow.

Most existing works, such as [19], [20], are based on the statistical analysis of the TFs in the time series that has occurred, ignoring the complex spatial dependence between vehicles in TN, which makes it difficult to achieve better performance. To capture the complex nonlinear characteristics in the traffic data, Deep Neural Networks (DNNs) [21] have been employed to extract these features [22]. Among them, Convolutional Neural Networks (CNNs) [23] have a larger receptive field due to operations such as pooling, which in turn allows the neighborhood relationship and local spatial characteristics to be preserved, so it is widely utilized to extract the spatial dependence of TFs [9]. Graph Neural Networks (GCNs) build graph features based on the adjacency relationship of vehicles in TN, and then models the spatial dimensional topological relationship [24]. Recurrent Neural Networks (RNNs) can model long-term or short-term time series with sequential relationships and are suitable for modeling the time series of TFs [25]. Several other variants, such as the Residual Network (ResNet) [26], Long Short-Term Memory networks (LSTM) [15], [27], Transformer [28], etc., are also used for TFP modeling and analysis. In addition, there are also hybrid methods like [29] that combines a variety of different DNNs to separately model the spatial and temporal data.

However, the accuracy of TFP is not satisfactorily high due to the lack of in-depth consideration of the specific spatial correlations of different observation points. Specifically, some methods only hold a naive assumption that the traffic nodes are equally important. Nevertheless, these nodes that affect each other have **dynamic spatial correlations**, *i.e.*, the correlations among different nodes at the same time slice vary a lot. Since the self-attention mechanism [30] can differentially focus on the importance of different data points, it has been widely used in classification, prediction, question answering, and other tasks. Moreover, Graph Attention Networks (GATs) introduce attention to graph topology data, adaptively assign importance weights to adjacent points, and differentially aggregate the feature information of adjacent points [31]. Therefore, GAT significantly improves the expressiveness of graph neural network models. Based on the above motivation and inspired by GATs, we propose to employ the GATs to capture the complicated and dynamic relationship between different observation points in the spatial dimension. In addition, the correlation between two

groups of the same traffic node at different times in the TN is also different, *i.e.*, the **dynamic time correlations**. Therefore, we also dynamically capture the attention weights between different time slices and fuse the associated information of adjacent regional times through the GATs. Furthermore, TFs also have **dynamic periodic correlations**, which should also be captured and fused in a dynamical manner.

Based on the above analysis, there are intricate structural relationships [32] among adjacent vehicles in a TN, which makes TFs have dynamic correlations in space, time, and periodicity. To model this dynamic association, we propose a novel model entitled *Fully Dynamic Self-Attention Spatio-Temporal Graph Networks* (FDSA-STG), which is inspired by the attention mechanism [30] while being a multi-layer deep learning framework. The main contributions of this paper are as follows.

- We propose a novel model for accurate TFP problem entitled FDSA-STG, which jointly modifies the GATs and the self-attention mechanism that fully dynamically focuses and integrates spatial, temporal, and periodic correlations. In addition, it can differentially capture the global and local dependencies of diverse characteristics in TN. Specifically, it carries out real-time differentiated attention to distinguishing regions and time slots, thereby fitting the original intricate structural relationship of TFs, so that the prediction performance is significantly improved.
- A tailor-made neural network structure is designed, which includes the spatial graph attention component (SGAT) for extracting spatial characteristics, the temporal graph attention component (TGAT) for extracting regional temporal characteristics, and the fusion layer for fusing three groups of periodic structural characteristics. These components are specifically designed by combining different characteristics of TFs, and completely dynamic self-attention mechanisms are correspondingly designed to focus on differences. This fine-grained and fully dynamic self-attention mechanism enables FDSA-STG to have superior prediction stability and present a more balanced evaluation curve avoiding error accumulation with time.
- Extensive experimental results on real-world datasets testify that FDSA-STG exceeds other baseline models while showing improved performances on popular evaluation matrices at the same time.

The remainder of this paper is organized as follows. Related work is reviewed in Section II. In Section III, the definition of the problem is formulated for clarity. The proposed FDSA-STG consists of four parts, which are articulated with details in Section IV. In Section V, we explain the experimental process and comparison results in all aspects to illustrate the effectiveness of FDSA-STG, which is followed by summary and future work in Section VI.

## II. RELATED WORK

The solutions for the TFP problem roughly evolve from traditional statistical analysis to machine learning-based methods. In this section, we correspondingly introduce these two main categories in detail.

### A. Traditional Statistical Analysis

The most representative methods include the Historical Average (HA), AutoRegressive Integrated Moving Average (ARIMA) [19], and Vector AutoRegressive (VAR) [20]. HA conducts prediction using the mean value of TFs over some time slots that has occurred. ARIMA simulates the TFs of a time series through mathematical modeling. VAR conducts prediction by linearly fitting the characteristics of TFs. Despite this, none of them take the complex nonlinear characteristics of TFs into account. Therefore, satisfactory prediction results cannot be obtained in various cases.

### B. From Machine Learning to Deep Learning

Machine learning and deep learning technologies [33] have shown excellent results in fitting the intricate structural characteristics of data and have received widespread attention [34]–[36]. Since CNNs can retain the local characteristics of neighborhood connections and space, and has strong abstract representation ability for local operations, it has been widely employed to extract the spatial features of TFs [9]. Besides, RNNs can better model the problems of time series, while LSTM [37] is widely used to extract the time features of TFs [27]. The work [25] captures the spatio-temporal correlation through a two-dimensional network composed of a large amount of memory units. There are also some other deep learning methods, such as the work [22] based on stacked autoencoders to learn traffic characteristics in a greedy hierarchical manner. In order to better fit the spatio-temporal characteristics, there are also some hybrid methods, such as DNN-BTF [38] using CNNs to extract spatial characteristics and RNNs to extract temporal characteristics. The work [29] extracts the characteristics of TFs by combining CNNs and LSTM, and combines the attention mechanism [30] to focus on the importance of different periods. Multi-level Attention Network (GeoMAN) [39] combines attention and recurrent neural network to perform global and local attention while extracting different characteristics of TF well. In addition, the GCNs [40] is widely exploited because of its excellent ability to process data with graph-structured characteristics, and can commendably fit the temporal and spatial characteristics of TPs. For example, Attention-Based Spatial-Temporal Graph Convolutional Network (ASTGCN) [24] combined with GCNs and attention effectively captures the dynamic spatio-temporal correlation.

Based on the above literature review, we found that insufficient attention has been paid to the different correlation characteristics of TFs, including spatial, temporal, and periodic characteristics. Motivate by this, we innovatively integrate the fine-grained self-attention mechanism into TFP, with which the model achieves a fully-dynamic differentiated focus on different regions and different moments. The prediction performance is subsequently promoted as analyzed and illustrated in following sections.

### III. PROBLEM DEFINITION

To better clarify, we present the problem definition from the perspectives of data definition and flow prediction definition.

TABLE I  
THE NOTATIONS USED IN THIS WORK

Notation	Definition
$V$	Nodes of TN
$E$	Edges of TN
$t_d_i^k$	TFs of node $i$ at time slice $k$
$y^k$	TFs of $V$ at time slice $k$
$Y^{k:t}$	TFs of $V$ at time interval $t$ before time slice $k$
$\Delta$	Sampling frequency
$q$	Number of Sampling
$k_p$	Current time
$k_p^d$	Same time as $k_p$ on the previous day
$k_p^w$	Same time as $k_p$ on the same day of the previous week
$Y_{resent}^{k_p:2q}$	TFs of recent periodicity
$Y_{daily}^{k_p:2q}$	TFs of daily periodicity
$Y_{weekly}^{k_p:2q}$	TFs of weekly periodicity

The flow prediction problem is divided into three components, which are recent periodicity, daily periodicity, and weekly periodicity. These three components are able to capture all the TF patterns while considering sudden TF to a certain degree.

