

Short-time Traffic Flow Prediction Based on High-order Graph Convolutional Networks

Qingrong Wang, Rong He, Changfeng Zhu, Huihui Rao

Abstract—Forecasting traffic patterns is an essential component of intelligent transport systems. It helps forecast traffic congestion, improve traffic management, and make roads safer. However, most existing studies focus on the direct traffic data dependencies between road segments but lack sufficient attention to the implied higher-order dependencies within the road network. Furthermore, although existing models attempt to capture spatio-temporal relationships by integrating different components, they are still inadequate in effectively modeling the interaction between spatio-temporal relationships. We propose a higher-order graph convolutional network model (STK-HGCN) to address the above issues. Firstly, the model extracts and fuses temporal low-order and high-order information using temporal convolutional layers; in terms of spatial aspects, we capture low-order spatial information by constructing geographic and functional graphs and introduce hypergraphs to create the implied high-order information between road networks, which are then convolved to obtain the high-order spatio-temporal dependencies. Secondly, we construct a spatial-temporal interaction module with a "sandwich" structure to capture spatial-temporal interaction features effectively. Finally, we performed comparative experiments on the METR-LA and PEMS-BAY datasets, and the results indicate that, compared to the optimal baseline model, the STK-HGCN model improves MAE and RMSE by 0.08, 0.13, 0.15, and 0.37, respectively, while reducing MAPE by 0.25% and 0.18%, respectively.

Keywords—hypergraph convolution, spatio-temporal interaction, spatio-temporal convolution, traffic flow prediction

I. INTRODUCTION

DRIVEN by the strong trend of urbanization, traffic flow continues to grow and expand, and the resulting urban traffic congestion has become a significant challenge for today's society. Therefore, the management and optimization of traffic flow are crucial in traffic planning and road construction. In particular, short-term traffic flow prediction enables the management to accurately grasp the time and

location of the occurrence of traffic congestion, formulate targeted traffic diversion programs to prevent the proliferation and deterioration of congestion, and simultaneously provide real-time road condition information to the public, helping them choose appropriate travel routes to avoid congested areas [1].

Currently, the development of traffic flow prediction tasks can be summarised in four phases: prediction models based on statistical analysis, machine learning models, deep learning models, and hybrid model phases [2]. Early statistical methods [3] have certain shortcomings due to their limitations, making it challenging to effectively address non-routine and unexpected traffic conditions. In contrast, machine learning-based methods [4]-[7] are widely used because they better represent nonlinear data. However, although machine learning methods have certain advantages in compensating for the nonlinearities and irregularities of statistical-based methods, they usually can only extract shallow features, and it is difficult to extract the deeper characteristics required in the traffic flow, and thus perform poorly in coping with tasks such as traffic flow prediction.

The swift progress in deep learning presents novel research methods to tackle this challenge. For example, using Long Short-Term Memory Networks [8],[9] (LSTMs), Recurrent Neural Networks [10],[11] (RNNs), and Gated Recurrent Units [12],[13] (GRUs), the temporal features in traffic data can be effectively learned. These methods can accurately predict traffic flow by effectively capturing long-term dependencies and temporal evolution features in sequence data. Researchers have implemented Convolutional Neural Networks [14] (CNNs) to effectively manage the spatial characteristics of traffic information; where they chose to store the traffic data in the form of a two-dimensional matrix and, at the same time, automatically captured the spatial features with the help of the translation invariant qualities of CNNs. Although this approach has achieved some success in capturing spatiotemporal features, it is difficult to apply to traffic networks with irregular spatial distributions. Moreover, single-model approaches are unable to account for both the spatial and temporal features of traffic patterns.

With a deeper understanding of traffic networks' intricacies and changing patterns, spatio-temporal map data has become an essential form of data representation in traffic prediction. Therefore, related scholars have established various combinatorial models on spatio-temporal graph data. By integrating Graph Convolutional Network (GCN) and Temporal Convolutional Network (TCN), they suggested several hybrid methods for predicting traffic flow using GCN. These methods preserve the traffic network's graph structure while effectively capturing the time-dependent patterns of traffic flow. For example, literature [15] introduced the

Manuscript received May 9, 2024; revised September 15, 2024. This work was supported in part by the National Natural Science Foundation of China (No. 72161024), "Double-First Class" Major Research Programs, Educational Department of Gansu Province (No. GSSYLXM-04).

Qingrong Wang is a professor at School of Electronic and Information Engineering, Lanzhou Jiaotong University, Lanzhou 730070, China. (E-mail: 329046272@qq.com).

Rong He is a postgraduate student at School of Electronic and Information Engineering, Lanzhou Jiaotong University, Lanzhou 730070, China. (Corresponding author, E-mail: 3391355291@qq.com).

Changfeng Zhu is a professor at School of Traffic and Transportation, Lanzhou Jiaotong University, Lanzhou 730070, China. (E-mail: cfzhu003@163.com).

Huihui Rao is a postgraduate student at School of Electronic and Information Engineering, Lanzhou Jiaotong University, Lanzhou 730070, China. (E-mail: rhhyhlh@163.com).

Diffusion Convolutional Recurrent Neural Network (DCRNN) model, which employs a random wandering strategy to capture spatial correlation effectively and also replaces the traditional matrix multiplication operation in the GRU with a diffusion convolution operation, which achieves the effective capture of temporal features; literature [16] proposes the Graph WaveNet (GWN) method, which integrates the idea of diffusion convolution in extracting spatial features and introduces the adaptive connection matrix, which effectively overcomes the limitations of the fixed topology in the extraction of spatial features in the traditional method; In literature [17], a Spatio-Temporal Graph Convolutional Network (STGCN) was introduced, utilizing a series of spatiotemporal blocks to capture the spatiotemporal correlations within graph data. A composite architecture containing gated temporal convolution and spatial graph convolution is designed in each spatiotemporal block to accurately characterize temporal dynamics and spatial features in traffic flow prediction tasks, thus effectively improving prediction accuracy. Although the above study captured the spatiotemporal features well by combining the models, the spatial and temporal components in the models are independent of each other, which makes the models limited in modeling the interactions between space-time dimensions, and unable to capture the complex correlations between them adequately. Therefore, modeling the complex spatio-temporal interaction features remains challenging for transport networks.

In addition, over time, the traffic flow in the same region shows continuous changes. Also, it affects the traffic flow in the neighboring areas, making the interaction between regions non-linear [18]. At the same time, regions are also affected by complex spatial relationships, including physical spatial relationships (neighboring regions in the traffic network) and semantic spatial relationships (similar functional areas in different regions). However, most existing research relies on the traditional graph structure to represent direct traffic data dependencies between two regions. Still, this approach fails to fully capture the potential higher-order dependencies within the road network.

Currently, the introduction of hypergraphs provides an effective solution to this problem, with the core advantage of encoding multiple feature nodes simultaneously and containing higher-order node and edge space relationships. Therefore, more and more researchers have applied hypergraphs to traffic prediction. Literature [19] proposed a Hypergraph Neural Network (HGNN) framework that efficiently captures complex data correlations among multiple nodes with the help of hypergraph structure. Literature [20] has made significant strides in forecasting traffic flow by employing a hypergraph framework to represent local spatio-temporal relationships within traffic data. This approach leverages hypergraph convolution along with a sliding-window information fusion mechanism to effectively identify and harness spatio-temporal characteristics. Considering that real-world graph-structured data fully represent the spatial relationships of road networks, literature [21] proposes a Directed Hypergraph Attention Network (DHAT), which suggests a directed hypergraph convolution method based on directed hypergraphs to make full use of the traffic sequences complex spatial relationships

among them better to get the higher-order spatial relationships of topological structures. However, the existing hypergraph neural network models are often limited by relying only on the initial hypergraph structure and failing to fully consider the dynamic changes that may occur during the training process, which limits their performance and application scope to some degree.

To address these challenges, we introduce a novel temporal prediction model using a higher-order graph convolutional network (STK-HGCN). Specifically, in the temporal dimension, the model employs higher-order differential convolution and improved gated temporal convolution using the attention mechanism to extract temporal features. For the spatial dimension, STK-HGCN models spatial correlation into three aspects: geospatial correlation, functional spatial correlation, and higher-order spatial correlation; based on this, STK-HGCN further constructs geographic graphs, functional graphs and higher-order hypergraphs and employs a two-layer spatial convolutional structure to obtain the higher-order spatial information among nodes. At the same time, STK-HGCN uses a Kalman filter to correct the convolution output, from which the essential laws in traffic data are mined, aiming to reduce the bias caused by too many model parameters. Finally, the spatial and temporal convolution layers are constructed as a "sandwich" structure to form a spatio-temporal interaction module to obtain the interaction information of the node data.

II. ATHEORETICAL FOUNDATION

A. Traditional and hypergraphs

Traditional Graph

A traffic road network can be described as a traditional graph $G_G=[V_G, E_G]$, where $V_G=[v_1, v_2, \dots, v_i, v_j, \dots, v_n]$ represents the nodes within the road network, and $E_G=[e_1, e_2, \dots, e_m]$ represents the set of edges, where m denotes the total number of edges. The adjacency matrix A represents the connections among nodes in Fig. 1. In the adjacency matrix, if there is a connection between nodes, the corresponding element is assigned a value of 1. Conversely, if there is no connection, the element is assigned a value of 0. Thus, we can define its adjacency matrix as follows:

$$A_{ij} = \begin{cases} 1, & e(v_i, v_j) \in E_G \\ 0, & e(v_i, v_j) \notin E_G \end{cases} \quad (1)$$

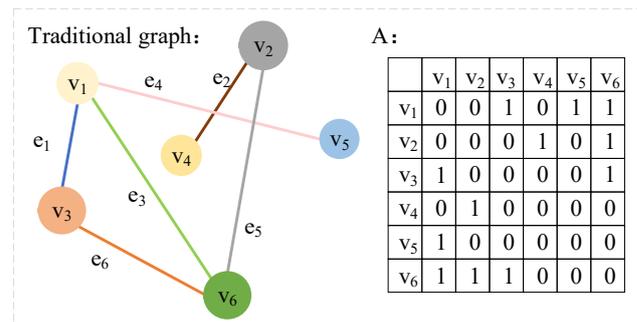


Fig. 1. Traditional graph

In transportation networks, traditional graphs are usually used to describe the road system structure, where nodes denote intersections or traffic hubs and edges denote road connection relationships. However, the road network structure cannot adequately cover the complex relationships of various traffic modes and flows in complex transportation networks. Therefore, the introduction of hypergraphs becomes crucial. Hypergraphs allow edges to connect two or more nodes and can more flexibly represent the associations between various transportation modes, such as bus routes, subway routes, and other transportation services. Through hypergraphs, routes and stations of different transportation modes can be linked to more accurately simulate the actual operation of urban transportation networks.

Hypergraph

A hypergraph can be depicted as $G_H=[V_H, E_H]$, where $V_H=[v_1, v_2, \dots, v_i, \dots, v_n]$ signifies the nodes within the road network and $E_H=[e_1, e_2, \dots, e_m]$ illustrates the higher-order hyperedges. The hypergraph's structure is illustrated in Fig. 2 and is described by the association matrix H . An element $h(v_i, e_j)$ of the association matrix H is 1 if node i is contained in hyperedge j ; otherwise, it is 0. Additionally, each node and hyperedge is associated with a degree matrix. D_v refers to the degree matrix corresponding to the nodes, whereas D_e denotes the degree matrix associated with the hyperedges within the network. In these matrices, $d(v)$ and $d(e)$ denote the degree matrix elements for nodes and hyperedges, respectively:

$$h(v_i, e_j) = \begin{cases} 1, & v_i \in e_j \\ 0, & v_i \notin e_j \end{cases} \quad (2)$$

$$d(v) = \sum_{e \in E_H} \omega(e) h(v, e) \quad (3)$$

$$d(e) = \sum_{v \in V_H} h(v, e) \quad (4)$$

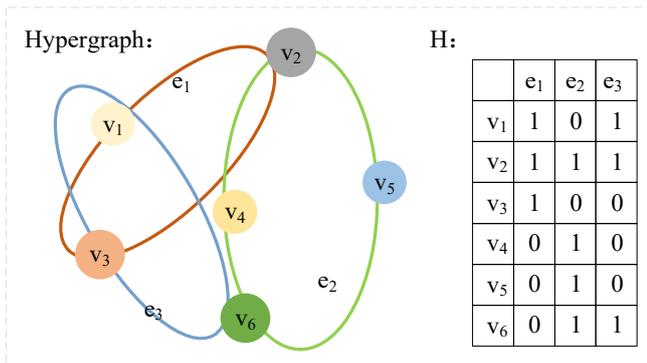


Fig. 2. Hypergraph

One must model the data to perform an effective convolution operation on the hypergraph, like in GCN. Most existing models use a distance-based approach to construct the hypergraph. Still, the traffic data contains a large amount of noise and outliers, resulting in inaccurate numbers and distances of regional nodes. Meanwhile, traffic time series data also contains dynamic relational graph structures, thus limiting the accuracy of data modeling and affecting the performance of hypergraph learning. The Graph Attention Network (GAT) excels in dynamic learning and effectively captures first-order relationships within the data. Compared with existing methods, GAT can be integrated into the

network framework for unified training, which enables fine-grained optimization of dynamic hypergraph relations. Therefore, this study employs the GAT to create hyperedges and, in turn, develop hypergraph correlation matrices derived from traffic time series data.

Firstly, the traffic road network data embedded values X are entered into GAT, and the relationships between nodes are modeled using the attention mechanism to generate updated node representations. This approach significantly enhances the model's capacity to capture and interpret temporal changes in graph data, thereby enhancing its overall predictive accuracy. After training, the node representations generated by GAT are used as embeddings $M=[m_1, \dots, m_N]$, which contain the original node features and temporal information.

$$m_i = \text{GAT}(X) = \sigma \left(\sum_{j \in N(i)} \alpha_{ij} W X \right) \quad (5)$$

$$\alpha_{ij} = \frac{\exp(\text{LeakyReLU}(a_{ij}[W_i z_i \oplus W_j z_j]))}{\sum_{k \in N(i)} \exp(\text{LeakyReLU}(a_{ik}[W_i z_i \oplus W_k z_k]))} \quad (6)$$

Among them, $N(i)$ is the neighbor set of node i , and σ is the sigmoid function. α_{ij} represents the importance weight from node i to j . W_i , W_j , W_k , and a_{ij} , a_{ik} are learnable parameters.

Next, the node embedding M is mapped into an association matrix H_D via a fully connected layer (FC) and a Softmax function to obtain a representation of the relationship between the nodes and the hyperedges:

$$H_D = \text{FC}(M) = \text{softmax}(\tanh(W_f M + b_f)) \quad (7)$$

Where H_D denotes the dynamic hypergraph embedded in the traffic time series data and serves to characterize the higher-order relationships between regional nodes of the traffic road network. W_f denotes the weight matrix and b_f represents the deviation vector.

B. Problem Analysis

Spatial and Temporal Characterization of Traffic Flow

Traffic flow exhibits significant spatial and temporal attributes. In the time dimension, traffic flow has apparent time dependence, i.e., the traffic flow at any given moment significantly influences the flow in subsequent moments. At the same time, in space, sensors are usually distributed in different road sections for real-time monitoring of parameters such as traffic flow and speed. Due to the continuity of traffic flow and the influence between neighboring road sections, the data collected by these sensors tend to be similar. For example, when a road section experiences traffic slowdown or congestion, its neighboring upstream and downstream sections will also be affected. As shown in Fig. 3, this figure illustrates the speed change profiles recorded by three neighboring sensors (numbered 1, 2, and 3) over a day. The figure shows that the speed change curves of these three sensors show a high degree of similarity during peak traffic hours, reflecting that synchronous fluctuations characterize the traffic flow during peak hours. Consequently, it is crucial to capture spatial and temporal patterns for precise traffic forecasting. This necessitates comprehensive analysis and exploration of the spatio-temporal characteristics inherent in traffic flow data.

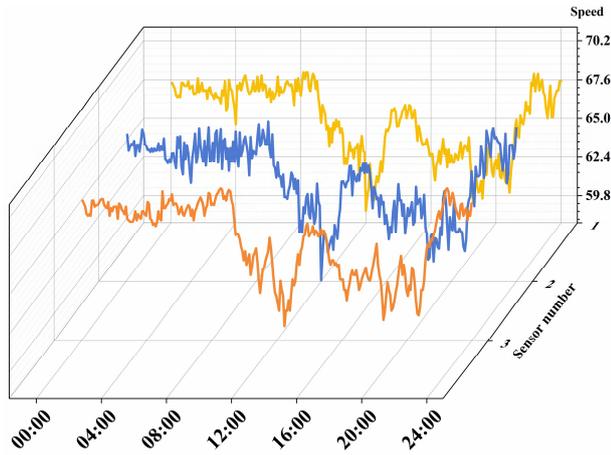


Fig. 3. Traffic speed distribution of adjacent sensors

Higher-order Spatio-temporal Interactions

A key challenge in traffic flow prediction lies in accurately capturing the interactive effects of its complex spatio-temporal features. In a traffic system, the traffic flow at each regional node is shaped by its current conditions and the future flow from adjacent nodes. Again, in Fig. 4(a), when traffic congestion occurs in the commercial area at time step t_1 , its traffic state directly influences the state at time step t_2 . In the busy areas of the city, it even influences the state at time step t_3 , which indicates a higher-order temporal correlation between the regions across time dimensions; in the spatial dimension, as shown in Figs. 4(b) and (c), there are apparent geospatial correlations among the commercial, residential, school, and hospital-prominent geospatial associations. Meanwhile, significant semantic correlations are also shown between areas with similar functions in different regions. However, the difficulty in predicting traffic flow stems from the distinct nature of temporal and spatial dependencies, along with their interactions over time and space. For example, when traffic congestion occurs in a commercial area at the current time step, its traffic conditions may directly impact the entire commercial area and its surrounding neighborhood at the same time step. They may also affect the traffic conditions in the neighborhood and similar functional regions at the next step.

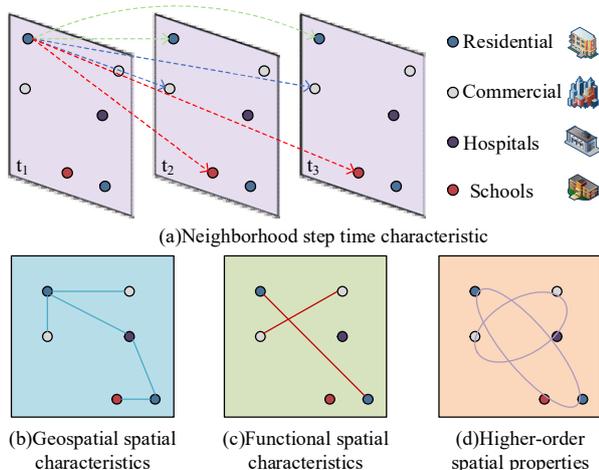


Fig. 4. Complex spatial and temporal characteristics of traffic data

In addition, we should not ignore the correlation between more than two regional nodes in the transport network. As shown in Fig. 4(d), the business district and its surrounding regional nodes within a particular spatial range form a

hyperedge, which considers the node relationships across multiple spatial latitudes, thus effectively representing the common attributes and associations among the regional nodes of the road network. Therefore, traffic flow prediction needs to integrate information in both space-time dimensions and fully use the dynamic interaction features and higher-order information of spatio-temporal data to enhance further the model's capacity to capture higher-order spatial correlations.

In summary, this paper integrates the spatiotemporal data observed on the traffic road network G over t periods into a matrix X , $X = (X^{(t-T+1)}, X^{(t-T+2)}, \dots, X^{(t)}) \in \mathbb{R}^{T \times N \times F}$, where N denotes the overall count of nodes within the traffic road network, and F represents each node's feature dimensions. $Y = (X^{(t+1)}, X^{(t+2)}, \dots, X^{(t+P)}) \in \mathbb{R}^{P \times N \times F}$ represents the traffic data for P 's future time steps. Thus, we can describe traffic flow prediction as predicting traffic conditions at each node for several future time steps based on traffic observations at each node for several historical time steps. The function $f()$ can represent this relationship.

$$Y = f(X, G) \quad (8)$$

III. MODEL CONSTRUCTION

Figure 5 illustrates the STK-HGCN model's structure, consisting of three key components: the input module, the spatio-temporal interaction module, and the output module. Firstly, the input module effectively represents information across both temporal and spatial dimensions. In addition, during the training process, we use historical data as the temporal convolution layer's input, ensuring that the model can fully retain the original temporal feature information. In this manner, we can effectively mine the temporal features in traffic data, ensuring the model receives accurate information input in spatial and temporal dimensions.

Second, the spatio-temporal interaction module adopts a unique "sandwich" structure to capture the spatio-temporal interaction features affecting traffic flow. It consists of multiple ST-Blocks, the core components of which are temporal and spatial convolutional layers. The temporal convolutional layer (GH-TCN) consists of two main components: Higher Order Differential Convolution (HDTCN) and Gated Temporal Convolution Improved Using the Attention Mechanism (Gate-ATCN). Specifically, Gate-ATCN provides basic temporal features for the model by capturing low-order information in the temporal dimension. HDTCN provides deeper temporal features for the model by exploiting temporal differences to dig deeper and present complex higher-order temporal correlations between regions across the temporal dimension.

The spatial convolution layer (KH-GCN), combined with the input adjacency matrix and temporal feature information, has constructed a spatial graph convolution method (H-GCN). H-GCN consists of graph convolution (GCN) and hypergraph convolution (HGCN), in which the geospatial correlation and functional spatial correlation of geographic and functional graphs are captured using GCN, and the dynamic higher-order features of hypergraphs in the traffic road network are captured using HGCN, to capture the higher-order spatial information in the historical data

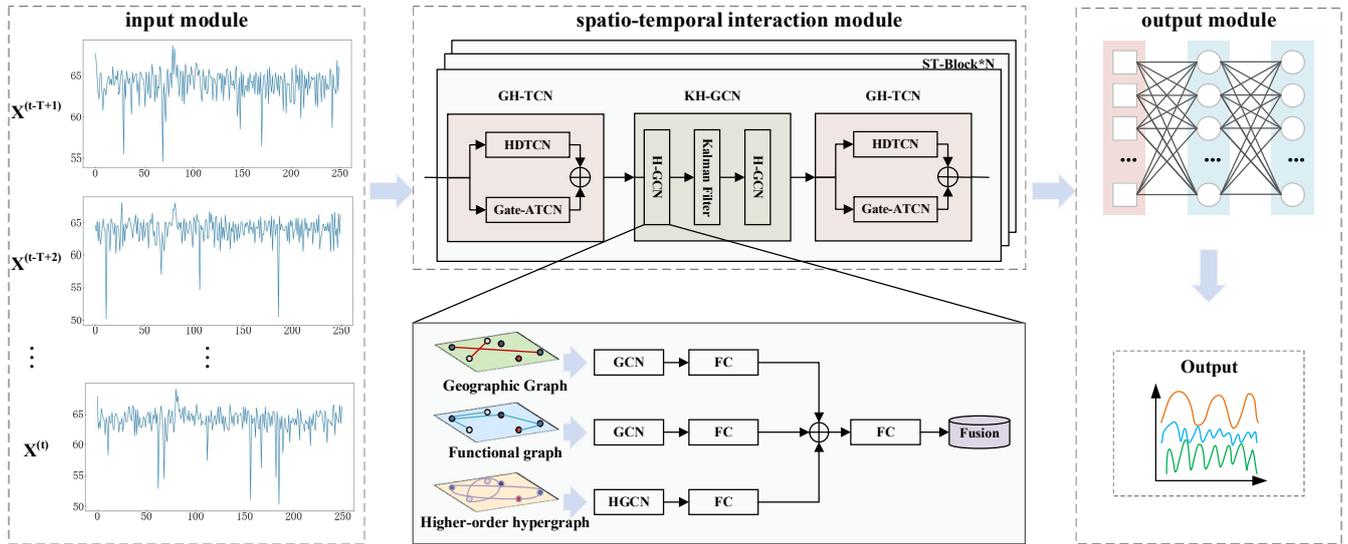


Fig. 5. STK-HGCN model

accurately. To mitigate the model's bias due to too many parameters, the model introduces a Kalman filter to correct the output of the spatial convolution. The adjusted feature data is then passed into the subsequent temporal convolution layer for further processing. In this process, the spatial convolution layer plays the role of continuing from the preceding and introducing the following to the next between the temporal convolution layer, which enables the feature information to propagate between the two layers, thus deeply exploring the correlation of traffic data in the spatial and temporal dimensions and their interactive features. Finally, a two-layer FC is utilized to generate the prediction results, thereby enhancing the precision and dependability of the results.

A. Input Module

Along the temporal dimension, the historical information is directly used as input into the temporal convolution layer to ensure that the model fully captures and retains the original temporal features throughout the training process. In terms of the spatial dimension, the adjacency matrix of the road network's graph structure typically represents the geospatial distance relationships among the nodes within the network. Furthermore, the trending similarity of different nodes in terms of traffic flow indicates that they may have similar characteristics regarding location and function. Also, higher-order spatial correlations exist between multiple regional nodes, which is indispensable in traffic prediction. Given this, this paper focuses on processing spatial characteristic inputs to capture spatio-temporal characteristics more accurately.

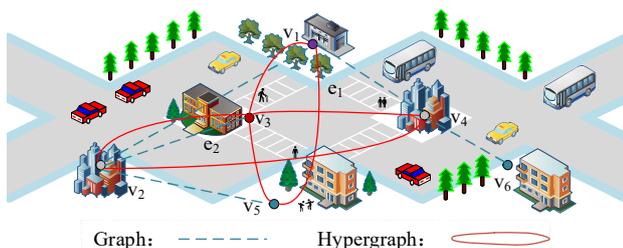


Fig. 6. Traditional and hypergraphic graphs in road networks

As shown in Fig. 6, in the context of transport road networks, the traditional graph modeling approach focuses

only on the direct relationship between two-by-two nodes, ignoring the complex, global connectivity information that may exist between multiple regions. However, in the framework of hypergraphs, edge e_2 can connect nodes v_2 , v_3 , v_4 , and v_5 simultaneously, demonstrating its powerful ability to capture higher-order relationships among multiple nodes. This paper constructs geographic and functional graphs from various dimensions to capture the low-order information and portray this complex network more comprehensively and in-depth. It introduces a higher-order hypergraph to reveal the higher-order connectivity among the nodes. Typically, the adjacency matrix of the geospatial graph is derived from the Euclidean distance between nodes, as shown below:

$$A_{ij}^g = \begin{cases} \exp(-\frac{d_{ij}^2}{\sigma^2}), & \text{if } \exp(-\frac{d_{ij}^2}{\sigma^2}) \geq \varepsilon \\ 0, & \text{otherwise} \end{cases} \quad (9)$$

Where hyperparameters ε and σ^2 are responsible for regulating the sparsity of the matrix, while d_{ij} represents the distance separating nodes i and j .

To capture the functional relationships between regional nodes in the traffic data, we employ the Dynamic Time Warping (DTW) algorithm to construct the adjacency matrix of the functional graph, which serves as a representation of these connections. which is:

$$A_{ij}^f = \begin{cases} 1, & D(X^i, Y^j) > \varepsilon \\ 0, & \text{otherwise} \end{cases} \quad (10)$$

$$D(i, j) = \text{dist}(x_i, y_j) + \min(D(i-1, j), D(i, j-1), D(i-1, j-1)) \quad (11)$$

Where X^i is the time series of the i th node, and $D(i, j)$ denotes the shortest distance between $X = (x_1, x_2, \dots, x_m)$ and $Y = (y_1, y_2, \dots, y_n)$ given two sub-time series. $\text{dist}(x_i, y_j)$ is the absolute distance of x_i and y_j , and ε is a hyperparameter used to regulate the sparsity of the adjacency matrix.

The high-order information graph fully considers the non-direct connection relationship between regional nodes in the construction process to more accurately capture the higher-order correlation between more than two regional nodes. This design enables the high-order information graph

to explore more complex traffic flow patterns and predict dynamic changes more accurately. Therefore, based on the association matrix H_D of the hypergraph structure in Chapter II, we obtain the normalized hypergraph adjacency matrix \hat{Z} in this paper. Its specific formula is as follows:

$$A^{H_D} = HWH^T - Dv \quad (12)$$

$$\tilde{A}^{H_D} = A^{H_D} + I_n \quad (13)$$

$$\hat{Z} = \tilde{D}^{-\frac{1}{2}} \tilde{A}^{H_D} \tilde{D}^{-\frac{1}{2}} \quad (14)$$

Where I_n the identity matrix of order N and A^{H_D} represents the neighborhood of the hypergraph. W serves as the diagonal matrix, which contains the weight information of the hyperedge. The Dv matrix then reflects the degree of each vertex in the hypergraph. The \tilde{A}^{H_D} matrix is an adjacency matrix of the hypergraph, enhanced with a self-loop. This augmentation enables the simultaneous incorporation of features from both the current node and its neighboring nodes within the traffic network, thereby facilitating effective information aggregation.

B. Spatio-Temporal Interaction Module

The spatio-temporal interaction module employs a sequence of ST-Blocks, each incorporating a unique 'sandwich' architecture comprised of two GH-TCN layers and one KH-GCN layer. The KH-GCN layer connects the two GH-TCN layers, enabling the propagation of feature information between them and enhancing the interaction and fusion of information across layers. Simultaneously, the model employs a fully convolutional architecture to extract and integrate features at various levels through successive convolutional layers. This approach enables the model to grasp a broader range of contextual information, thus enhancing its capacity to uncover spatial and temporal interactions.

Time Convolution Layer

(1) Gated time convolution

TCNs are extensively employed in time series prediction tasks and can output sequences of arbitrary lengths to the same size. To address the issue of a restricted field of view inherent in the original TCN method, researchers [22]-[24] implemented Dilated Causal Convolution (DCC) to improve the TCN's capacity for capturing long-range dependencies in time series data. TCN with DCC can effectively capture the temporal dependence between traffic states of neighboring time steps and is especially good at capturing the continuity and change trends in the short term.

As shown in Fig. 7, each layer of the TCN with DCC contains a one-dimensional convolution block (1-DConv) where the dilation factor d grows exponentially. Each residual block consists of a DCC layer, a weight normalization layer (WeightNorm), a Dropout layer, and a ReLU activation function layer. The residual blocks are connected through residual connections. Despite its excellent performance in several aspects, it still faces some challenges. One of the most prominent issues is the lack of ability to dynamically adjust the importance of sequences, which may result in the model not performing flexibly and accurately enough when dealing with complex and variable time-series data.

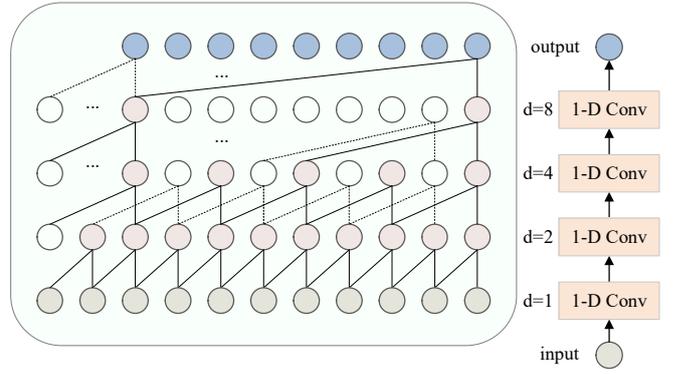


Fig. 7. Structure of TCN with DCC

Therefore, this paper further improves it to compensate for the deficiency of TCN with DCC in dynamically adjusting the importance of sequence data. By introducing an attention mechanism in the residual connection, the causal convolutional layer's expansion can dynamically change the significance of the input data. In Fig. 8, the expanded causal convolutional network (ATCN) with the added attention mechanism adds an Attention Layer before the expanded causal convolutional layer in each residual structure. The Attention Layer's main function is to emphasize key information in the input data, highlighting the essential differences between data points. After the Attention Layer's weighting process, the sequence data's vital information is augmented and passed to the Expansion Causal Convolution Layer for additional processing. By employing this approach, the model can better distinguish the importance of sequence data after each hidden layer processing, which enhances the ability of the Expanded Causal Convolutional Layer to learn the temporal features. At the same time, it can efficiently learn the complex feature relationships between the sequences and pass this information to the next layer for further processing.

Subsequently, we introduce the gating mechanism based on ATCN, leading to the construction of a gated time convolution based on the attention mechanism (Gate-ATCN). This approach effectively regulates information processing within every layer of the temporal convolution. The Gate ATCN in Fig. 9 consists of two ATCN modules. One module employs DCC and hyperbolic tangent activation functions, which perform dilated activation. The other module employs DCC and sigmoid activation functions, which regulate the rate of information transfer. Subsequently, by integrating the outputs of the two modules, we get the low-order temporal characteristics of regional nodes within the road network. Let the input of layer l be $X^{(l)}$, $l \in \{0, \dots, L\}$, and L represents the overall count of layers within the spatiotemporal block, which is computed as follows:

$$X^{(0)} = \text{1DConvs}(X) \quad (15)$$

$$X^{(l)} = \frac{\exp(W_a X^{(l)} + b_a)}{\sum_{i=1}^l \exp(W_a X^{(i)} + b_a)} X^{(l)} \quad (16)$$

$$X_a^{(l)} = g(c_g(X^{(l)}, \Theta_g)) \odot \sigma(c_\sigma(X^{(l)}, \Theta_\sigma)) \quad (17)$$

Where 1Dconvs refers to a convolution operation with a kernel size of 1×1 , which serves to normalize the data. $X^{(l)}$ is the output after the attention layer. $c_g(\cdot)$ and $c_\sigma(\cdot)$ are one-dimensional convolution operators, Θ_g and Θ_σ are

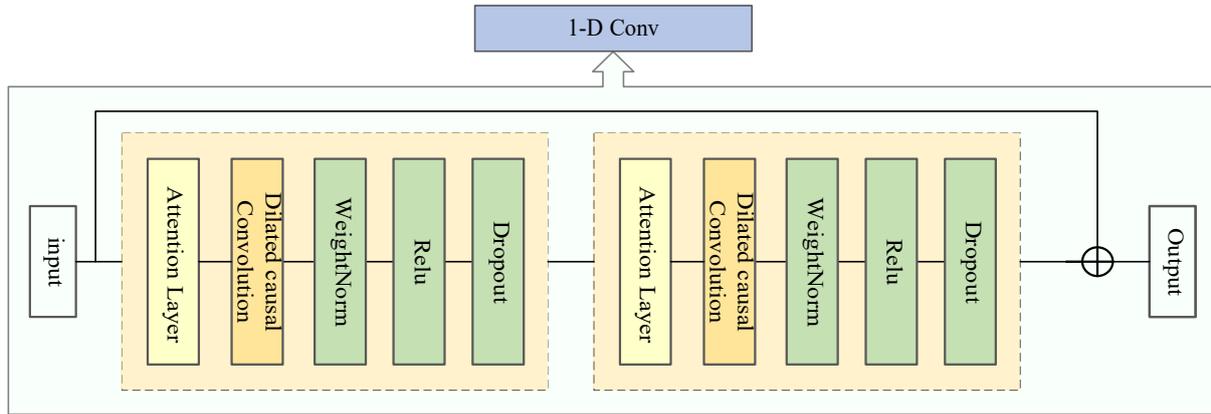


Fig. 8. Structure of ATCN residual connections

the parameters, \odot stands for the product of elements, the activation function $g(\cdot)$ corresponds to the ReLU of the output layer, and $\sigma(\cdot)$ is the sigmoid function.

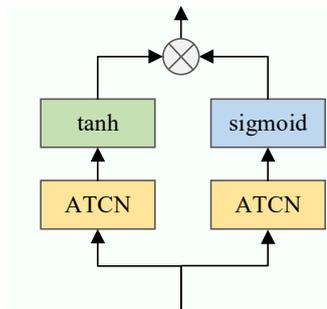


Fig. 9. Gate-ATCN architecture diagram

(2) Higher order differential convolution

As traffic flow exhibits complex nonlinearity and dynamics, the current traffic condition often has a significant impact on the traffic condition at the subsequent time step. As a result, building on the Gate-ATCN model presented in this paper, we design a higher-order differential temporal convolution module (HDTCN). This module effectively leverages the time difference data between the current and preceding specific time steps, ensuring a more comprehensive utilization of temporal information throughout the process. The particular structure is shown in Fig. 10, where a jump connection is used between the higher-order differential operation and the time-convolution layer to help the model perform backpropagation and thus alleviate the gradient vanishing problem.

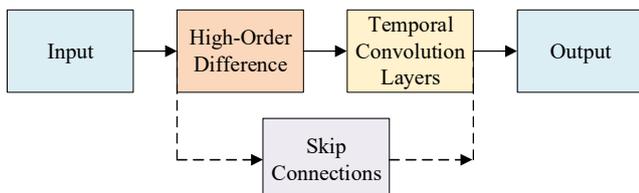


Fig. 10. HDTCN architecture diagram

The HDTCN can capture the dynamic changes of the regional nodes in the traffic road network in different time dimensions, thus effectively revealing the higher-order temporal correlations among traffic conditions. Similarly, let the input of layer l be $X^{(l)}$, and the time difference of $X^{(l)}$ is expressed as $X_{t,h}^{(l)}$:

$$X_{t,h}^{(l)} = \begin{cases} X_{t,h}^{(l)} = 0, t = 1, \dots, k \\ X_{t,h}^{(l)} = X_t^{(l)} - X_{t-k}^{(l)}, t = k + 1, \dots, P \end{cases} \quad (18)$$

Where k is the difference in step size. As the number of layers l increases, the differential signal changes from first order to higher order. In this way, the higher-order differential information of the time signal in the traffic network is increased.

Similar to Gate-ATCN, the gating mechanism used to enhance higher-order temporal difference information can be represented as:

$$X_{t,h}^{(l)} = \frac{\exp(W_a X_{t,h}^{(l)} + b_a)}{\sum_{l=1}^L \exp(W_a X_{t,h}^{(l)} + b_a)} X_{t,h}^{(l)} \quad (19)$$

$$X_h^l = \text{ReLU}(c_g(X_{t,h}^{(l)}, \Theta_{gh})) \odot \sigma(c_\sigma(X_{t,h}^{(l)}, \Theta_{\sigma h})) \quad (20)$$

Where Θ_{gh} and $\Theta_{\sigma h}$ are the convolutions of c_{gh} and $c_{\sigma h}$ parameters, respectively. Since there are many 0 elements in the differential signal, the activation function in Eq. ReLU is used to favor the transfer of useful differential information.

(3) Temporal feature fusion

Equations (2) and (4) extract low-order and high-order time-related information from spatio-temporal data, respectively, which are consistent in dimension. We fuse the results of the two formulas to obtain comprehensive time-related details and final temporal feature information. The formula for the calculation is given by:

$$X_T^l = X_a^l \odot W_a + X_h^l \odot W_h \quad (21)$$

X_T^l not only contains information directly related to the original time but also incorporates deeper and higher-order information, thus comprehensively reflecting the complexity and diversity of the time dimension in traffic flow prediction.

Spatial Convolution Layers

In a traditional graph, edges usually connect only two nodes, and their representation is relatively simple. In hypergraphs, hyperedges allow two or more nodes to connect, thus providing more prosperous and complex connectivity patterns. Therefore, to capture complex higher-order information and traditional graph structures' geographic and functional spatial properties, this paper utilizes graph convolution and higher-order hypergraph convolution to effectively capture spatial correlations based on the dynamic hypergraph constructed in Chapter II.

This paper initially employs GCN to capture both the geographic and functional spatial features from the corresponding graphs. The GCN updates the feature representation of target nodes by leveraging both the adjacency and feature matrices between nodes, along with aggregating the features of neighboring nodes. This enables the GCN to fully leverage the topology and spatial relationships of the road network, allowing it to effectively capture the network's spatial features. The formula for its calculation is given below:

$$y_g(t) = GCN(X(t)) = A_{ij}^g X(t) \Phi' \quad (22)$$

$$y_f(t) = GCN(X(t)) = A_{ij}^f X(t) \Phi' \quad (23)$$

Where A_{ij}^g and A_{ij}^f are the adjacency matrices of the geographic and functional maps, X is the input matrix of features, and Φ_1 is the parameter to be determined.

Subsequently, hypergraph convolution is utilized to identify the evolving higher-order spatial characteristics of regional nodes in the transportation network. The hypergraph convolution is $HConv(\cdot)$, whose inputs include the feature matrix $X = [X_1, \dots, X_T]$ and the adjacency matrix \hat{Z} . The method assigns traffic data from different time series to each node of the hypergraph structure. It performs a hypergraph convolution on each node to aggregate the hyperedge features and node features of the higher-order hypergraph of the road network to achieve high-order spatial correlation capturing. To wit:

$$y_i'(t) = HConv(X_i(t), \Theta, \hat{Z}) = \sigma \left(\sum_{n=1}^{|V_H|} \sum_{m=1}^{|E_H|} \hat{Z}(x_i, \varepsilon_m) \hat{Z}(x_j, \varepsilon_m) X_i(t) \Theta' \right) \quad (24)$$

Where $|V_H|$ and $|E_H|$ are the numbers of road network hypergraph nodes and edges, respectively, y_i' represents the output of the hypergraph convolution, σ corresponds to the sigmoid function, and Θ' denotes the parameter that is optimized during training.

Since the dimension of the adjacency matrix is different between the hypergraph and the traditional graph, for the output result of the convolution operation, this paper uses an FC to achieve the dimension conversion. Then, the converted result is fused to form a new feature vector. At the same time, an FC is used to map this new feature vector to the dimension required in the next step for further processing.

$$y_g'(t) = \sigma(y_g(t)W_g + b_g) \quad (25)$$

$$y_f'(t) = \sigma(y_f(t)W_f + b_f) \quad (26)$$

$$y_H'(t) = \sigma(y_i'(t)W_H + b_H) \quad (27)$$

$$y(t) = \text{Tanh} \left[\left[y_g'(t), y_f'(t), y_H'(t) \right] W_{fusion} \right] + b_{fusion} \quad (28)$$

The activation function y_g' for y_f' , y_H' , and σ is ReLU to ensure the nonlinear nature of the spatial properties while avoiding the problem of vanishing gradients. In the FC layer of the fusion result, Tanh is chosen as the activation function, which can map the input values to the range of $[-1, 1]$ and help speed up the convergence of the model.

In real scenarios, traffic data exhibit non-linear and unstable characteristics. Therefore, it is often tricky for different hybrid models to achieve good compatibility and complementarity in complex scenes and multi-parameter situations. This study applies the Kalman filter to minimize the model's deviation caused by multiple parameters. They are correcting the convolution output results to extract spatial features further to reveal the traffic data's essential features.

In the Kalman filtering process, in Fig. 11, the initial traffic data sequence produced by the first layer of H-GCN is labeled as $y(t)$, After applying the Kalman filter, the resulting value is referred to as $y(t)'$, this value, $y(t)'$, is then fed into the subsequent layer of H-GCN for further processing.

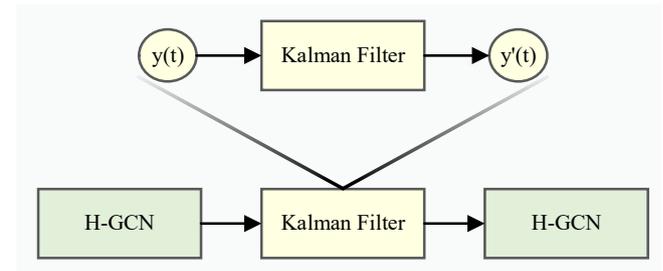


Fig. 11. Structure of spatial convolution layer

Traffic data at a given time t is inherently influenced by conditions both preceding and following that moment, and can therefore be expressed as follows:

$$y'(t) = \partial_0 y(t-1)^2 + \partial_1 y(t)^2 + \partial_2 y(t+1)^2 + \delta_0(t)^2 \quad (29)$$

Where ∂_0 , ∂_1 , and ∂_2 are the parameter matrices before and after t , i.e., $t-1$, t , and $t+1$, respectively. δ_0 is the artificially introduced observation noise, defined as the covariance matrix.

C. Output Module

Given that fully connected networks excel at identifying intricate patterns and nonlinear connections in traffic data, this study develops the output module by layering these networks.

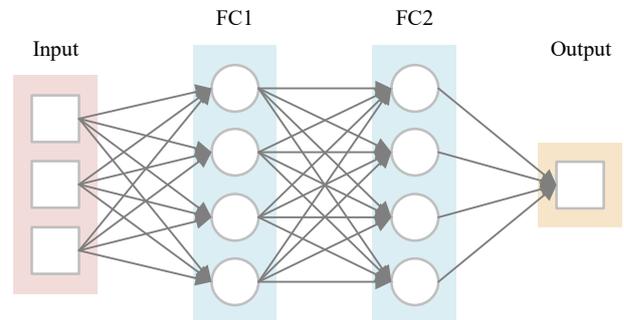


Fig. 12. Structure of the output module

As shown in Fig. 12, the output representation from the spatio-temporal interaction module serves as its input. Among them, the main task of the first layer of an FC network is to extract the nonlinear features from the output of the spatio-temporal interaction module. By leveraging the nonlinear transformation capabilities of the activation function, the initial layer of a fully connected network can

transform input data into a high-dimensional feature space, thereby capturing a more precise representation of the data's underlying structure. The second layer of the fully connected network is responsible for fusing the feature information from the previous output. The linear combination of weight matrices and bias terms further integrates the interactions between the features, enabling the extraction of predictions and the output of predicted sequences $Y = (Y_1, Y_2, Y_3, \dots, Y_N)$.

IV. EXPERIMENTAL RESULTS AND ANALYSIS

A. Data Description

This paper uses two representative traffic datasets publicly available in DCRNN: METR-LA and PEMS-BAY. The METR-LA dataset focuses on motorway traffic conditions across Los Angeles County, spanning March to June 2012, with data gathered by 207 sensors. The PEMS-BAY dataset, by contrast, emphasizes traffic conditions in the Bay region and contains data collected by 325 loop sensors from January to May 2017—Table I details information about these two datasets.

TABLE I
STATISTICAL INFORMATION ON DATA SETS FOR METR-LA AND PEMS-BAY

DATA	NODES	EDGES	TIME STEPS
METR-LA	207	1515	34272
PEMS-BAY	325	2369	52116

B. Experimental Setup

Parameter Setting

We used Python 3.7 in this experiment, built the framework based on PyTorch, with the development environment PyCharm, and computed on a GPU (RTX 3090 24GB). We utilized the METR-LA and PEMS-BAY, partitioning them temporally into training, validation, and test sets with a 7:1:2 ratio. To thoroughly capture the spatio-temporal characteristics, we developed a spatio-temporal interaction module with a three-layer ST-Block structure. For both temporal convolution and spatial convolution, we chose 3*3 convolutional cores. For the temporal convolution layer, we used Gate-ATCN and set its expansion factors to 1, 2, and 4 in order. The prediction step size is 3, and the learning rate is set to 0.001, with a batch size of 32.

Parametric Analysis

Since the learning rate, batch size, and the number of attention heads in the hypergraph construction method significantly impact the model performance during the experiments, the above parameters are analyzed using experimental search.

(1) Attention heads

In constructing a hypergraph dynamically in GAT, the choice of the number of attention heads is a crucial parameter, which will directly affect the degree of attention of each regional node in the hypergraph to its related nodes. Therefore, we further analyze the model's training performance on the PEMS-BAY by testing the variation in the count of attention heads.

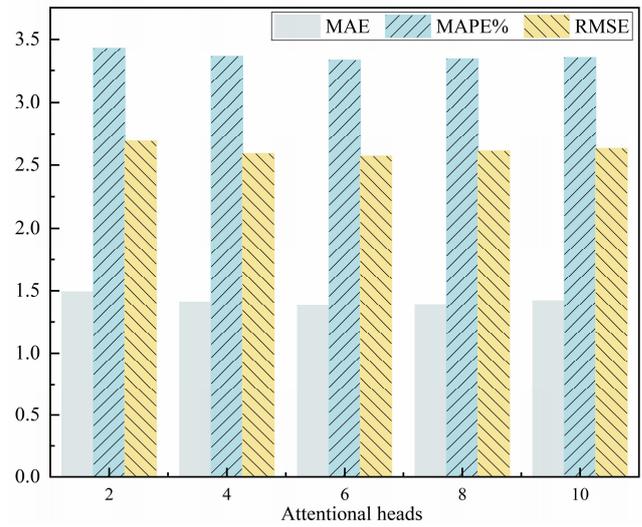


Fig. 13. Effect of the number of attentional heads on the predictive performance

In Fig. 13, the model exhibits suboptimal performance with 2 or 4 attention heads. Notably, model accuracy improves as attention heads increase, with optimal performance observed at 6 heads. However, further increases in attention heads lead to a decline in prediction accuracy. This degradation can be attributed to excessive model complexity resulting from an overabundance of attention heads. Consequently, this study identifies 6 attention heads as the optimal configuration for maximizing prediction accuracy.

(2) Learning rate

The learning rate is an essential parameter in neural network training, and a suitable learning rate can significantly speed up the training process and help enhance the model's performance. Therefore, for the PEMS-BAY dataset, we analyze the model's training effect by testing the learning rate variation in the set of {0.0001, 0.001, 0.005, 0.01} and further in-depth.

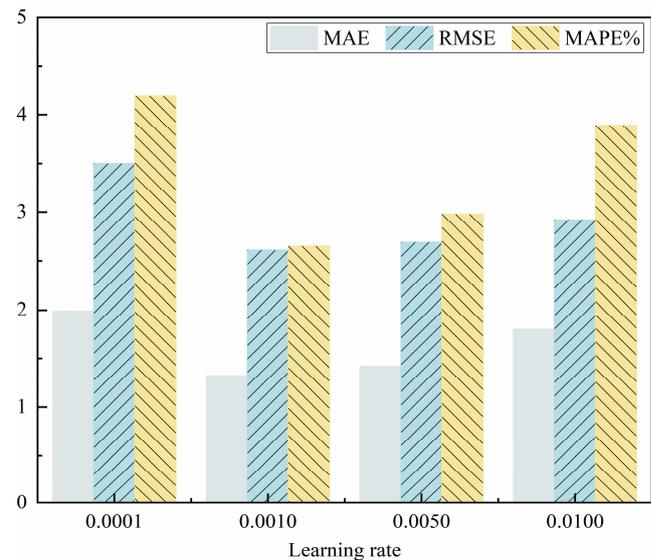


Fig. 14. Effect of the number of learning rates on the predictive performance

As illustrated in Fig. 14, the model's predictive performance is unsatisfactory when the learning rate is 0.0001. However, when it was increased to 0.001, the model's performance was significantly improved. At that time, the MAE, RMSE, and MAPE values were all minimized, and the

model performed optimally. When the learning rate was further increased to 0.005 and 0.01, the model's performance showed a decreasing trend. It may be because a learning rate that is too high tends to make the model fall into an overfitting predicament, resulting in its impaired generalization ability. Therefore, we determine the learning rate in this paper to be 0.001.

(3) Batch size

The batch size, determining how many samples are introduced to the neural network during each training iteration, plays a crucial role in shaping not only the speed of training but also the convergence behavior and the model's ability to generalize effectively to new data. To investigate these effects, we analyzed the model's training performance on the PEMS-BAY, with the learning rate fixed at 0.001. We explored variations in batch size within the range of {8, 16, 32, 64}.

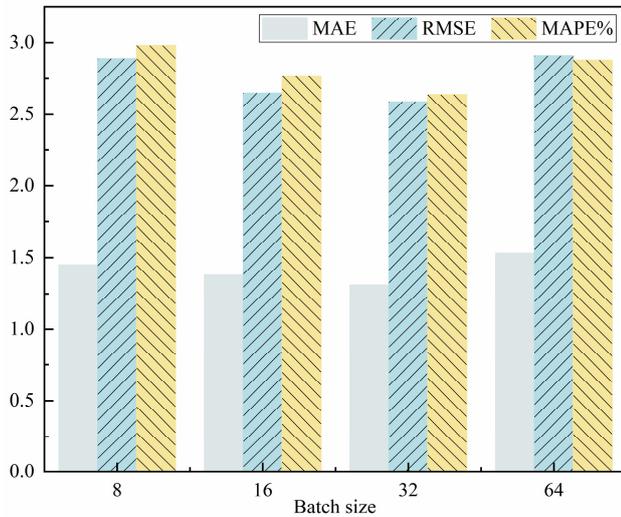


Fig. 15. Effect of the number of batch sizes on the predictive performance

As shown in Fig. 15, the model performs poorly when the batch size is 8 or 16. However, the model achieves peak performance when the batch size is increased to 32. Notably, the model's performance declines when the batch size is increased further to 64. The primary reason is that a large batch size can lead to model overfitting on the training data, ultimately diminishing its overall performance. Therefore, in this paper, we assign the value of 32 to this.

C. Evaluation Metrics

To thoroughly assess the efficacy of various prediction methods, this study employs three principal metrics: Mean Absolute Error (MAE), Mean Absolute Percentage Error (MAPE), and Root Mean Square Error (RMSE). Among them, MAE visualizes the actual magnitude of the prediction error, MAPE focuses on revealing the relative degree of deviation between the predicted and actual values, and RMSE is derived from the square root of the mean squared error. In the context of these metrics, a lower RMSE value signifies superior predictive accuracy.

$$MAE(y, \hat{y}) = \frac{1}{N} \sum_{i=1}^N |y_i - \hat{y}_i| \quad (30)$$

$$MAPE(y, \hat{y}) = \frac{100\%}{N} \sum_{i=1}^N \left| \frac{y_i - \hat{y}_i}{y_i} \right| \quad (31)$$

$$RMSE(y, \hat{y}) = \sqrt{\frac{1}{N} \sum_{i=1}^N (y_i - \hat{y}_i)^2} \quad (32)$$

Where $y = y_1, \dots, y_n$ denotes the true value, $\hat{y} = \hat{y}_1, \dots, \hat{y}_n$ represents the model's predicted value, and N refers to the index of the observed sample.

D. Baseline Modelling

To evaluate the efficacy of the method introduced in this study, the traditional method (e.g., HA) and machine learning algorithms (e.g., SVM, LSTM) are analyzed experimentally, respectively. In addition, this paper also compares the proposed method with deep learning methods based on GCN and RNN (such as STGCN, ASTGCN, and DCRNN) and the model DHAT using hypergraph convolution.

HA: This model employs the historical mean of traffic data over time as a prediction result by building on a smooth traffic time series.

SVM: This model solves the nonlinear problem of traffic data by introducing a kernel function and establishing a hyperplane in a high-dimensional space.

LSTM: This model effectively captures and remembers long-term dependencies in traffic sequence data, thus enabling traffic prediction over some time in the future.

STGCN: The model uses a purely convolutional architecture, in which the spatial convolutional layer effectively captures spatial information using spectral domain map convolution, while the temporal convolutional layer employs gated operations to extract the dynamic characteristics in the time aspect.

ASTGCN: This model employs an attention mechanism to dynamically generate a mask matrix, effectively capturing traffic data's dynamic characteristics.

DHAT: This model introduces directed hypergraph convolution and combines it with an attention mechanism to effectively realize the prediction task.

DCRNN: This model utilizes the graph's random wandering approach and effectively captures spatiotemporal features by constructing diffusion convolution gated loop units.

E. Analysis of Experimental Results

From the perspective of comprehensive model performance, this paper compares the proposed STK-HGCN model with the baseline models of the above six. We used 15 minutes of data from the METR-LA and PEMS-BAY datasets. Table II and Table III display the details.

As shown in Tables II and III, the HA model demonstrates the lowest predictive accuracy on the PEMS-BAY and METR-LA datasets, which is mainly attributed to the fact that the HA method adopts the historical mean as the prediction result, which can achieve relatively good results on smooth datasets, the prediction performance falls short when applied to the intricate, nonlinear time-series characteristics of traffic flow. The SVM approach, compared to the traditional HA model, can deal with nonlinear data, but its spatiotemporal characteristics of traffic data are not enough for mining, and the prediction effect is poor. The LSTM model accounts for the temporal characteristics of

TABLE II

ANALYSIS OF VARIOUS MODELS' PERFORMANCE ON THE PEMS-BAY

Model	MAE	RMSE	MAPE%
HA	2.78	5.59	6.08
SVM	1.75	3.60	3.81
LSTM	1.64	3.15	3.42
STGCN	1.47	2.98	3.02
ASTGCN	1.48	3.04	3.05
DHAT	1.45	2.90	2.97
DCRNN	1.45	2.95	2.81
STK-HGCN	1.30	2.58	2.63

TABLE III

ANALYSIS OF VARIOUS MODELS' PERFORMANCE ON THE METR-LA

Model	MAE	RMSE	MAPE%
HA	4.80	10.00	11.70
SVM	3.59	8.45	9.30
LSTM	3.10	7.88	8.98
STGCN	2.80	5.64	7.32
ASTGCN	2.59	5.28	6.33
DHAT	2.70	5.41	6.65
DCRNN	2.67	5.36	6.59
STK-HGCN	2.51	5.15	6.08

nodes within the traffic network. However, it overlooks the influence of spatial correlations, which ultimately hinders its predictive accuracy. Methods based on TCN and GCN offer a marked advantage over traditional approaches. Notably, the ASTGCN model incorporates an attention mechanism, unlike the STGCN, enabling it to dynamically assign varying weights to each module and capture the evolving correlations between nodes. DHAT leverages hypergraph convolution in conjunction with an attention mechanism to effectively model the spatio-temporal dependencies inherent in traffic data. DCRNN employs random walks on the graph to explore its structure, achieved through the construction of diffusion convolution-gated recurrent units, thereby enabling the capture of complex spatiotemporal features.

Compared to the previous models, the model in this paper incorporates hypergraph and time difference information to illustrate the potential higher-order details of the transportation network. It also builds the spatial-temporal interaction module of the "sandwich" structure. This effectively enhances the model's ability to process spatial-temporal interaction information, allowing STK-HGCN to more effectively capture both the dynamics of spatiotemporal interaction information and the influence of higher-order spatial connections within the road network structure. As a result, the model's predictive accuracy is significantly enhanced. In the 15-minute prediction range, on the PEMS-BAY, the STK-HGCN model improved MAE by 0.15 and RMSE by 0.37, with a significant reduction of 0.18% in MAPE. On the METR-LA dataset, the STK-HGCN model improved MAE by 0.08 and RMSE by 0.13, with a significant reduction of 0.25% in MAPE.

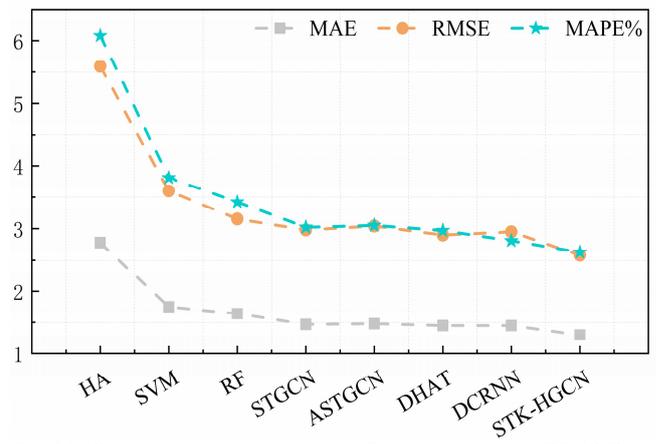


Fig. 16. Comparative error assessment of various models for PEMS-BAY

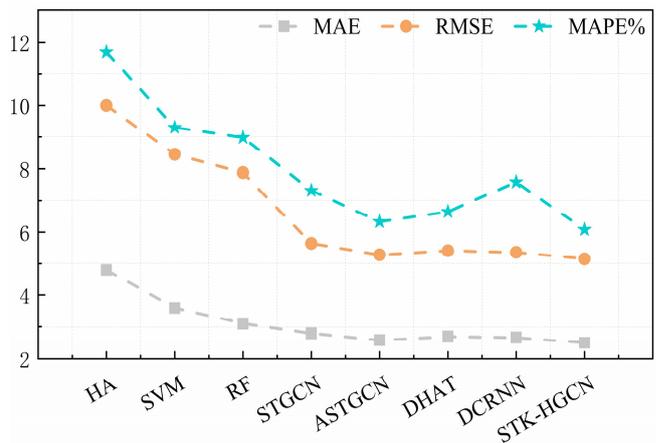


Fig. 17. Comparative error assessment of various models for METR-LA

The model's prediction performance in this paper and the above baseline model are visualized and analyzed. As shown in Fig. 16 and 17, the Figures show the comparative analysis of the errors of the PEMS-BAY and the METR-LA datasets, respectively. Through these two visualizations, we can observe the advantage of the STK-HGCN model in prediction performance. Since the STK-HGCN model considers the effects of higher-order information and spatio-temporal interaction information, its prediction performance shows obvious superiority compared to other baseline models. This outcome further confirms the strong efficacy of the STK-HGCN model in significantly improving prediction accuracy.

To gain a deeper understanding of the STK-HGCN model's performance, we randomly select a specific sensor on the PEMS-BAY and METR-LA datasets and visualize its weekly traffic speed. The critical metric of speed enables traffic flow dynamics to be analyzed visually. In the speed graph, lower speeds represent traffic congestion in the current period, while higher speeds indicate unimpeded traffic flow. As shown in Fig. 18 and 19, the overall fitting effect of the model is excellent because the STK-HGCN model fully considers the higher-order dependencies and the feature extraction of the traffic network is more comprehensive. Especially in the peak region where the data volume fluctuates wildly, the STK-HGCN model can still match the actual data well, showing its strong adaptability and stability. With the increasing prediction duration, the STK-HGCN model can still better fit daily data trends, proving its ability to capture higher-order information and spatiotemporal interactivity between nodes in different regions.

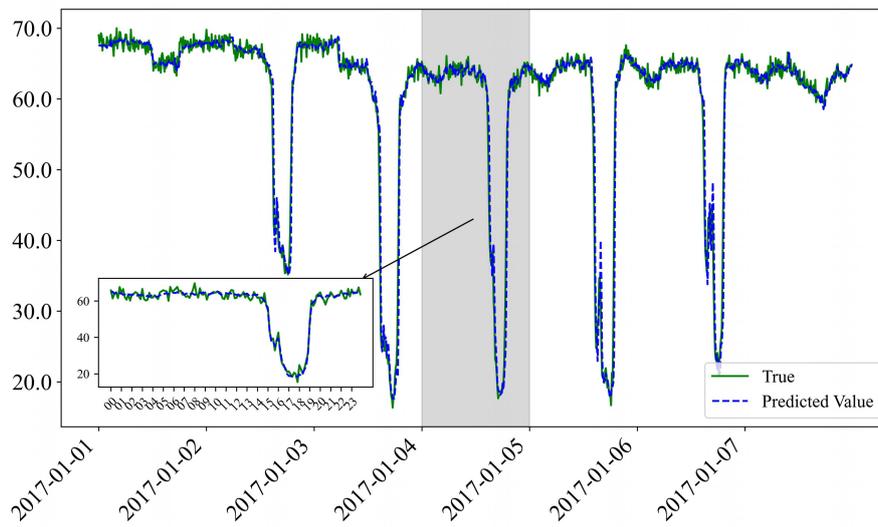


Fig. 18. STK-HGCN traffic data prediction for a sensor week on the PEMS-BAY dataset

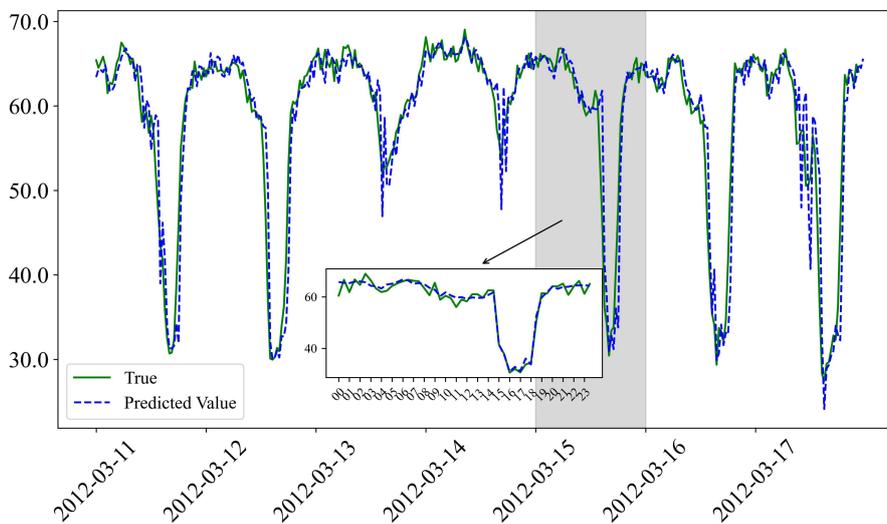


Fig. 19. STK-HGCN traffic data prediction for a sensor week on the METR-LA dataset

In addition, ASTGCN, STGCN, DHAT, and DCRNN, which have relatively good prediction results, are selected to compare with the models in this paper, and we forecast traffic data for the next 15, 30, and 45 minutes by using a segment of historical traffic data on the METR-LA, respectively. The corresponding visualization results are in Fig. 20 to Fig. 22. By analyzing the presented results in Fig., It has been observed that the STGCN model shows some capability in

capturing underlying trends in traffic data. but its prediction performance is poor. The ASTGCN model employs a fixed spatial convolution kernel, which cannot dynamically adjust to variations in the spatial relationships among different nodes, and the prediction effect is not good. In contrast, the DHAT model improves prediction accuracy. However, it uses a directed hypergraph that remains constant under different traffic conditions, which limits its performance

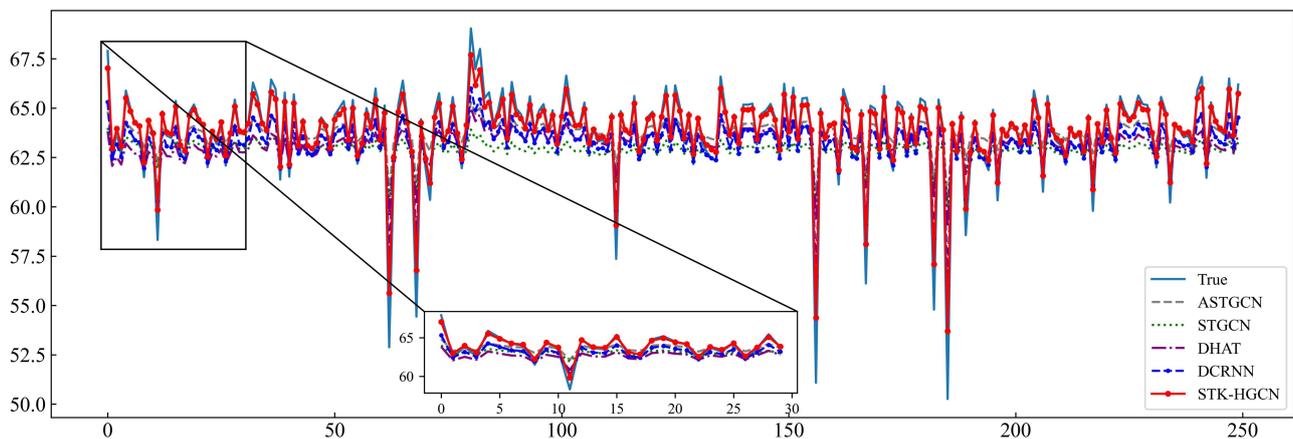


Fig. 20. Predicted results for 15 minutes

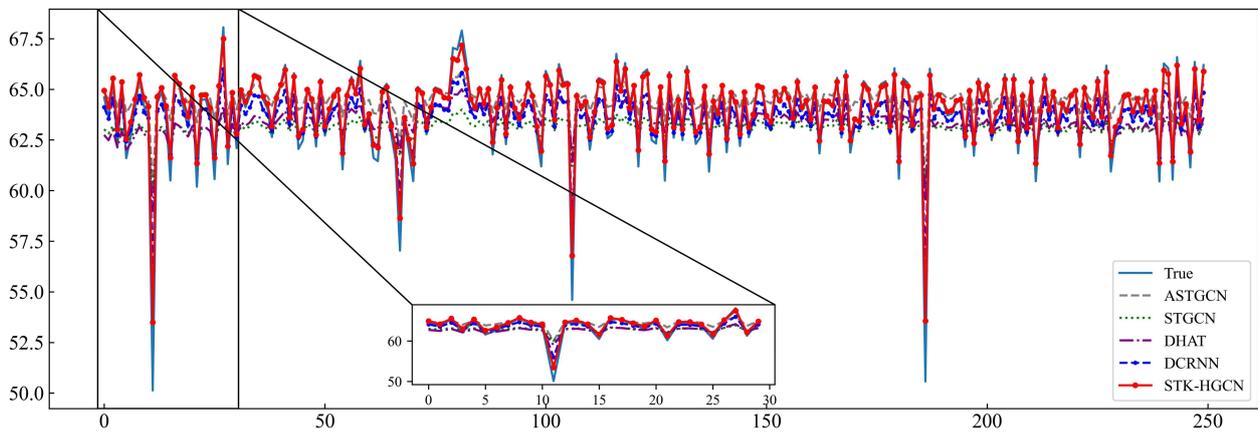


Fig. 21. Predicted results for 30 minutes

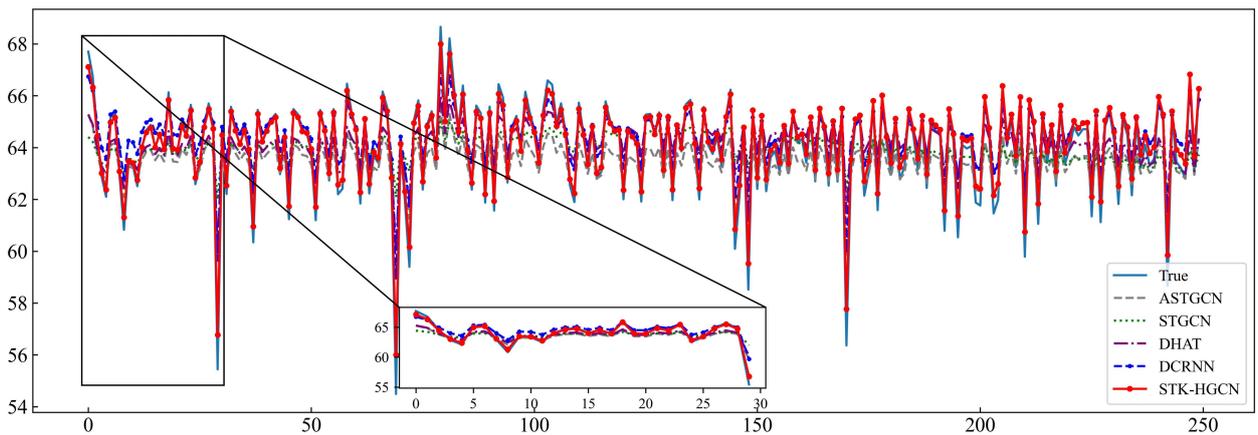


Fig. 22. Predicted results for 45 minutes

in modeling dynamic spatial relationships. In this case, the DCRNN model utilizes the topological graph structure to capture the spatio-temporal properties, further improving its prediction performance. Compared to other comparative models, The DCRNN captures the spatial dependencies inherent in the transportation network's topology by modeling bi-directional stochastic diffusion across the graph, which enables it to capture the spatio-temporal relationships between nodes more accurately. The model introduced in this study integrates spatiotemporal information and interaction features to fit the actual data better, showing that the STK-HGCN model better captures spatiotemporal correlations between different nodes. The model shows stable performance in predicting 15, 30, and 45 minutes, which can deeply explore the spatio-temporal characteristics of traffic data and effectively reflect the actual road network information.

F. Ablation Experiments

Expanding upon the baseline models, this section assesses the performance of each element of the STK-HGCN framework via ablation experiments.

- (1) M-TH: removing temporal higher-order correlation features from the STK-HGCN model;
- (2) M-SH: removing spatial higher-order correlation features in the STK-HGCN model;
- (3) M-P: removing geospatial correlation in the STK-HGCN model;
- (4) M-F: removing functional spatial correlations in the STK-HGCN model;

- (5) M-K: removing the Kalman filter component in the STK-HGCN model.

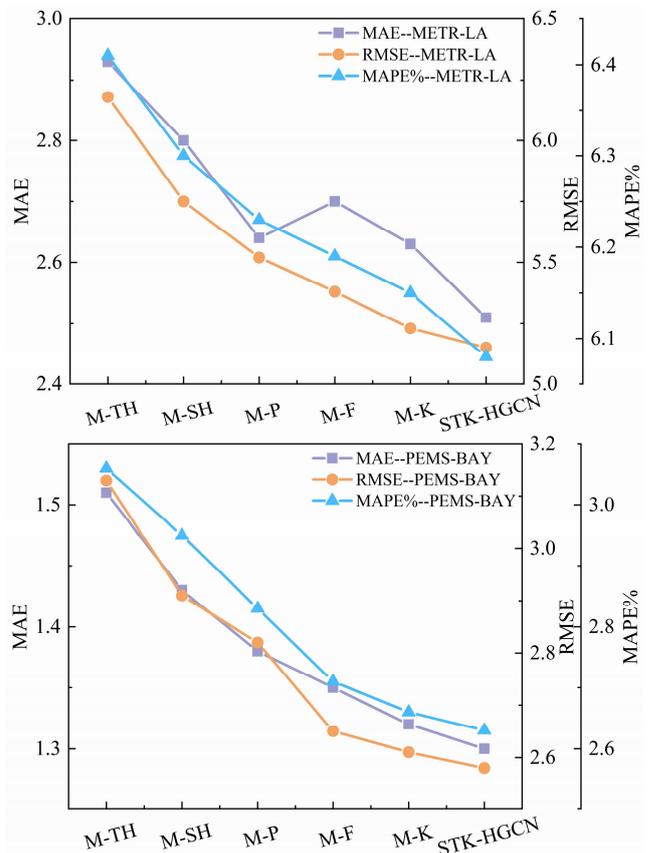


Fig. 23. Comparative analysis of ablation experiments

Fig. 23 illustrates the comparison results of STK-HGCN and its five variant models on RMSE, MAE, and MAPE. The results indicate that the STK-HGCN consistently surpasses all other variants across the three evaluation metrics. This finding underscores the efficacy of incorporating the model's distinct components. When the geographic and functional maps are removed, the model prediction performance decreases, indicating that it is crucial to consider geospatial correlation and functional spatial correlation. When temporal and spatial higher-order correlations are removed, the predictive performance of both M-TH and M-SH decreases significantly. However, in contrast, the prediction using only temporal higher-order correlation is considerably better than spatial higher-order correlation. This could be because traffic flow prediction is fundamentally a temporal forecasting problem, making temporal feature extraction crucial to the prediction results. Furthermore, the model performance is also impaired when the Kalman filter is removed, which demonstrates the necessity of introducing the Kalman filter for model robustness. In this paper, the model considers spatio-temporal low-order and high-order correlations and introduces the Kalman filter, effectively improving prediction accuracy.

V. CONCLUSION

We present STK-HGCN, a model for short-term traffic flow forecasting that leverages a hypergraph convolutional network. This framework enhances the precision and efficiency of predicting traffic flow.

(1) To effectively capture complex higher-order traffic information across time and space, we designed the GH-TCN and KH-GCN modules. The GH-TCN module uses higher-order temporal differential convolution and gated temporal convolution based on the attention mechanism to capture temporal features. In the spatial dimension, we model the traffic data as a geographic graph, functional graph, and high-order hypergraph, and we use a dual-layer graph convolution method in the KH-GCN module.

(2) To fully consider the interaction characteristics of spatio-temporal information and the robustness, we construct a spatio-temporal interaction module with a "sandwich" structure to integrate spatio-temporal features and obtain spatio-temporal interaction information. At the same time, we introduce a Kalman filter to correct the convolutional output and alleviate the model's bias due to the introduction of too many parameters.

(3) We constructed the M-TH, M-SH, M-P, M-F, and M-K models by removing each component to analyze the impact of different model components. The evaluation metrics show that considering the higher-order spatio-temporal information and interaction characteristics significantly enhances the predictive performance of the STK-HGCN model. The experimental results are even superior with further introduction of the Kalman filter. Additionally, the STK-HGCN model outperforms the comparison model in evaluation metrics such as MAE, RMSE, and MAPE, particularly in 15-minute, 30-minute, and 45-minute predictions, which align best with the original data.

While our model demonstrates improved predictive performance relative to the comparison model, it does not address the issue of data noise. Traffic data collection can be

influenced by varying levels of noise, representing a limitation of our current approach. Future work will focus on mitigating data noise to further enhance the model's prediction accuracy.

REFERENCES

- [1] J. Yao, R. Chu, T. Shi, P. Wang, and X. Zhao, "Review on machine learning-based traffic flow prediction methods," *Journal of Traffic and Transportation Engineering*, vol. 23, no. 3, pp. 44-67, 2023.
- [2] Q. Wang, X. Chen, C. Zhu, and W. Chai, "Short-term Traffic Flow Prediction Based on the SGA-KGCN-LSTM Model," *Engineering Letters*, vol. 31, no. 3, pp. 1221-1235, 2023.
- [3] Z. Yaying and H. Guan, "Traffic Flow Prediction Model Based on Deep Belief Network and Genetic Algorithm," *IET Intelligent Transport Systems*, vol. 12, no. 6, pp. 533-541, 2018.
- [4] J. Wu, H. Zhang, and X. Ran, "Nonparametric Regressive Short-term Traffic Flow Forecast Algorithm Based on Data Reduction and SVM," *Journal of Highway and Transportation Research and Development*, vol. 37, no. 7, pp. 129-134, 2020.
- [5] F. Harrou, A. Zeroual, and Y. Sun, "Traffic Congestion Monitoring Using an Improved KNN Strategy," *Measurement*, vol. 156, pp. 107534, 2020.
- [6] Y. Gu, W. Lu, X. Xu, L. Qin, Z. Shao, and H. Zhang, "An Improved Bayesian Combination Model for Short-Term Traffic Prediction With Deep Learning," *IEEE Transactions on Intelligent Transportation Systems*, vol. 21, no. 3, pp. 1332-1342, 2020.
- [7] Z. Sun, Y. Hu, W. Li, S. