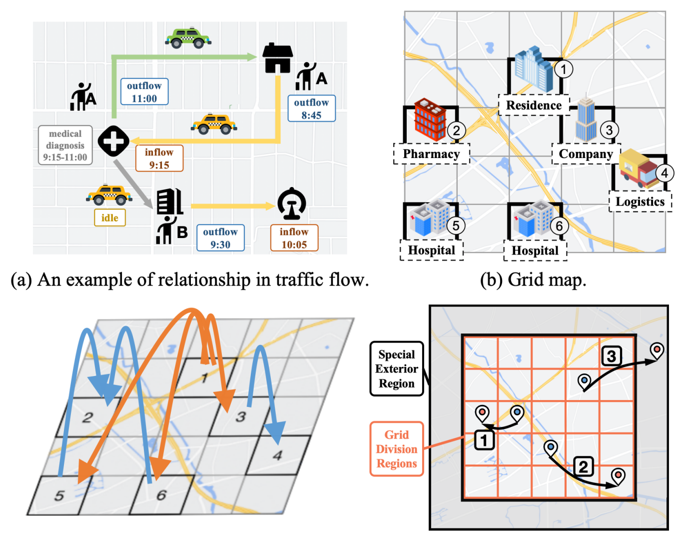
2023 IEEE International Conference on Big Data (BigData) | 979-8-3503-2445-7/23/$31.00 ©2023 IEEE | DOI: 10.1109/BigData59044.2023.10386746

Spatial-Temporal Flow Holistic Interaction Graph Convolution Network for Bidirectional Traffic Flow Forecasting

Canyang Zhang[1], Qi Zheng[1] and Yaying Zhang[1,2]✉
[1] Key Laboratory of Embedded System and Service Computing, Ministry of Education, Tongji University
[2] Shanghai Artificial Intelligence Laboratory
Shanghai, China
{zcy, zhengqi97, yaying.zhang}@tongji.edu.cn

***Abstract*—Traffic flow forecasting is indispensable in modern urban life. Considering the complexity, variability and strong timeliness of traffic flow, traffic flow forecasting is a worth exploring but challenging research field. To achieve better traffic flow forecasting effect, we focus on two critical aspects that assume noteworthy importance: i) the features inside the traffic outflows and inflows. ii) the supplementary information regarding exterior region which is the area outside the grid division regions. To address these challenges, we propose a novel deep learning model Spatial-Temporal Flow Holistic Interaction Graph Convolution Network (STHGCN). In STHGCN, graph convolution based modules are applied through multi-step simulation. An exterior region feature estimation module is designed to estimate the influence of the special exterior region through the characteristics of complete trajectories, which enables a more comprehensive reasoning for traffic flow forecasting in grid division regions. Furthermore, a flow feature fusion integrator and stackable convolution modules are proposed to aggregate the intermediate features extracted from various perspectives, which simulate the constantly-updating and interlinked states of traffic flows through the process of multi-layer feature separation and fusion. We conduct extensive experiments on real-world traffic datasets and our proposed model outperforms all baselines.**



(a) An example of relationship in traffic flow. (b) Grid map. (c) Regional flow transfers. (d) Different types of transfers.

Fig. 1. Examples of traffic transfers and traffic flow relationship.

I. Introduction

As a crucial component of intelligent transportation systems, traffic flow prediction has significant implications for various aspects of urban life. Despite the significant progress in traffic flow forecasting through numerous studies, we still have two noteworthy observations.

Firstly, there are inter-relationships and intra-relationships in traffic flows. The *inter-relationships* refer to the interaction of traffic inflow and outflow among regions (inflow-outflow). There exists a strong relationship between traffic inflow and outflow which is due to the fact that regional traffic flow data is usually derived from trajectory information statistics. As shown in Fig 1(a), passenger A hails a taxicab at 8:45 from home to hospital and arrives at 9:15. The boarding and disembarkation of this trip corresponds to an outflow of the departure region and an inflow of the target region with cause-and-effect sequence. After the arrival, the passenger receives a medical diagnosis and leaves at 11:00. The taxi driver drives around and receives the next order at a nearby hotel at 9:30. In this situation, the inflow of hospital region at 9:15 leads to the outflow of the nearby regions at 9:30 and 11:00. The *intra-relationships* in traffic flows refer to the correlations among either the inflows of regions or the outflows of regions (inflow-inflow, outflow-outflow). For example, there are similarities in the intra-relationships within traffic flows. As shown in Fig. 1(b) and Fig. 1(c), in the morning of weekday, the inflow patterns of region 3, 5, 6 are similar while the outflow patterns are not.

Secondly, most of the previous traffic flow forecasting models merely pay attention to the grid division regions. However, they neglect the traffic transfers between the grid division regions and the area outside. As shown in Fig. 1(d), transfer 1 and 2 are typical transfers among grid division regions.

This work was partly supported by the National Key Research and Development Program of China under Grant 2022YFB4501700 and the Foundation of State Key Laboratory of Pollution Control and Resource Reuse, Tongji University(No.2022-4-ZD-05).

While transfer 3, occurring between grid division regions and the special exterior region, warrants extra attention. We can regard the outside area as a large special region. In the context of traffic prediction, this region operates analogously to others, exhibiting characteristics such as traffic flow volume and interactions with neighboring regions.

To address these challenges, we propose a **S**patial-**T**emporal Flow **H**olistic Interaction **G**raph **C**onvolution **N**etwork (STHGCN), the main contributions of this work are summarized as follows:

- The inter-relationships and intra-relationships in regional traffic flows are captured for modeling the complex spatial-temporal dependency.
- A special exterior region is taken into account. The supplementary traffic features of such region are estimated and used to assist the flow prediction of grid division regions.
- We conduct experiments on four real-world traffic datasets. The results show that our proposed model obtains significant improvements over the baseline methods.

II. Related Work

Traffic flow forecasting has always been a hot topic. Typically, such forecasting endeavors commence by considering both the temporal and spatial dimensions.

For temporal feature modeling of traffic data, RNN [1] and its variants are commonly used. Temporal Convolutional Networks (TCN) employs convolution structure at temporal processing [2]. Yu et al. adopt pure convolution structure to extract spatio-temporal information from the traffic graph data [3]. Oord et al. propose dilated causal convolution for the efficient capture of long-range temporal dependencies [4]. Researchers also utilize the transformer model for time series processing [5]. For spatial feature modeling, the graph convolutional neural network has been widely applied. For traffic data, Li et al. propose a convolutional model by analoging the diffusion model [6]. More recently, transformer-based models show promising results in capturing traffic spatial feature [7]. In this paper, our proposed model STHGCN employs the TCN method based on causal convolution for temporal feature processing and graph convolution method for spatial feature processing.

For regional traffic flow prediction, there are also abundant research achievements. Considering the short-range and long-range spatial dependencies, Zhang et al. use convolution-based residual networks to simulate such dependencies among traffic regions [8]. In order to make up for the lack of dynamics analysis, Yao et al. propose a model for the spatial dynamics and the fluctuations of periodical dependencies [9]. To achieve better temporal analysis, Zhang et al. utilize a multi-scale self-attention network and temporal hierarchy aggregation layer to process temporal signals [10]. To jointly capture spatial-temporal dependencies between domains, Li et al. propose a light-weight transformer scheme [11].

Previous research have extensively explored the forecasting of traffic inflow and outflow, with many studies dedicated to this area. Among them, some studies pay attention to the relationship within traffic flows. Zhang et al. explore the mutual transitions among taxicab pick-ups and drop-offs and propose mutual transition aware framework [12]. Zhao et al. model three kinds of relationship in traffic flow respectively and integrated them at the final prediction block to obtain the prediction flow volume [13]. In our proposed STHGCN, the inter-relationships and intra-relationships at traffic flows are modeled with a stackable multi-layer structure. The proposed model conducts the traffic feature diversion and fusion in each layer, enhancing the capture of the dynamic and time-sensitive entanglement relationship.

In addition, recent studies seldom consider the traffic conditions of the area outside the grid division regions, which can be considered as a special exterior region. With a broader scope of consideration, STHGCN improves prediction accuracy by analyzing trajectories between the grid division regions and the exterior region.

III. Preliminary

A. Definition 1 (Spatial Region)

A traffic predication area can be partitioned into an $I \times J$ grid map based on the longitude and latitude where each grid denotes a spatial region and all grids are disjoint. The grid region at i^{th} row and j^{th} column can be denoted as $g_{i,j}$. Regarding each grid as a node, the grid map can be transformed into a graph $\mathcal{G} = (V, A)$, where V is the set of nodes, each of which corresponds to a unique grid or region, $|V| = N = I \times J$. $A \in \mathbb{R}^{N \times N}$ corresponds to the importance of one node to another.

B. Definition 2 (Traffic Inflow and Outflow)

Traffic inflow and outflow can be described as the movement of vehicles or people into and out of a specific area, such as a road segment, an intersection, or a given traffic region. Practically regional traffic inflows and outflows are often calculated from trajectories. Similar to [14], we represent the flow data as spatial data in time series, including the source $s = (\tau_s, x_s, y_s)$ and the destination $d = (\tau_d, x_d, y_d)$, where τ is timestamp and (x, y) is a geospatial point. Let $\mathcal{P}$ denotes a set of traffic flow transfers consisting of $(start, end)$ pairs. For a grid node $g_{i,j} \in \mathcal{G}$, its inflow and outflow during the interval t are respectively defined as:

$$X_t^{in,i,j} = |\{((s,d) \in \mathcal{P} : (x_d, y_d) \in g_{i,j} \wedge \tau_d \in t)\}|, \quad (1)$$

$$X_t^{out,i,j} = |\{((s,d) \in \mathcal{P} : (x_s, y_s) \in g_{i,j} \wedge \tau_s \in t)\}|. \quad (2)$$

C. Definition 3 (Graph Convolution)

The graph convolution operation we used can be defined as $\Theta \star_{\mathcal{G}}$, a multi-step diffusion graph convolution operation based on two adjacency matrices, which can be formulated as:

$$\Theta \star_{\mathcal{G}} (F, A_1, A_2) = \sum_{k=0}^{K-1} (\theta_{k,1}(A_1)^k + \theta_{k,2}(A_2)^k) F. \quad (3)$$

where F is the input data and A_1,A_2 are input adjacency matrices. $\theta_{k,1}$ and $\theta_{k,2}$ are the weight matrices which are not shared between graph convolutions. K represents the maximum diffusion step.

Hereafter, to facilitate the understanding of notations used in this paper, in and out appearing in the superscript of a notation indicate $inflow$ and $outflow$ respectively. The superscript $inter$, $intra$ imply the inter-relationship and intra-relationship notions, and e indicates exterior region related notions.

In our model, traffic inflow and outflow are symmetrically processed. For brevity, in the formulas that have symmetry processing, **we use superscript α to represent one flow and β to represent the other flow**. For example, when α stands for inflow, then β is outflow and vice versa.

D. Problem Definition

For the flow prediction area, given spatial region graph G and the traffic flow observations of all the grids in historical consecutive T time intervals $[X_{t-T}, \ldots, X_{t-1}]$, the traffic forecasting problem is to predict the future traffic flows of the grid area $X_t \in \mathbb{R}^{2\times I\times J}$ in the next time interval.

IV. Methodology

The overall framework of our proposed STHGCN is shown in Fig. 2. Historical traffic flow data and time period information are put into the time period fusion block for time periodic information integration. There are stackable Traffic Flow Separation and Interaction Graph Convolution Modules. A single TSIM implements one Exterior Region Feature estimation Module , two Inflow/Outflow Feature Extraction Modules and a Flow Feature Fusion Inergrator. The information from multilayer TSIMs is passed to the prediction block, which in turn integrates the skip connection features of all layers to generate the prediction.

A. Time Period Fusion Block

We focus on the *time position information*, which refers to the temporal alignment or timestamps associated with traffic data. The time period fusion block integrates the input regional traffic flow data and the time period information. All graph nodes in each time slice share the same time period information. In the form of one-hot code, the time period information encoding is comprised of two temporal hierarchies, which can be denoted as $E_d \in \mathbb{R}^{T\times 7}$ and $E_s \in \mathbb{R}^{T\times s_n}$ (7 is the number of days in a week and s_n represents the number of time slices in a day, e.g. s_n =288 if the time slice is 5 minutes).

In the time period fusion block, the embeded data M^{in}, $M^{out} \in \mathbb{R}^{T\times N\times C}$ can be obtained by the raw traffic flow data $X = \{X_{n-T}, \cdots, X_{n-1}\}$ and time period information $\{E_d, E_s\}$. T is the length of the time slices and C is the dimension of hidden features.

$$M^\alpha = Conv(X^\alpha)||\mathcal{R}(Conv(E_d))||\mathcal{R}(Conv(E_s)). \quad (4)$$

where $Conv$ represents the convolution operation [1]. $\mathcal{R}$ represents the dimension replication operation and $||$ is the concatenation operation. M^α are used as the input of the first TSIM later.

B. Exterior Region Feature Estimation Module (ERFM)

The area outside the grid division regions is regarded as a special exterior region. Since there commonly exists traffic flow transfers between grid regions and the exterior region, the traffic feature estimation and analysis of the exterior region deserve attention. The exterior region feature estimation module estimates the auxiliary features of inflow and outflow in the special exterior region, i.e. the inflow auxiliary feature $U^{in} \in \mathbb{R}^{T\times 1\times C}$ and the outflow auxiliary feature $U^{out} \in \mathbb{R}^{T\times 1\times C}$, as in Fig. 3. These two features are estimated from the input data D^{in}, $D^{out} \in \mathbb{R}^{T\times N\times C}$ of TSIM. The input of the first TSIM are the output of the above-mentioned time period fusion block, i.e. M^{in} and M^{out}, and the input of the other TSIMs are the output of the previous TSIM.

$$U^\alpha = Conv(\mathcal{M}(Conv(D^\beta), Conv(D^\alpha))), \quad (5)$$

where $\mathcal{M}$ is a replaceable merge operation in different scenarios and we use subtraction operation in this paper.

Take traffic inflow as an example. Two special adaptive adjacency matrices A_e^{in}, $A_e^{out} \in \mathbb{R}^{1\times N}$ are introduced to facilitate the simulation of traffic flow interaction between the exterior region and the grid division regions. A_e^{in} quantifies the influence of the external auxiliary inflow feature U^{in}, which originates from the exterior region on each grid division region. The influence of external auxiliary flow feature R^{in}, $R^{out} \in \mathbb{R}^{T\times N\times C}$ on N regions can be formulated as follows:

$$R^\alpha = W_e^\alpha A_e^\alpha U^\alpha. \quad (6)$$

where W_e^α is a weight matrix.

C. Inflow/Outflow Feature Extraction Module (FEM)

Inflow/Outflow feature extraction module handles the feature derivation and process in inflows or outflows for the inter-relationship and intra-information in regional traffic inflow/outflows.

Taking traffic inflow as an example, its inter-relationship with outflow and its intra-relationship with the other inflow in other regions can be separated out, as shown in Fig. 4. An interactive feature adjacency matrix $A_{inter}^{in} \in \mathbb{R}^{N\times N}$ from inflow to outflow, along with a fixed static geographic adjacency matrix $A_g \in \mathbb{R}^{N\times N}$, are used to discover the hidden interaction entanglement information in traffic flows. An inflow inner feature adjacency matrix $A_{intra}^{in} \in \mathbb{R}^{N\times N}$ and A_g are used to achieve the mining of inner features in traffic inflow. For the three matrices mentioned above, A_g is a matrix constructed according to the geographic adjacency relationship of traffic regions in the real data set. A_{inter}^{in} and A_{intra}^{in} are two adaptive adjacency matrices whose structures are consistent with the self-adaptive adjacency matrix proposed by [15]. For example, A_{inter}^{in} is derived from two node embedding dictionaries $E_{inter}^{in,1} \in \mathbb{R}^{N\times c'}$ and $E_{inter}^{in,2} \in \mathbb{R}^{c'\times N}$ with learnable parameters, which can be be expressed as:

$$A_{inter}^\alpha = SoftMax(ReLU(E_{inter}^{\alpha,1} E_{inter}^{\alpha,2})), \quad (7)$$

[1]The parameters of $Conv$ appearing in this paper are not shared with each other. For brevity, we tend not to state the specific parameters every time a convolution operation occurs.

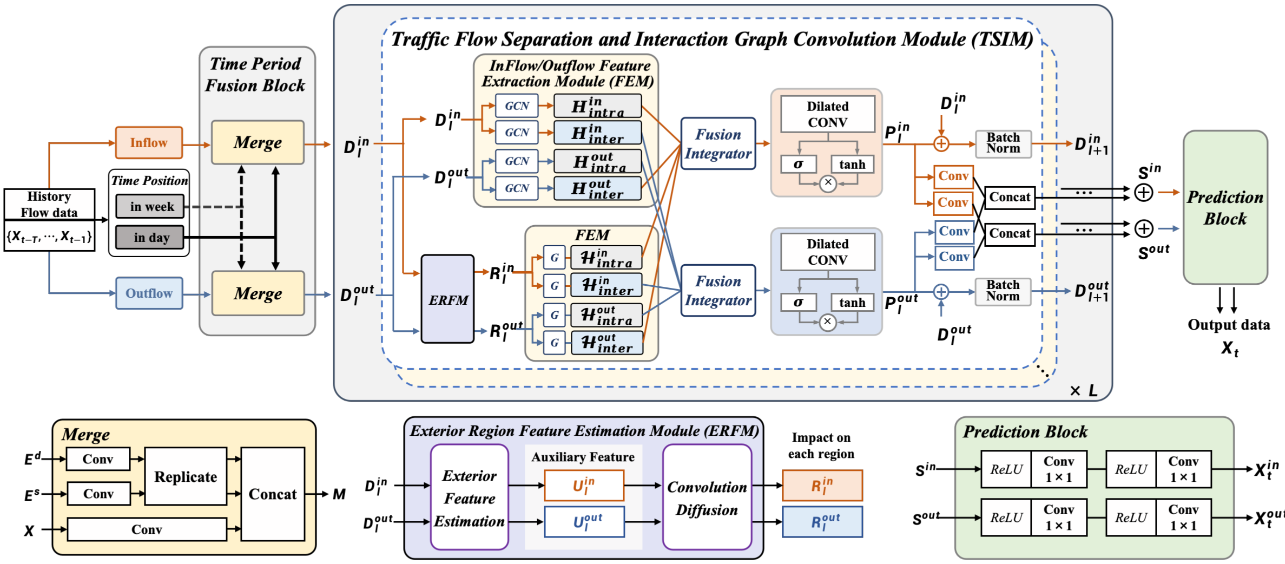


Fig. 2. Overall architecture of **S**patial-**T**emporal Flow **H**olistic Interaction **G**raph **C**onvolution **N**etwork (STHGCN).

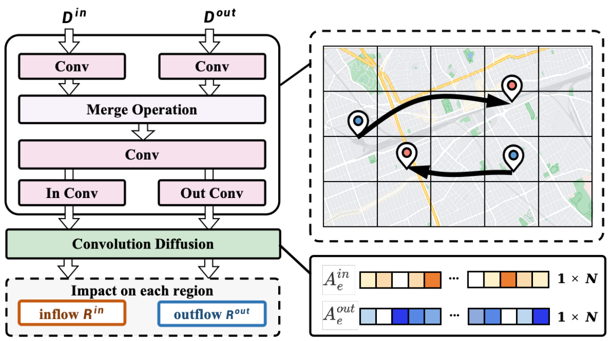


Fig. 3. The architecture of Exterior Region Feature Estimation Module.

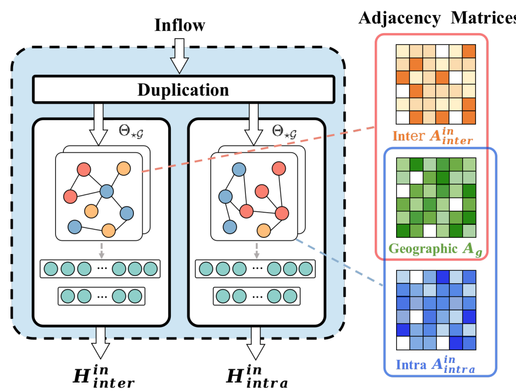


Fig. 4. The architecture of the Inflow/Outflow Feature Extraction Module (FEM). Here we present the inflow analysis as an example.

where $SoftMax$ represents the softmax function and $ReLU$ represents the relu activation function.

The data D^{in}, $D^{out} \in \mathbb{R}^{T\times N\times C}$ passed into TSIM are processed by an Inflow/Outflow Feature Extraction Module (FEM). The inter-information $H^{\alpha}_{inter} \in \mathbb{R}^{T\times N\times C}$ of traffic flow and the intra-information $H^{\alpha}_{intra} \in \mathbb{R}^{T\times N\times C}$ that starts from one of the flows and affects the other flow in grid division regions can be formulated as:

$$H^{\alpha}_{inter} = \Theta \star_{\mathcal{G}} (D^{\alpha}, A_g, A^{\alpha}_{inter}), \tag{8}$$

$$H^{\alpha}_{intra} = \Theta \star_{\mathcal{G}} (D^{\alpha}, A_g, A^{\alpha}_{intra}), \tag{9}$$

In summary, the processing of the TSIM input data D^{in} and D^{out} by FEM can be expressed as follows:

$$H^{in}_{intra}, H^{in}_{inter}, H^{out}_{intra}, H^{out}_{inter} = FEM(D^{in}, D^{out}), \tag{10}$$

where FEM represents a directional flow feature extraction module.

In order to more accurately simulate delay and diffusion of the traffic data, we also apply diffusion operations to external auxiliary information of the exterior region, which enables the capture of multi-step influences. As shown in Fig. 2, the output of the exterior region feature estimation module (ERFM) acts as the input of FEM. For external auxiliary inflow feature in exterior region, it generates the inter-information $\mathcal{H}^{in}_{inter} \in \mathbb{R}^{T\times N\times C}$ from inflow to outflow and the inflow intra-information $\mathcal{H}^{in}_{intra} \in \mathbb{R}^{T\times N\times C}$. In the same way, the external auxiliary outflow features in exterior region generates the inter-information $\mathcal{H}^{out}_{inter} \in \mathbb{R}^{T\times N\times C}$ from outflow to inflow and the outflow intra-information $\mathcal{H}^{out}_{intra} \in \mathbb{R}^{T\times N\times C}$. Such operations can be formulated as follows:

$$\mathcal{H}^{in}_{intra}, \mathcal{H}^{in}_{inter}, \mathcal{H}^{out}_{intra}, \mathcal{H}^{out}_{inter} = FEM(R^{in}, R^{out}). \tag{11}$$

D. Traffic Flow Feature Fusion Integrator

The feature fusion integrator is proposed to simulate the interweaving and complex characteristics of traffic flow and facilitate the subsequent stackable structure. A fusion method is introduced to integrate the four parts of information extracted from the traffic flows. The new traffic flow information

$\mathcal{N}^\alpha \in \mathbb{R}^{T\times N\times C}$ can be derived by H^β_{inter}, H^α_{intra}, $\mathcal{H}^\beta_{inter}$ and $\mathcal{H}^\alpha_{intra}$, which can be formulated as:

$$\begin{aligned}\mathcal{N}^\alpha = W_1^\alpha \odot H^\beta_{inter} + W_2^\alpha \odot H^\alpha_{intra} \\ + W_3^\alpha \odot \mathcal{H}^\beta_{inter} + W_4^\alpha \odot \mathcal{H}^\alpha_{intra},\end{aligned} \tag{12}$$

where $W_1^\alpha, W_2^\alpha, W_3^\alpha, W_4^\alpha$ are the weight matrices for traffic flow information fusion.

E. Traffic Flow Separation and Interaction Graph Convolution Module (TSIM)

A single TSIM implements an ERFM and two FEMs. The two FEMs fulfill distinct roles. One is to handle the intricate flow features of grid division regions, the other is to deal with the auxiliary flow features of the exterior region. To have better performance, a TSIM also has two structures besides the above components.

TCN: To address the trend characteristics between consecutive time slices of traffic flow, we incorporate Temporal Convolutional Networks (TCN) to handle them. Taking inflow as an example, assume the dilation rate is dr, then we can get the processed data P^{in}, $P^{out} \in \mathbb{R}^{(T-dr)\times N\times C}$ through TCN, which can be expressed as follows:

$$P^\alpha = \tanh(W^t_{i,1} \star \mathcal{N}^\alpha) \odot \sigma(W^t_{i,2} \star \mathcal{N}^\alpha), \tag{13}$$

where $\star$ is the dilated convolution operation, σ is the sigmoid function and $\tanh$ is the tanh function. $W^t_{i,1}$, $W^t_{i,2}$, $W^t_{o,1}$ and $W^t_{o,2}$ are weight matrices.

Skip Connections: To mitigate overfitting and enable more efficient learning in deep neural networks, residual connections are utilized. In a single TSIM, the input data D^{in} and D^{out} of TSIM and the data P^{in} and P^{out} processed by TCN are connected to form the residual data as the input of the next TSIM, which can be formulated as:

$$D^\alpha_{l+1} = BN(D^\alpha_l + \Phi(P^\alpha_l)), \tag{14}$$

where Φ represents the truncation operation to align dimension, BN represents the batch normalization, D^{in}_l is the input of the l^{th} TSIM, and D^{in}_{l+1} denotes the input of the next TSIM.

Furthermore, we establish skip connections between TSIMs. The output S^{in}, $S^{out} \in \mathbb{R}^{(T-dr)\times N\times C'}$ of a single TSIM are obtained by the combination of P^{in} and P^{out} through concatenation as follows:

$$S^\alpha = Concat(Conv(P^\alpha), Conv(P^\beta)). \tag{15}$$

where $Concat$ represents the concatenation operation.

F. Prediction Block

To achieve the traffic flow forecasting of the next time slice, we design the prediction block. The input of the prediction block is the integration results of data S in each TSIM. Convolution and ReLU activation function are used for dimension processing and feature extracting to obtain the target result X_t.

$$X^\alpha_t = Conv(ReLU(Conv(ReLU(\sum\nolimits_{l=1}^{L} S^\alpha_l)))). \tag{16}$$

where L represents the number of TSIMs. S^α_l denotes the S^α in the l^{th} TSIM.

V. Experiments

A. Experiment Setting

Dataset. Our experiments are conducted on four real-world traffic dataset: NYC-Taxi, NYC-Bike [9], CHI-Taxi and DC-Taxi [16]. The grid size of the datasets is 1km $\times$ 1km. For NYC datasets, we spilt the final 20 days for testing, the 8 days before for validation, and the rest (32 days) for training. For DC and CHI datasets, we spilt the final 2 months for testing, the 2 months before for validation, and the rest (8 months) for training.

Preprocessing and Evaluation Metric. Standardization is adopted at the data processing stage. The loss function of our model is MAE. RMSE and MAPE are used to evaluate the performance of the model. A threshold θ is set to filter out samples when testing the model [2]. Based on the distribution of the datasets, we set θ to 10 for the NYC datasets and to 2 for the other datasets.

Implementation. Our proposed model is implemented with PyTorch. The batch size is set to 16. The number of the input time slices T is set to 4 and the number of the TSIMs is set to 2. Our codes are available at https://github.com/hardworkingYang/STHGCN.

B. Baselines

We select representative baselines for experimental comparison: HA, ARIMA [18], ST-Resnet [8], DMVST-Net [17], Graph-WaveNet [15], STDN [9], DSAN [19], STGODE [20], DMSTGCN [21]. ST-SSL [22], ST-TIS [11]. These baselines includes traditional methods and comparable deep learning methods, which achieve outstanding even state-of-the-art performance.

C. Prediction Accuracy

The result of the experiment on NYC datasets is shown in table I. In the experiment, STHGCN achieves the best results compared with all baseline models, which shows that our model has a good performance on real traffic flow prediction scenarios. The result of the experiment on CHI and DC datasets is shown in table II. Our model also performs better on these two datasets from different geographical sources compared with four representative methods. The result indicates that our model has relatively good versatility that can play a role in both few-shot datasets and dense datasets from different urban areas.

Compared with the traditional naive prediction models like HA and ARIMA, STHGCN captures spatial features in traffic data, and thus has a better performance. Compared with ST-ResNet, DMVST-Net and STDN, our STHGCN uses adaptive adjacency matrices for graph diffusion convolution, which provides more flexibility in capturing potential connections between global regions. Compared with the graph convolution based models Graph Wavenet and DMSTGCN, STHGCN

[2] As in the methods adopted by Yao et al. [17] [9], the threshold setting is common in both industry and academia because low traffic volume are of little interest in real-world applications.

TABLE I
THE TRAFFIC FORECASTING PERFORMANCE COMPARISON OF STHGCN AND BASELINES ON TWO NYC TRAFFIC DATASETS.

Datasets	NYC-Taxi				NYC-Bike			
Metrics	RMSE		MAPE (%)		RMSE		MAPE (%)	
Methods	In	Out	In	Out	In	Out	In	Out
HA	33.83	43.82	21.14	23.18	11.93	12.49	27.06	27.82
ARIMA	27.25	36.53	20.91	22.21	11.25	11.53	25.79	26.35
ST-ResNet	22.23±0.30	27.23±0.25	21.40±0.99	21.79±0.47	8.18±0.26	8.30±0.32	25.88±1.22	26.32±1.41
Graph-WaveNet	20.47±0.12	25.83±0.13	16.58±0.07	16.34±0.08	8.19±0.05	9.20±0.05	21.93±0.15	22.66±0.15
STDN*	19.05±0.31	24.10±0.25	16.25±0.26	16.30±0.23	8.15±0.15	8.85±0.11	20.87±0.39	21.84±0.36
STGODE	23.10±0.02	26.56±0.04	17.90±0.08	17.84±0.01	8.36±0.03	9.28±0.03	22.21±0.17	23.48±0.06
DMSTGCN	18.79±0.14	24.39±0.16	15.73±0.24	15.40±0.34	8.32±0.12	8.89±0.09	22.21±0.22	22.75±0.38
ST-SSL	21.78±1.12	25.21±0.33	15.26±0.26	15.39±0.26	8.76±0.08	9.65±0.26	21.27±0.39	21.80±0.35
ST-TIS*	17.73±0.23	21.96±0.13	14.65±0.32	14.83±0.76	7.57±0.04	7.73±0.10	18.64±0.23	18.58±0.19
STHGCN	**16.94±0.09**	**21.88±0.08**	**13.30±0.02**	**13.19±0.03**	**6.82±0.05**	**7.39±0.03**	**17.50±0.14**	**17.92±0.09**

The results highlighted in bold are the best results in the experiment while the underlined results indicate the second best results.
The results of models marked with an asterisk (*) are cited from previous papers conducted under the same experimental conditions [9] [11].

TABLE II
PERFORMANCE COMPARISON OF STHGCN AND BASELINES ON CHI AND DC TAXI DATASETS..

Datasets	CHI-Taxi				DC-Taxi			
Metrics	RMSE		MAPE (%)		RMSE		MAPE (%)	
Methods	In	Out	In	Out	In	Out	In	Out
ST-ResNet	18.75	22.00	41.24	43.96	23.55	34.90	39.92	44.32
Graph-WaveNet	12.98	15.41	30.21	31.51	6.51	9.02	28.03	29.81
STGODE	15.10	15.60	42.19	42.78	8.85	10.15	38.55	35.26
DMSTGCN	12.26	15.14	29.72	30.88	6.27	8.64	33.68	32.83
STHGCN	**11.19**	**13.88**	**27.42**	**28.06**	**5.79**	**7.98**	**26.90**	**27.91**

takes into account the issue of time periodicity and the impact of exterior region traffic flow so that it can be aware of the specific location of the input time slice and achieve better prediction effect.

Compared with the advanced ordinary differential equation based models like STGODE, self-supervised models like ST-SSL and transformer based models like ST-TIS, STHGCN has better traffic flow forecasting performance and outperforms them. Collectively, STHGCN does not isolate traffic inflow and outflow and considers the inter-relationship and intra-relationship in traffic flows. In addition, STHGCN uses the prediction auxiliary information based on the traffic flow of the exterior region, broadening its scope of consideration and enhancing performance.

D. Ablation Study

Ablation study is conducted to analyze the effects of the components of STHGCN. We design three STHGCN variants as follows:

- **w/o-time-period:** removes the time period fusion block.
- **w/o-inout:** removes the Inflow/Outflow Feature Extraction Modules (FEM) in STHGCN.
- **w/o-special-region:** removes the Exterior Region Feature Estimation Modules (ERFM) in STHGCN.

The ablation study is tested on NYCTaxi and NYCBike datasets and the result is shown in Fig. 5. It shows that each module of our model is indispensable and can effectively improve the traffic forecasting effect. In *w/o-time-period*, due

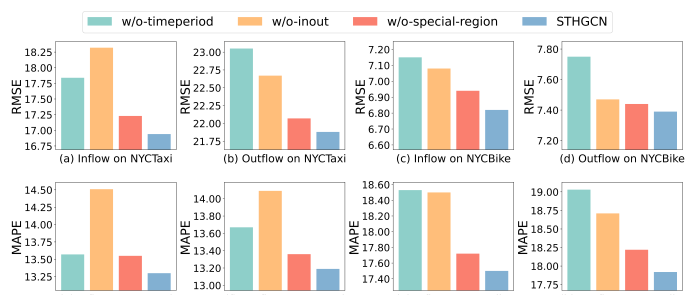


Fig. 5. The results of ablation study.

to the lack of time period fusion block, the temporal periodical sensitivity of the model weakens, which leads to the decline in the accuracy of traffic flow prediction. In *w/o-inout*, without the modeling of inter-relationship and intra-relationship in traffic flow, the adaptability of the model to complex traffic data decreases a lot. In *w/o-special-region*, lacking the supplementary features from the special exterior region, the model ignores some traffic flow volume fluctuation caused by cross-regional traffic transfers, and thus its performance is worse than that of STHGCN.

E. Time and Memory Efficiency Comparison

We select five representative baselines to conduct additional experiment together with STHGCN for time and memory efficiency comparison on NYC-Taxi dataset. The result of the experiment is shown in Table III. The inference time in the table is measured on the test set with the batch size set to 16. Compared with the lightweight optimized transformer based like model ST-TIS, STHGCN need less inference time and less GPU space while training. In comparison with the models that are also based on graph convolution like Graph-Wavenet and DMSTGCN, the inference time of STHGCN is at an intermediate level due to the additional modeling of inter-relationship and intra-relationship in traffic flows and the estimation part of exterior region flow. And STHGCN occupies less GPU space compared to the two models, thanks to the reduction of stackable layers without compromising the prediction performance. As opposed to the self-supervised and ODE models, STHGCN improves the prediction effect with less inference time and less GPU space occupied. To sum up, STHGCN attains superior predictive performance on the grid traffic dataset by taking less space and time resources.

TABLE III
TIME-MEMORY EFFICIENCY EXPERIMENT ON NYC-TAXI DATASET.

Method	Traning time (s)	Inference time (s)	Memory (MB)	Inflow RMSE	Outflow RMSE
Graph-WaveNet	**3.89**	**0.9**	2133	20.47	24.81
STGODE	5.87	1.42	<u>1929</u>	23.10	26.56
DMSTGCN	6.48	1.41	2795	18.79	24.39
ST-SSL	8.34	2.59	3347	21.78	25.21
ST-TIS	20.86	5.16	6815	<u>17.73</u>	<u>21.96</u>
STHGCN	<u>5.2</u>	<u>1.17</u>	**1833**	**16.94**	**21.88**

VI. CONCLUSION

In this work, we propose a novel Spatial-Temporal Flow Holistic Interaction Graph Convolution Network model for traffic flow prediction. It utilizes a time period fusion method for time periodic capture. It takes into consideration the influence of the exterior region and utilizes an ERFM to estimate its impact on the grid division regions as auxiliary information. It uses TSIMs to model intricate traffic flow information from the perspective of graph convolution. And STHGCN employs a traffic flow feature fusion integrator to accurately simulate the interweaving relationships among regional traffic flows. With layer-by-layer separation and fusion of inflow and outflow information, STHGCN achieves better traffic flow prediction performance. Experimental results on four real-world datasets show that STHGCN outperforms the compared baselines. In the future, we plan to investigate the distinct and dynamic contributions of two types of traffic flow and two kinds of traffic regions to the final prediction for further accuracy enhancement.

REFERENCES

[1] D. E. Rumelhart, G. E. Hinton, and R. J. Williams, "Learning internal representations by error propagation," California Univ San Diego La Jolla Inst for Cognitive Science, Tech. Rep., 1985.

[2] S. Bai, J. Z. Kolter, and V. Koltun, "An empirical evaluation of generic convolutional and recurrent networks for sequence modeling," *arXiv preprint arXiv:1803.01271*, 2018.

[3] B. Yu, H. Yin, and Z. Zhu, "Spatio-temporal graph convolutional networks: A deep learning framework for traffic forecasting," *27th International Joint Conference on Artificial Intelligence*, p. 3634–3640, 2018.

[4] A. v. d. Oord, S. Dieleman, H. Zen, K. Simonyan, O. Vinyals, A. Graves, N. Kalchbrenner, A. Senior, and K. Kavukcuoglu, "Wavenet: A generative model for raw audio," *9th ISCA WKSH on Speech Synthesis WKSH*, 2016.

[5] S. Guo, Y. Lin, H. Wan, X. Li, and G. Cong, "Learning dynamics and heterogeneity of spatial-temporal graph data for traffic forecasting," *IEEE Transactions on Knowledge and Data Engineering*, vol. 34, no. 11, pp. 5415–5428, 2021.

[6] Y. Li, R. Yu, C. Shahabi, and Y. Liu, "Diffusion convolutional recurrent neural network: Data-driven traffic forecasting," *International Conference on Learning Representations (ICLR)*, 2018.

[7] J. Jiang, C. Han, W. X. Zhao, and J. Wang, "Pdformer: Propagation delay-aware dynamic long-range transformer for traffic flow prediction," *Proceedings of the AAAI conference on artificial intelligence*, 2023.

[8] J. Zhang, Y. Zheng, and D. Qi, "Deep spatio-temporal residual networks for citywide crowd flows prediction," in *Thirty-first AAAI conference on artificial intelligence*, 2017.

[9] H. Yao, X. Tang, H. Wei, G. Zheng, and Z. Li, "Revisiting spatial-temporal similarity: A deep learning framework for traffic prediction," in *Proceedings of the AAAI conference on artificial intelligence*, vol. 33, no. 01, 2019, pp. 5668–5675.

[10] X. Zhang, C. Huang, Y. Xu, L. Xia, P. Dai, L. Bo, J. Zhang, and Y. Zheng, "Traffic flow forecasting with spatial-temporal graph diffusion network," in *Proceedings of the AAAI conference on artificial intelligence*, vol. 35, no. 17, 2021, pp. 15 008–15 015.

[11] G. Li, S. Zhong, X. Deng, L. Xiang, S.-H. G. Chan, R. Li, Y. Liu, M. Zhang, C.-C. Hung, and W.-C. Peng, "A lightweight and accurate spatial-temporal transformer for traffic forecasting," *IEEE Transactions on Knowledge and Data Engineering*, 2022.

[12] Y. Zhang, B. Wang, Z. Shan, Z. Zhou, and Y. Wang, "Cmt-net: A mutual transition aware framework for taxicab pick-ups and drop-offs co-prediction," in *Proceedings of the Fifteenth ACM International Conference on Web Search and Data Mining*, 2022, pp. 1406–1414.

[13] Y. Zhao, Y. Lin, Y. Zhang, H. Wen, Y. Liu, H. Wu, Z. Wu, S. Zhang, and H. Wan, "Traffic inflow and outflow forecasting by modeling intra- and inter-relationship between flows," *IEEE Transactions on Intelligent Transportation Systems*, vol. 23, no. 11, pp. 20 202–20 216, 2022.

[14] J. Zhang, Y. Zheng, J. Sun, and D. Qi, "Flow prediction in spatio-temporal networks based on multitask deep learning," *IEEE Transactions on Knowledge and Data Engineering*, vol. 32, no. 3, pp. 468–478, 2019.

[15] Z. Wu, S. Pan, G. Long, J. Jiang, and C. Zhang, "Graph wavenet for deep spatial-temporal graph modeling," *Proceedings of the 28th International Joint Conference on Artificial Intelligence*, p. 1907–1913, 2019.

[16] Y. Jin, K. Chen, and Q. Yang, "Selective cross-city transfer learning for traffic prediction via source city region re-weighting," in *Proceedings of the 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, 2022, pp. 731–741.

[17] H. Yao, F. Wu, J. Ke, X. Tang, Y. Jia, S. Lu, P. Gong, J. Ye, and Z. Li, "Deep multi-view spatial-temporal network for taxi demand prediction," in *Proceedings of the AAAI conference on artificial intelligence*, 2018.

[18] S. V. Kumar and L. Vanajakshi, "Short-term traffic flow prediction using seasonal arima model with limited input data," *European Transport Research Review*, vol. 7, no. 3, pp. 1–9, 2015.

[19] H. Lin, R. Bai, W. Jia, X. Yang, and Y. You, "Preserving dynamic attention for long-term spatial-temporal prediction," in *Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, 2020, pp. 36–46.

[20] Z. Fang, Q. Long, G. Song, and K. Xie, "Spatial-temporal graph ode networks for traffic flow forecasting," in *Proceedings of the 27th ACM SIGKDD conference on knowledge discovery & data mining*, 2021, pp. 364–373.

[21] L. Han, B. Du, L. Sun, Y. Fu, Y. Lv, and H. Xiong, "Dynamic and multi-faceted spatio-temporal deep learning for traffic speed forecasting," in *Proceedings of the 27th ACM SIGKDD conference on knowledge discovery & data mining*, 2021, pp. 547–555.

[22] J. Ji, J. Wang, C. Huang, J. Wu, B. Xu, Z. Wu, J. Zhang, and Y. Zheng, "Spatio-temporal self-supervised learning for traffic flow prediction," *Proceedings of the AAAI conference on artificial intelligence*, 2023.