#### A. Data Definition

In this work, the relevant notations used are explained in Table I. Taking the spatial traffic road network as a reference, the relevant data applied in this research is described as follows:  $V$  is the set of nodes in the network, which comes from the traffic location at the time of data collection, such as intersections, high-altitude detection, and so on. We assume that there are a total of  $n$  nodes in  $V$ . Because of the complex relationship between nodes, we exploit graph features to manipulate traffic data.  $E \in \mathcal{R}^{n \times n}$  is the adjacency relationship between nodes. If  $E_{ij} = 1$ , it means that there exists an association relationship between nodes  $i$  and  $j$ . In other words, the traffic flow between points  $i$  and  $j$  will affect each other. In addition, there are  $m$  types of traffic data in each node, which together play a certain role in the real-time traffic of the nodes.

It should be noted that TF is a kind of spatio-temporal graph data. As shown in Fig. 2, the vertical dimension is the time dimension, and inside it, each time slice is the spatial dimension. Let  $t_d_{ij}^k$  denotes the TF of the  $j$ -th type of the  $i$ -th node at the current time  $k$ . Therefore, let  $t_d_i^k \in \mathcal{R}^{1 \times m}$  denote the TFs of node  $i$  at time  $k$ , the TFs of  $V$  at the time  $k$  can be defined as

$$y^k = \{t_d_i^k\} \in \mathcal{R}^{n \times m}. \quad (1)$$

As a result, for  $V$ , the TFs in time interval  $t$  before time  $k$  can be defined as a matrix denoted by (2).

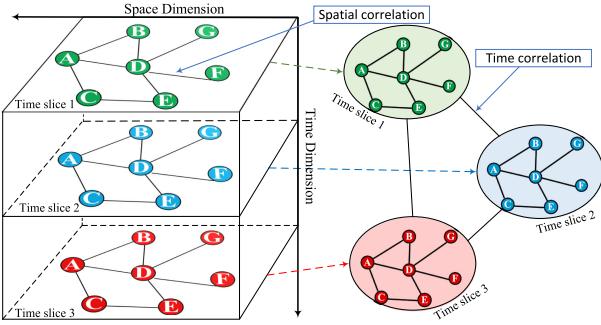


Fig. 2. Spatio-temporal graph data. The abscissa represents the space dimension, and the ordinate represents the time dimension. Each layer represents a time slice, i.e., a certain moment. The interior of each layer represents the spatial topological relationship of the TN at the current moment, i.e., the spatial correlation. Meanwhile, each time slice is a spatial map. There is also a correlation between different time slices, and there is also a time topology relationship between time slices, i.e., the time correlation. The combination of the two forms the topological structure of the Spatio-temporal graph representing TF.

$$\begin{aligned} \mathbf{Y}^{k:t} &= [\mathbf{y}^{k-t}, \dots, \mathbf{y}^{k-1}, \mathbf{y}^k]^T \in \mathcal{R}^{t \times n \times m} \\ &= \begin{bmatrix} t_d_1^{k-t} & t_d_2^{k-t} & \dots & t_d_n^{k-t} \\ \vdots & \vdots & \ddots & \vdots \\ t_d_1^{k-1} & t_d_2^{k-1} & \dots & t_d_n^{k-1} \\ t_d_1^k & t_d_2^k & \dots & t_d_n^k \end{bmatrix} \quad (2) \end{aligned}$$

### B. Flow Prediction

In this work, to explore the periodic characteristics, assuming that the sampling frequency is  $\Delta$  and the number of sampling is  $q$ , given the  $\mathbf{Y}^{k_g:q\Delta}$  of the TN at any time  $k_g$ , where the past sampling times are  $(k_g - q\Delta, \dots, k_g - \Delta, k_g)$ , the goal is to predict the  $\mathbf{Y}^{k_p:r\Delta}$  of the subsequent interval  $t_p$  at the current time  $k_p$ , i.e., the TFs of moments  $(k_p, k_p + \Delta, \dots, k_p + r\Delta)$ , where  $r$  is the number of sampling in prediction. In this paper, we set  $q = 12$  and  $\Delta$  is set as 5 minutes. It is worth noting that, to facilitate the clarity, we omit  $\Delta$ , such that  $k - t\Delta$  is denoted as  $k - t$ .

Correspondingly, we select three types of representative TFs with recent periodicity, daily periodicity, and weekly periodicity from the time dimension, which are respectively expressed as follows.

1) *Recent Periodicity*: As we all know, the variation trend of the TF value at the current moment is often closely related to the TF value at the previous time interval. Inspired by it, the TFs with recent periodicity are extracted from the previous  $2q$  time interval of the current time  $k_p$ , which is symbolized as

$$\mathbf{Y}_{resent}^{k_p:2q} = \begin{bmatrix} t_d_1^{k_p} & t_d_2^{k_p} & \dots & t_d_n^{k_p} \\ \vdots & \vdots & \ddots & \vdots \\ t_d_1^{k_p-1} & t_d_2^{k_p-1} & \dots & t_d_n^{k_p-1} \\ t_d_1^{k_p-q} & t_d_2^{k_p-q} & \dots & t_d_n^{k_p-q} \\ \vdots & \vdots & \ddots & \vdots \\ t_d_1^{k_p-2q} & t_d_2^{k_p-2q} & \dots & t_d_n^{k_p-2q} \end{bmatrix}. \quad (3)$$

TABLE II  
PERIODIC TFs VALUES, TAKING 12:00 ON TUESDAY, 23TH NOVEMBER, 2021 AS AN EXAMPLE

Periodic TFs	Time Periods
Recent	10:00 - 12:00 on Tuesday, 23th November, 2021
Daily	11:00 - 13:00 on Monday, 23th November, 2021
Weekly	11:00 - 13:00 on Tuesday, 16th November, 2021

It should be noted that the subsequent fusion process of the time dimension will be described in Section IV-B.

2) *Daily Periodicity*: Inspired by the frequent peak periods in the morning and evening caused by people's similar work patterns every day, TFs at the same time on different days often have the same characteristics. Therefore, the TFs with daily periodicity are extracted from each  $q$  time interval before and after the same time  $k_p^d$  on the previous day, and it is symbolized as

$$\mathbf{Y}_{daily}^{k_p^d:2q} = \begin{bmatrix} t_d_1^{k_p^d+q} & t_d_2^{k_p^d+q} & \dots & t_d_n^{k_p^d+q} \\ \vdots & \vdots & \ddots & \vdots \\ t_d_1^{k_p^d} & t_d_2^{k_p^d} & \dots & t_d_n^{k_p^d} \\ \vdots & \vdots & \ddots & \vdots \\ t_d_1^{k_p^d-(q-1)} & t_d_2^{k_p^d-(q-1)} & \dots & t_d_n^{k_p^d-(q-1)} \\ t_d_1^{k_p^d-q} & t_d_2^{k_p^d-q} & \dots & t_d_n^{k_p^d-q} \end{bmatrix}. \quad (4)$$

3) *Weekly Periodicity*: Similarly, TFs on the same day in different weeks often have similar characteristics. Therefore, the TFs with weekly periodicity are extracted from the  $q$  time intervals before and after the same time  $k_p^w$  on the same day of the previous week, which is symbolized as

$$\mathbf{Y}_{weekly}^{k_p^w:2q} = \begin{bmatrix} t_d_1^{k_p^w+q} & t_d_2^{k_p^w+q} & \dots & t_d_n^{k_p^w+q} \\ \vdots & \vdots & \ddots & \vdots \\ t_d_1^{k_p^w} & t_d_2^{k_p^w} & \dots & t_d_n^{k_p^w} \\ \vdots & \vdots & \ddots & \vdots \\ t_d_1^{k_p^w-(q-1)} & t_d_2^{k_p^w-(q-1)} & \dots & t_d_n^{k_p^w-(q-1)} \\ t_d_1^{k_p^w-q} & t_d_2^{k_p^w-q} & \dots & t_d_n^{k_p^w-q} \end{bmatrix}. \quad (5)$$

Taking the prediction on the TF at 12:00 on Tuesday, 23th November, 2021, and subsequent time intervals as an example, the time periods to which the three types of representative periodic TFs values belong are shown in Table II.

## IV. FDSA-STG MODELING

In this section, we present the framework of FDSA-STG, which is shown in Fig. 3. The main components are as follows.

- Three groups of similar network architectures to extract spatio-temporal correlation, where the SGAT is mainly used to extract spatial features, and the TGAT is mainly used to obtain the regional periodic features.

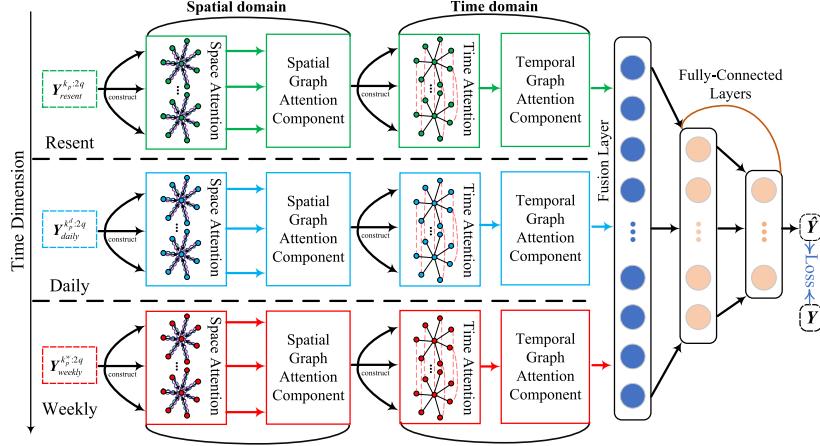


Fig. 3. The framework of FDSA-STG with a spatial graph attention component, a temporal graph attention component, and a fusion layer. The two-layer fully-connected layer will perform the traffic flow prediction with gradient back-propagation.

- The fully-connected layer of fusion features is utilized to fuse three groups of time-periodic TFs (resent, daily, weekly).
- A two-layer fully connected layer is employed for traffic prediction. In the end, the errors are compared between the predicted TF and the real TF, and then the gradients are used for back-propagation.

#### A. Spatial Graph Attention Component

Meanwhile, there are intricate structural relationships between observation points, and graph features can properly describe this relationship. Since all the SGATs in FDSA-STG have similar structures, we take the spatial dimension of recent periodicity as an example for illustration. First, we need to construct the adjacency matrix edge  $E$  according to the layout of TN. For example, Fig. 4(a) shows an example of the construction process. Therefore, the adjacency set  $\mathcal{A}_i$  of traffic node  $i$  is

$$\mathcal{A}_i = \{j | E_{ij} = 1, 1 \leq j \leq n\}. \quad (6)$$

It should be noted that inspired by the concept of “channel” in the traditional convolution operation, in this component, we regard each sampling time slice ( $\Delta$ ) as a “time channel”. That is, the time dimension corresponds to the channel dimension. In addition, the correlation between two groups of the same traffic nodes at different times in the TN is different. For example, in Fig. 2, the relevance of traffic nodes “A” and “B” at different times slice is different. Therefore, based on the above inspirations, we design “graph attention layers” for each time channel, which is utilized to extract the spatial characteristics of each time slice. The specific SGAT design is shown in Fig. 5, which is composed of  $2q$  graph attention layers, which are finally connected and output in the time dimension.

Correspondingly, in the TN of each time slice  $k$ , the graph attention layers is composed of two graph attention single-layer connected in series. In each layer, the normalized weight coefficient  $atts_{ij}^k$  of the traffic node  $i$  and its adjacent nodes  $j$ , which

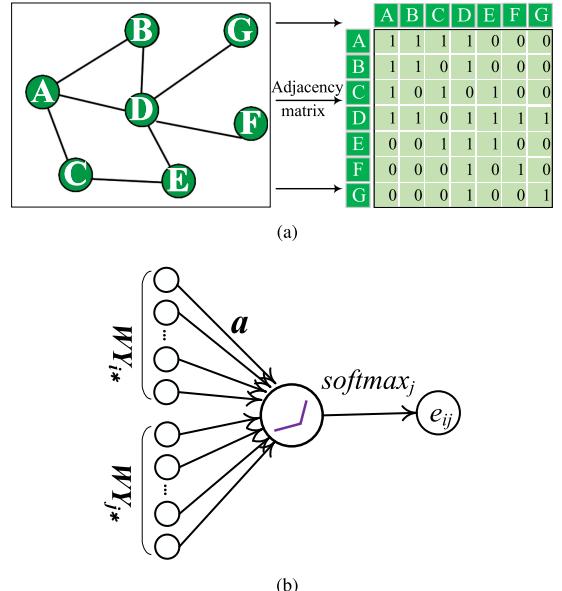


Fig. 4. Examples of parameter construction of SGAT. (a) The construction process of adjacency matrix. (b) The attention mechanism.

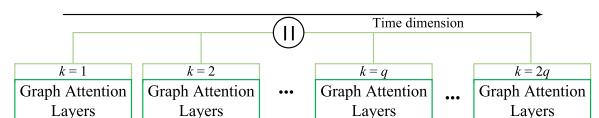


Fig. 5. Composition of SGAT.

implies the degree of importance, can be formulated as

$$atts_{ij}^k = \frac{\exp(\text{LeakyReLU}([\mathbf{Y}_{ki*}\mathbf{W}^k \| \mathbf{Y}_{kj*}\mathbf{W}^k] \mathbf{a}^T))}{\sum_{s \in \mathcal{A}_i} \exp(\text{LeakyReLU}([\mathbf{Y}_{ki*}\mathbf{W}^k \| \mathbf{Y}_{ks*}\mathbf{W}^k] \mathbf{a}^T))}, \quad (7)$$

where  $\mathbf{W}$  is the weight matrix of the layer, and its dimension is  $F \times F'$ .  $F$  and  $F'$  are the dimensions of input and output, respectively.  $\text{LeakyReLU}$  is the activation function, and its negative input slope is set as 0.2.  $\parallel$  is the operation of concatenation.  $\mathbf{a}$  is the attention mechanism shown in Fig. 4(b), which is initialized as a single-layer feed-forward layer. In summary, the forward propagation process of the graph attention layers of the time slice  $k$  is

$$\mathbf{Y}_{k*}^{(l+1)} = \sigma \left( \frac{1}{Z} \sum_{z=1}^Z \sum_{j \in \mathcal{A}_i} (\text{att}_{ij}^k)^z \mathbf{Y}_{kj*}^{(l)} (\mathbf{W}^k)^z \right), \quad (8)$$

where  $\sigma$  is the activation function,  $Z = 3$  is the number of attention heads, and the superscript  $z$  represents the  $z$ -th attention head. Therefore, the output of the SGAT can be denoted as

$$\mathbf{Y}'_{resent} = \parallel \{\mathbf{Y}_{k*}\}_{k=1}^{2q}. \quad (9)$$

Similarly, we can obtain the output of the SGAT of the daily periodicity and weekly periodicity,  $\mathbf{Y}'_{daily}$  and  $\mathbf{Y}'_{weekly}$ , respectively. It should be noted that  $\parallel$  means concatenation according to the time dimension.

### B. Temporal Graph Attention Component

After the SGAT, each traffic node in each time slice has captured the spatial characteristics of its own and adjacent points, and thereby updates its characteristics. To capture the dynamic time correlations, we propose a TGAT as shown in the middle part of Fig. 3. It dynamically captures the attention weight between different time slices (as shown in Fig. 2), and fuses the correlation information of regional times through the graph attention layers, which are composed of two graph attention single-layer connected in the time dimension. In each layer, the normalized weight coefficient  $\text{att}_{ij}$  of the time slice  $i$  and  $j$ , which means the degree of importance between each time slice in the  $2q\Delta$  interval after spatial feature fusion, can be defined as

$$\text{att}_{ij} = \frac{\exp(\text{LeakyReLU}([\mathbf{Y}'_{i*} \mathbf{W} \| \mathbf{Y}'_{j*} \mathbf{W}] \mathbf{a}^T))}{\sum_{u \in \mathcal{B}_i} \exp(\text{LeakyReLU}([\mathbf{Y}'_{i*} \mathbf{W} \| \mathbf{Y}'_{u*} \mathbf{W}] \mathbf{a}^T))}, \quad (10)$$

where  $\mathbf{Y}' \in \mathcal{R}^{t \times n \times m'}$  is the output of the SGAT, and  $m'$  is the new feature dimension size of each node after extraction, where  $t = 2q$ .  $\mathcal{B}_i$  is the set of different time slices in the time interval  $t$ , which is defined as

$$\begin{aligned} \mathcal{B}_i &= \{j | j = k_p - v\}, \\ \text{s.t. } & \begin{cases} 0 \leq v \leq 2q, & \text{Recent,} \\ -q \leq v \leq q, & \text{Daily or Weekly.} \end{cases} \end{aligned} \quad (11)$$

It should be noted that some parameters in (10) are in the same form of those in (7), but they are completely independent of each other.

At the same time, we fuse the correlation information of regional times through the TGAT, and the forward propagation process is

$$\mathbf{Y}'_{i*}^{(l+1)} = \sigma \left( \frac{1}{Z} \sum_{z=1}^Z \sum_{u \in \mathcal{B}_i} \text{att}_{ij}^z \mathbf{Y}'_{u*}^{(l)} \mathbf{W}^z \right), \quad (12)$$

where  $Z = 3$  is the number of attention heads. After the forward propagation of this component, we can obtain the TF matrix  $\mathbf{Y}'_{resent} \in \mathcal{R}^{t \times n \times m''}$  that has been fused with temporal dimensions. Similarly, we can obtain the output of the TGAT of the daily periodicity and weekly periodicity,  $\mathbf{Y}'_{daily}$  and  $\mathbf{Y}'_{weekly}$ , respectively.

### C. Fusion Layer

After obtaining the spatio-temporal joint features  $\mathbf{Y}'_{resent}$ ,  $\mathbf{Y}'_{daily}$ , and  $\mathbf{Y}'_{weekly}$ , we need to fuse the three groups of time-periodic. According to the actual situation, different periodic features have different effects on the predicted TF, and inspired by the attention mechanism [30], we apply different attention weights to the TF values of different time-periodic. It is worth noting that, to achieve the same traffic node with different weight distribution in different time-periodic. Inspired by the Hadamard product, we design three weight matrices to merge with them. The fused features  $\mathbf{Y}_{fusion}$  are expressed as

$$\begin{aligned} \mathbf{Y}_{fusion} &= \mathbf{W}_r \circ \mathbf{Y}'_{resent} + \mathbf{W}_d \circ \mathbf{Y}'_{daily} + \mathbf{W}_w \circ \mathbf{Y}'_{weekly}, \\ \text{s.t. } & \mathbf{W}_r + \mathbf{W}_d + \mathbf{W}_w = \mathbf{ones}, \end{aligned} \quad (13)$$

where  $\mathbf{W}_r \in \mathcal{R}^{t \times n \times m''}$ ,  $\mathbf{W}_d \in \mathcal{R}^{t \times n \times m''}$  and  $\mathbf{W}_w \in \mathcal{R}^{t \times n \times m''}$  are attention weight matrices.  $\mathbf{ones} \in \mathcal{R}^{t \times n \times m''}$  is the matrix with all elements being 1.  $\circ$  is the Hadamard product. An example of a two-dimensional matrix is shown in (14).

$$\begin{aligned} \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix} \circ \begin{bmatrix} b_{11} & b_{12} & b_{13} \\ b_{21} & b_{22} & b_{23} \\ b_{31} & b_{32} & b_{33} \end{bmatrix} \\ = \begin{bmatrix} a_{11}b_{11} & a_{12}b_{12} & a_{13}b_{13} \\ a_{21}b_{21} & a_{22}b_{22} & a_{23}b_{23} \\ a_{31}b_{31} & a_{32}b_{32} & a_{33}b_{33} \end{bmatrix}. \end{aligned} \quad (14)$$

### D. Fully-Connected Layers

We use a two-layer connection layer to predict the TF value  $\hat{\mathbf{Y}} \in \mathcal{R}^{t \times n \times m}$ , and the forward calculation process is formulated as

$$\hat{\mathbf{Y}} = fc_2(fc_1(\mathbf{Y}_{fusion})), \quad (15)$$

where  $fc_1$  and  $fc_2$  are the first and last fully connected layers, respectively.

## V. PERFORMANCE EVALUATION AND ANALYSIS

To evaluate the performance of FDSA-STG and prove its effectiveness, we conduct extensive experiments on real datasets using the Performance Measurement System, which is provided by the California Department of Transportation (California PeMS). The detailed configuration and the corresponding instructions of the experiments are articulated in the following subsections. The evaluation results on prediction performance are superior compared to current representative models.

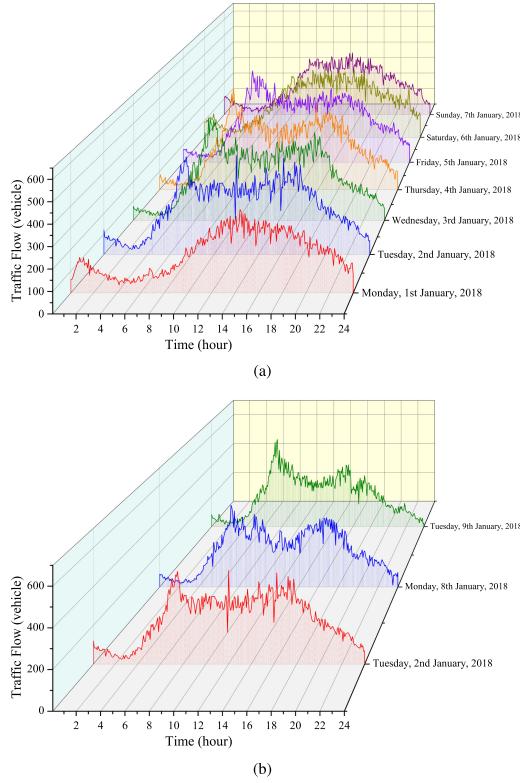


Fig. 6. The TFs at the first observation point from Monday, 1st January, 2018 to Tuesday, 9th January, 2018.

#### A. Dataset Briefing

PeMS contains real datasets of more than 39,000 detectors collected on highways distributed in many important metropolitan areas of California. In fact, it is widely used in the open system of TFP. The data is collected from the sensors every 30 seconds and aggregated in a 5-minute time slice, so each observation point contains 288-time slices of data per day. We use three-dimensional features: traffic flow information, average vehicle speed, and average lane occupancy rate. Each TF contains the location information of the sensor. In addition, we adopt the same settings as [24], and have respectively selected data sets constructed from the San Francisco Bay area in the third largest district and the San Bernardino area in the eighth district: PeMSD4 and PeMSD8, which are described as follows.

*PeMSD4:* This dataset is collected from January 2018 to February 2018, lasting about two months with a total of 59 days. It is mainly collected from a total of 3,848 detectors on 29 highways. To ensure the effect of randomness, we have removed the redundant data collected by the detectors separated by 3.5 miles and preserved the TFs of 307 detectors. Fig. 6(a) shows the TFs at the first observation point from Monday, 1st January, 2018, to Sunday, 7th January, 2018. Fig. 6(b) illustrates the recent periodicity, daily periodicity, and weekly periodicity of TFs. We can observe that TFs show a strong periodic characteristic, and the degree of correlation between different periods is different. This is also one of the starting points for our expectation of self-attention, which dynamically pays attention to time characteristics and fuses periodic characteristics.

*PeMSD8:* This dataset is collected from July 2016 to August 2016 and lasts for two months with a total of 62 days, which is mainly collected from a total of 1,979 detectors deployed along 9 highways. Similarly, using the above method of removing redundant data, the TF data of 177 detectors are preserved.

Regarding the division of the dataset, it should be noted that the data of the first 50 days is used as the training set, and the remaining is used as the test set.

#### B. Evaluation Metrics

To compare with existing research, we choose the average absolute error (MAE) and root mean square error (RMSE), which are widely employed in TFP, as the evaluation matrices. They are formulated as

$$\text{MAE} = \frac{1}{n} \sum_{i=1}^n |t_{-d_i} - \hat{t}_{-d_i}|, \quad (16)$$

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^n (t_{-d_i} - \hat{t}_{-d_i})^2}, \quad (17)$$

where  $t_{-d}$  is the raw value of TF while  $\hat{t}_{-d}$  is the predicted value of TF.

#### C. Baseline Models

In this section, FDSA-STG is employed for experimental comparison with several existing advanced TFP methods to verify its effectiveness. The baselines used include traditional statistical analysis methods (HA, VAR [20], and ARIMA [19]) and deep learning-based methods (LSTM [37], GeoMAN [39], ASTGCN [24]).

It should be noted that, for the fairness of comparison, except for HA and GeoMAN, all baselines are based on the source codes enthusiastically provided by the original authors and run following the recommendations of the original paper. For HA, we use the average value of the historical time period for prediction. For GeoMAN, since the source code is not publicly available, we implemented it carefully according to the original paper's recommendations. In addition, all baselines are run on the same processed dataset and under the same GPU environment. To ensure the rigor of the experiment, we try our best to eliminate the influence of interference factors carefully. Moreover, all experiments are run in Anaconda 4.8.3 as well as Tesla P100 environment.

#### D. Performance Comparison

Table III illustrates the average performance of FDSA-STG and each baseline predicting the value of TFs in the next hour. The smaller the value of MAE and RMSE, the more accurate the prediction of the method while the smaller the discipline of error, the better the performance of the model. It can be found that FDSA-STG is superior to all baselines in terms of MAE and RMSE. For example, on PeMSD4, compared to the best baseline ASTGCN, its MAE is reduced by 2.04, and RMSE is reduced by 4.37. On PeMSD8, its MAE is reduced by 1.50, and RMSE is reduced by 3.02, which indicates that FDSA-STG generates

TABLE III  
THE PERFORMANCE COMPARISON OF PREDICTION IN THE NEXT HOUR

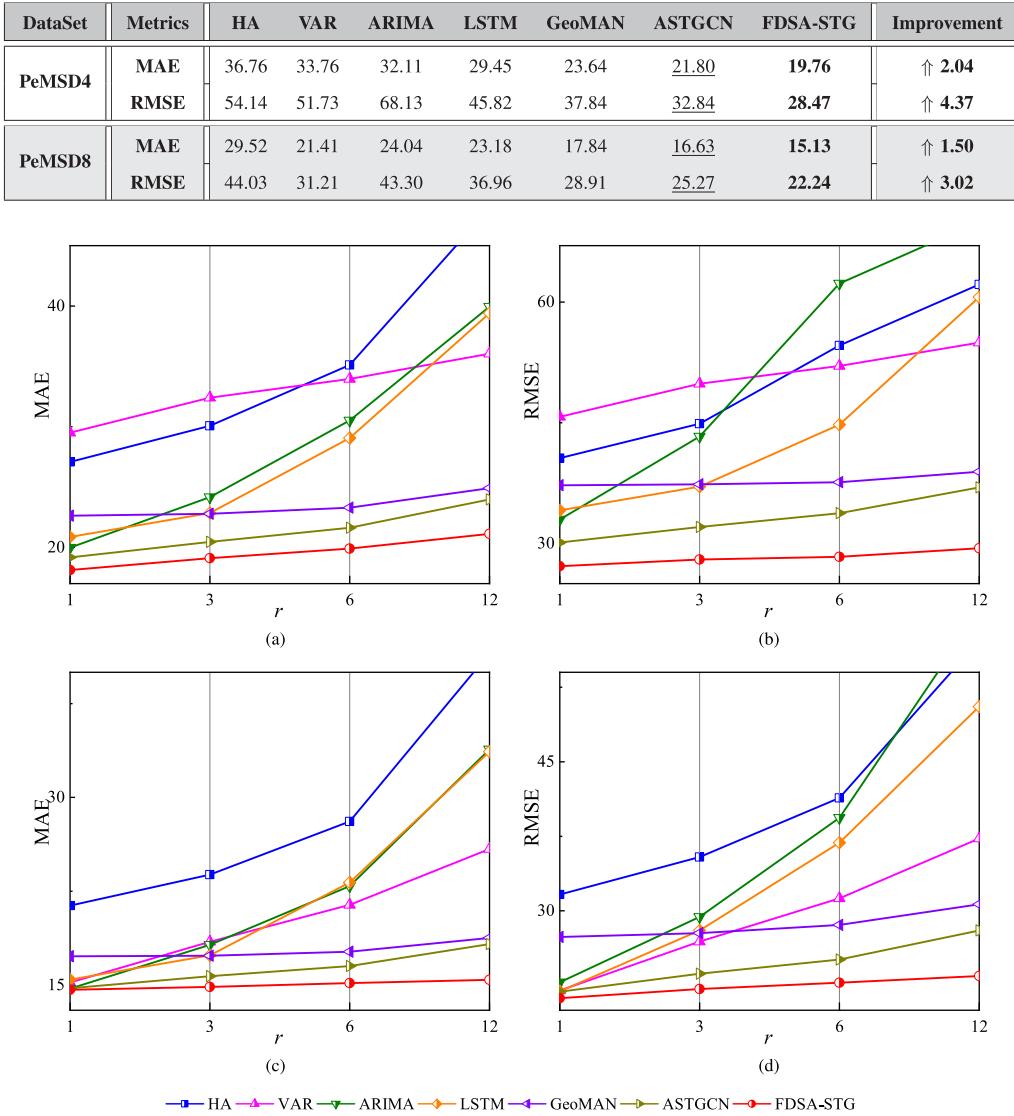


Fig. 7. The performance variation with the increase of the sampling interval  $r$ . (a) MAE @ PeMSD4. (b) RMSE @ PeMSD4. (c) MAE @ PeMSD8. (d) RMSE @ PeMSD8.

less errors compared to other baselines in the prediction process of TFs.

In addition, the sampling interval  $r$  is set to 1, 3, 6, 12, *i.e.*, we utilize the historical TF of the past hour to predict the TF of the next 5 minutes, 15 minutes, half an hour, and one hour. The performance variation of FDSA-STG and each baseline with the increase of  $r$  is shown in Fig. 7. As the prediction duration increases, the prediction accuracy of all baselines gradually decreases, and the error gradually increases. Especially for HA, ARIMA, VAR, and LSTM, the errors vary greatly. Among them, VAR considers more comprehensively than the other three in the time dimension, so the variation is relatively slight. Both GeoMAN and ASTGCN consider the spatio-temporal correlation at the same time, so the increase in the error of the performance variation with the increase of time is minor.

Due to FDSA-STG fully and dynamically paying attention to the different characteristics of TF in all aspects, it excellently fits the spatio-temporal characteristics, so the prediction error increases slowly at different time lengths and the most dependable performance is obtained.

#### E. Prediction Accuracy

We choose a random day for prediction and compare it with actual results to observe the specific effect of FDSA-STG's TFP. We take TFP from 5th February, 2018, to 11th February, 2018, on PeMSD4, as an example, and the comparison with the actual observed TFs is manifested in Fig. 8. It can be told that the TFs predicted by FDSA-STG have a similar pattern of variation to the actual observed TFs, which has already well matched with

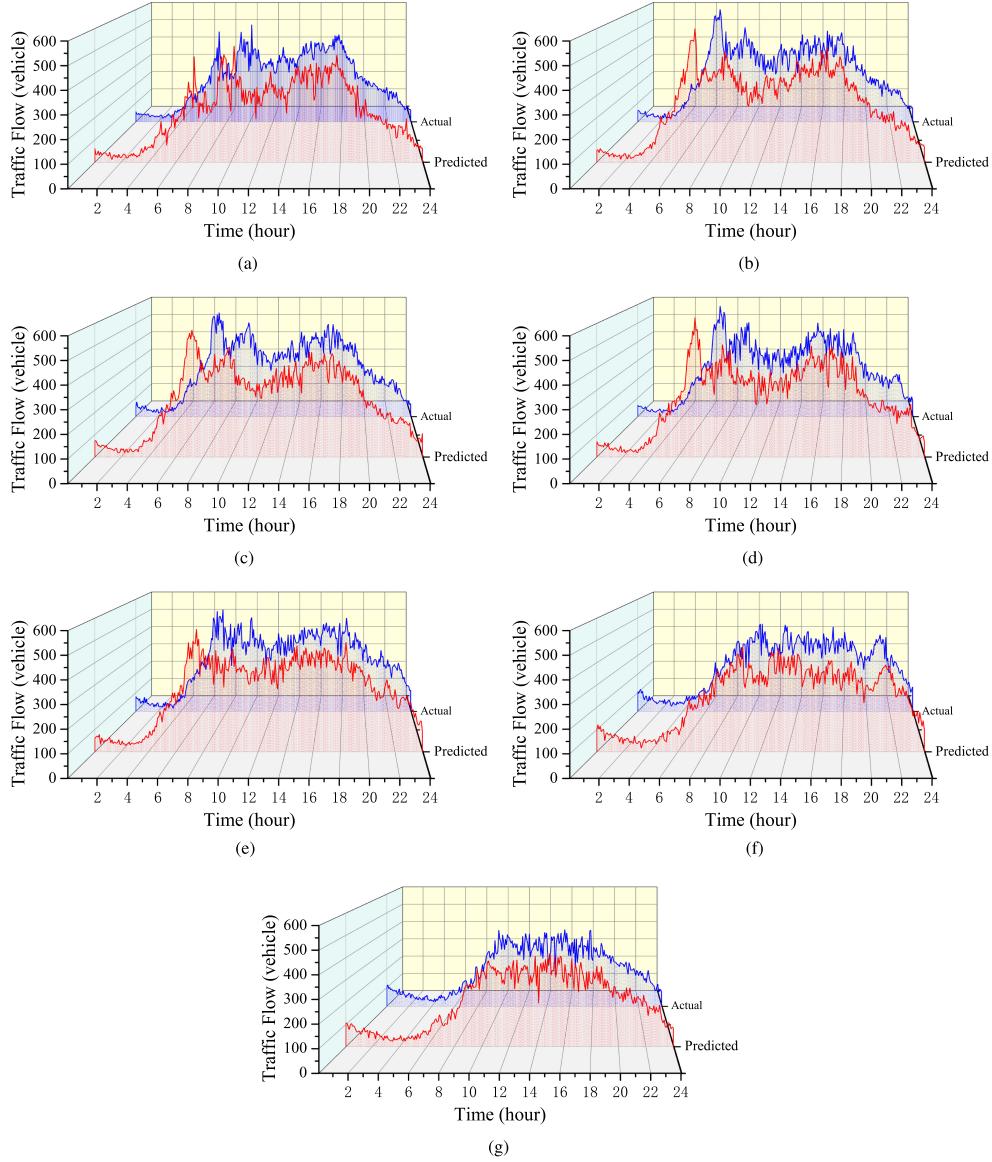


Fig. 8. Comparison between the TFs at the first observation point predicted by FDSA-STG and the actual observed TF from Monday, 5th February, 2018, to Sunday, 11th February, 2018. (a) Monday, 5th February, 2018. (b) Tuesday, 6th February, 2018. (c) Wednesday, 7th February, 2018. (d) Thursday, 8th February, 2018. (e) Friday, 9th February, 2018. (f) Saturday, 10th February, 2018. (g) Sunday, 11th February, 2018.

the actual characteristics. Even if there are slight fluctuations in some places, the magnitude of changes is not large and the changes are all within a minor range of actual observations. This is mainly derived from dynamic attention and integration of the periodic and various dimensions characteristics, and meticulous fitting of the true distribution characteristics of TFs. Consequently, the prediction FDSA-STG is completely valid, and it can effectively provide meaningful TFP services.

#### F. Computational Complexity

In addition, computational complexity is also an important indicator to measure model performance, which is mainly reflected by training time. Therefore, to further measure and illustrate the computational overhead of FDSA-STG, we conduct a comparative experiment of computational complexity. It should be noted

that the traditional statistical analysis method does not consider the nonlinear characteristics of the model, and the algorithm is relatively simple, so the computational complexity is low. However, after the previous analysis, traditional statistical analysis methods have greater prediction errors and poorer stability than deep learning-based methods. Therefore, we only compare with deep learning methods. In addition, to ensure the rigor and fairness of the results, FDSA-STG and the comparison baselines are run in the same experimental environment with the same GPU.

Specifically, we conduct comparative experiments on the PeMSD4 dataset, and record the result comparison of training time within an epoch in Table IV. LSTM has the lowest complexity, mainly due to its simple algorithm. However, it does not fully consider the spatial characteristics and dynamic correlation of TF, which makes the prediction performance of

TABLE IV  
COMPUTATIONAL COMPLEXITY COMPARISON OF FDSA-STG AND OTHER  
DEEP LEARNING-BASED BASELINES

Model	LSTM	GeoMAN	ASTGCH	FDSA-STG
Training Time	43.52 s	59.63 s	49.47 s	51.18 s

the algorithm relatively poor. GeoMAN is based on a multi-level attention mechanism to make its time consumption high. But it has to be said that GeoMAN has outstanding performance in geographic perception, the training time of FDSA-STG is slightly longer than that of ASTGCH, mainly due to the fact that we adopt a more complex attention mechanism to obtain more precise prediction accuracy. Due to the significant improvement in performance, the computational cost of FDSA-STG is within an acceptable range. Therefore, FDSA-STG has promising competitiveness compared with existing TFP algorithms.

## VI. SUMMARY AND FUTURE WORK

In this paper, we propose a traffic flow prediction method, entitled *Fully Dynamic Self-Attention Spatio-Temporal Graph Networks* (FDSA-STG), which alleviates the traffic pressure of regional emergencies and coordinate resource allocation decisions in advance. To address the problem that existing methods barely consider the dynamic correlation of spatial, temporal, and periodic characteristics, we correspondingly devise the spatial graph attention component, the temporal graph attention component, and the fusion layer to fit the original intricate structural relationship of the TN. In this way, FDSA-STG can dynamically focus and integrate the complex correlations in the spatial, temporal, and periodic characteristics of the TN, where three groups of similar structures are designed to extract the flow characteristics of recent periodicity, daily periodicity, and weekly periodicity. To testify the performance, extensive experiments are performed on the California PeMS dataset, while the results manifest that FDSA-STG exceeds existing leading research and exhibits superior performance in regards to each evaluation metric.

In the future work, we expect to advance FDSA-STG to the emergent fluctuation based on the existent works of general intelligent TFP. Specifically, sudden traffic accidents, temporary regional activities, road construction, etc., will have an enormous impact on regional traffic flow. In this case, how to accurately predict the TFs of each timestamp in different locations to provide meaningful emergency decision-making, is our next research focus. In addition, how to efficiently provide correct and reasonable optimal planning strategies for congested intersections is also one of the key issues in ITS. Therefore, we further plan to investigate the strategy of vehicle diversion on congested road sections, i.e., route planning, to help promote the development of ITS. Moreover, factors such as season and weather may have a particular relationship with TF, which is another promising research topic to follow. We also expect to use more advanced sensors to sample multi-dimensional traffic information in the future, to visualize the interesting effects of season, weather, and other factors of TFP.

## REFERENCES

- [1] A. Zhu, M. Ma, S. Guo, S. Yu, and L. Yi, "Adaptive multi-access algorithm for multi-service edge users in 5G ultra-dense heterogeneous networks," *IEEE Trans. Veh. Technol.*, vol. 70, no. 3, pp. 2807–2821, Mar. 2021.
- [2] F. Tang, B. Mao, N. Kato, and G. Gui, "Comprehensive survey on machine learning in vehicular network: Technology, applications and challenges," *IEEE Commun. Surv. Tut.*, vol. 23, no. 3, pp. 2027–2057, Jul.–Sep. 2021.
- [3] J. Liu, X. Wang, G. Yue, and S. Shen, "Data sharing in VANETs based on evolutionary fuzzy game," *Future Gener. Comput. Syst.*, vol. 81, pp. 141–155, 2018.
- [4] P. Zhang *et al.*, "Artificial intelligence technologies for COVID-19-like epidemics: Methods and challenges," *IEEE Netw.*, vol. 35, no. 3, pp. 27–33, May/Jun. 2021.
- [5] K. Niu, C. Cheng, J. Chang, H. Zhang, and T. Zhou, "Real-time taxi-passenger prediction with L-CNN," *IEEE Trans. Veh. Technol.*, vol. 68, no. 5, pp. 4122–4129, May 2019.
- [6] B. Yan, G. Wang, J. Yu, X. Jin, and H. Zhang, "Spatial-temporal chebyshev graph neural network for traffic flow prediction in IoT-based ITs," *IEEE Internet Things J.*, to be published, doi: [10.1109/IJOT.2021.3105446](https://doi.org/10.1109/IJOT.2021.3105446).
- [7] P. Zhang, C. Wang, G. S. Aujla, N. Kumar, and M. Guizani, "IoV scenario: Implementation of a bandwidth aware algorithm in wireless network communication mode," *IEEE Trans. Veh. Technol.*, vol. 69, no. 12, pp. 15774–15785, Dec. 2020.
- [8] J. Lu *et al.*, "Can steering wheel detect your driving fatigue?," *IEEE Trans. Veh. Technol.*, vol. 70, no. 6, pp. 5537–5550, Jun. 2021.
- [9] J. Wang, Y. Zhang, Y. Wei, Y. Hu, X. Piao, and B. Yin, "Metro passenger flow prediction via dynamic hypergraph convolution networks," *IEEE Trans. Intell. Transp. Syst.*, vol. 22, no. 12, pp. 7891–7903, Dec. 2021.
- [10] P. Zhang, C. Wang, C. Jiang, N. Kumar, and Q. Lu, "Resource management and security scheme of ICPSs and IoT based on VNE algorithm," *IEEE Internet Things J.*, to be published, doi: [10.1109/IJOT.2021.3068158](https://doi.org/10.1109/IJOT.2021.3068158).
- [11] Y. Zhao *et al.*, "A vehicle handling inverse dynamics method for emergency avoidance path tracking based on adaptive inverse control," *IEEE Trans. Veh. Technol.*, vol. 70, no. 6, pp. 5470–5482, Jun. 2021.
- [12] Q. Li, Z. Cao, M. Tanveer, H. M. Pandey, and C. Wang, "A semantic collaboration method based on uniform knowledge graph," *IEEE Internet Things J.*, vol. 7, no. 5, pp. 4473–4484, May 2020.
- [13] Q. Li, Z. Cao, J. Zhong, and Q. Li, "Graph representation learning with encoding edges," *Neurocomputing*, vol. 361, pp. 29–39, 2019.
- [14] X. Song, Y. Guo, N. Li, and L. Zhang, "Online traffic flow prediction for edge computing-enhanced autonomous and connected vehicles," *IEEE Trans. Veh. Technol.*, vol. 70, no. 3, pp. 2101–2111, Mar. 2021.
- [15] G. Gui, Z. Zhou, J. Wang, F. Liu, and J. Sun, "Machine learning aided air traffic flow analysis based on aviation Big Data," *IEEE Trans. Veh. Technol.*, vol. 69, no. 5, pp. 4817–4826, May 2020.
- [16] F. Wang *et al.*, "6G-enabled short-term forecasting for large-scale traffic flow in massive IoT based on time-aware locality-sensitive hashing," *IEEE Internet Things J.*, vol. 8, no. 7, pp. 5321–5331, Apr. 2021.
- [17] Z. Fang, S. Shen, J. Liu, W. Ni, and A. Jamalipour, "New NOMA-based two-way relay networks," *IEEE Trans. Veh. Technol.*, vol. 69, no. 12, pp. 15314–15324, Dec. 2020.
- [18] M. Li, P. Tong, M. Li, Z. Jin, J. Huang, and X.-S. Hua, "Traffic flow prediction with vehicle trajectories," in *Proc. AAAI Conf. Artif. Intell.*, 2021, pp. 294–302.
- [19] 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.
- [20] E. Zivot and J. Wang, "Vector autoregressive models for multivariate time series," in *Modeling Financial Time Series with S-Plus*, Berlin, Germany: Springer, 2006, pp. 385–429.
- [21] J. Liu *et al.*, "Intelligent jamming defense using DNN stackelberg game in sensor edge cloud," *IEEE Internet Things J.*, vol. 9, no. 6, pp. 4356–4370, Mar. 2022.
- [22] Y. Lv, Y. Duan, W. Kang, Z. Li, and F.-Y. Wang, "Traffic flow prediction with Big Data: A deep learning approach," *IEEE Trans. Intell. Transp. Syst.*, vol. 16, no. 2, pp. 865–873, Apr. 2015.
- [23] Q. Li, L. Li, W. Wang, Q. Li, and J. Zhong, "A comprehensive exploration of semantic relation extraction via pre-trained CNNs," *Knowl.-Based Syst.*, vol. 194, 2020, Art. no. 105488.
- [24] S. Guo, Y. Lin, N. Feng, C. Song, and H. Wan, "Attention based spatial-temporal graph convolutional networks for traffic flow forecasting," in *Proc. AAAI Conf. Artif. Intell.*, 2019, pp. 922–929.

- [25] Z. Zhao, W. Chen, X. Wu, P. C. Chen, and J. Liu, "LSTM network: A deep learning approach for short-term traffic forecast," *IET Intell. Transport Syst.*, vol. 11, no. 2, pp. 68–75, 2017.
- [26] J. Zhang, Y. Zheng, D. Qi, R. Li, X. Yi, and T. Li, "Predicting citywide crowd flows using deep spatio-temporal residual networks," *Artif. Intell.*, vol. 259, pp. 147–166, 2018.
- [27] F. Zhao, G.-Q. Zeng, and K.-D. Lu, "EnLSTM-WPEO: Short-term traffic flow prediction by ensemble LSTM, NNCT weight integration, and population extremal optimization," *IEEE Trans. Veh. Technol.*, vol. 69, no. 1, pp. 101–113, Jan. 2020.
- [28] H. Yan, X. Ma, and Z. Pu, "Learning dynamic and hierarchical traffic spatiotemporal features with transformer," *IEEE Trans. Intell. Transp. Syst.*, to be published, doi: [10.1109/TITS.2021.3102983](https://doi.org/10.1109/TITS.2021.3102983).
- [29] H. Zheng, F. Lin, X. Feng, and Y. Chen, "A hybrid deep learning model with attention-based conv-LSTM networks for short-term traffic flow prediction," *IEEE Trans. Intell. Transp. Syst.*, vol. 22, no. 11, pp. 6910–6920, Nov. 2021.
- [30] A. Vaswani *et al.*, "Attention is all you need," in *Proc. Adv. Neural Inf. Process. Syst.*, 2017, pp. 5998–6008.
- [31] P. Veličković, G. Cucurull, A. Casanova, A. Romero, P. Lio, and Y. Bengio, "Graph attention networks," 2017, *arXiv:1710.10903*.
- [32] P. Zhang, C. Jiang, X. Pang, and Y. Qian, "STEC-IoT: A security tactic by virtualizing edge computing on IoT," *IEEE Internet Things J.*, vol. 8, no. 4, pp. 2459–2467, Feb. 2021.
- [33] P. Zhang, C. Wang, C. Jiang, and Z. Han, "Deep reinforcement learning assisted federated learning algorithm for data management of IIoT," *IEEE Trans. Ind. Inform.*, vol. 17, no. 12, pp. 8475–8484, Dec. 2021.
- [34] B. Mao, F. Tang, Y. Kawamoto, and N. Kato, "AI models for green communications towards 6G," *IEEE Commun. Surv. Tut.*, vol. 24, no. 1, pp. 210–247, Jan.–Mar. 2022.
- [35] Z. M. Fadlullah *et al.*, "State-of-the-art deep learning: Evolving machine intelligence toward tomorrow's intelligent network traffic control systems," *IEEE Commun. Surv. Tut.*, vol. 19, no. 4, pp. 2432–2455, Oct.–Dec. 2017.
- [36] P. Zhang, Y. Wang, N. Kumar, C. Jiang, and G. Shi, "A security and privacy-preserving approach based on data disturbance for collaborative edge computing in social IoT systems," *IEEE Trans. Computat. Social Syst.*, vol. 9, no. 1, pp. 97–108, Feb. 2022.
- [37] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural computation*, vol. 9, no. 8, pp. 1735–1780, Nov. 1997.
- [38] Y. Wu, H. Tan, L. Qin, B. Ran, and Z. Jiang, "A hybrid deep learning based traffic flow prediction method and its understanding," *Transp. Res. Part C: Emerg. Technol.*, vol. 90, pp. 166–180, 2018.
- [39] Y. Liang, S. Ke, J. Zhang, X. Yi, and Y. Zheng, "GeoMAN: Multi-level attention networks for geo-sensory time series prediction," in *Proc. 27th Int. Joint Conf. Artif. Intell.*, 2018, pp. 3428–3434.
- [40] T. N. Kipf and M. Welling, "Semi-supervised classification with graph convolutional networks," 2016, *arXiv:1609.02907*.



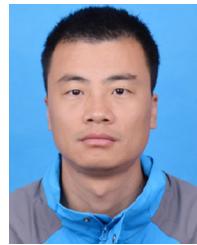
**Youxiang Duan** received the B.S. degree from the College of Computer and System Science, Nankai University, Tianjin, China, in 1986, and the Ph.D. degree from the School of Geosciences, China University of Petroleum (East China), Qingdao, China, in 2017. He is currently engaged in teaching and scientific research with the College of Computer Science and Technology, China University of Petroleum (East China). His main research interests include machine learning, network, and service computing.



**Ning Chen** (Graduate Student Member, IEEE) is currently working toward the master's degree from the College of Computer Science and Technology, China University of Petroleum (East China), Qingdao, China. His research interests include deep learning, network virtualization, and future network architecture.



**Shigen Shen** (Member, IEEE) received the B.S. degree in fundamental mathematics from Zhejiang Normal University, Jinhua, China, in 1995, the M.S. degree in computer science and technology from Zhejiang University, Hangzhou, China, in 2005, and the Ph.D. degree in pattern recognition and intelligent systems from Donghua University, Shanghai, China, in 2013. He is currently a Professor with the Department of Computer Science and Engineering, Shaoxing University, Shaoxing, China. His research interests include Internet of Things, cyber security, cloud computing, and game theory. He is currently a Member of the editorial review board of the *Journal of Organizational and End User Computing*.



**Peiyong Zhang** (Member, IEEE) received the Ph.D. degree from the School of Information and Communication Engineering, University of Beijing University of Posts and Telecommunications, Beijing, China, in 2019. He is currently an Associate Professor with the College of Computer Science and Technology, China University of Petroleum (East China), Qingdao, China. He has authored or coauthored multiple IEEE/ACM Transactions/Journal/Magazine papers since 2016, such as the IEEE TII, IEEE T-ITS, IEEE TVT, IEEE TNSE, IEEE TNSM, IEEE TETC, IEEE NETWORK, IEEE ACCESS, IEEE IoT-J, ACM TALLIP, COMPUT COMMUN, and IEEE COMMUN MAG. His research interests include semantic computing, future internet architecture, network virtualization, and artificial intelligence for networking. He was the Technical Program Committee of IEEE ICC'22, DPPR 2021, ISCIT 2016, ISCIT 2017, ISCIT 2018, ISCIT 2019, Globecom 2021, Globecom 2019, COMNETSAT 2020, SoftIoT 2021, IWCMC-Satellite 2019, IWCMC-Satellite 2020, and IWCMC-Satellite 2022.



**Youyang Qu** (Member, IEEE) received the B.S. degree in mechanical automation and the M.S. degree in software engineering from the Beijing Institute of Technology, Beijing, China, in 2012 and 2015, respectively, and the Ph.D. degree from the School of Information Technology, Deakin University, Melbourne, VIC, Australia, in 2019. He is currently a Postdoctoral Research Fellow. He has more than 40 publications, including high quality journals and conference papers, such as IEEE TII, IEEE TNSE, and IEEE IOTJ. His research interests include dealing with security and customizable privacy issues in blockchain, social networks, machine learning, and IoT. He is active in communication society and was an organizing committee Member in SPDE 2020, BigSecurity 2021, and Tridentcom 2021.



**Shui Yu** (Senior Member, IEEE) received the Ph.D. degree from Deakin University, Melbourne, VIC, Australia, in 2004. He is currently a Professor with the School of Computer Science, University of Technology Sydney, Sydney, NSW, Australia. He has authored or coauthored four monographs and edited two books, and more than 400 technical papers, including top journals and top conferences, such as IEEE TPDS, TC, TIFS, TMC, TKDE, TETC, ToN, and INFOCOM. His h-index is 64. His research interests include Big Data, security and privacy, networking, and mathematical modeling. Dr. Yu initiated the research field of networking for Big Data in 2013, and his research outputs have been widely adopted by industrial systems, such as Amazon cloud security. He is currently on a number of prestigious editorial boards, including IEEE COMMUNICATIONS SURVEYS AND TUTORIALS(Area Editor), *IEEE Communications Magazine*, and IEEE INTERNET OF THINGS JOURNAL. He was a Distinguished Lecturer of IEEE Communications Society during 2018–2021. He is a Distinguished Visitor of IEEE Computer Society, a voting Member of IEEE ComSoc Educational Services board, and an elected Member of the Board of Governor of IEEE Vehicular Technology Society.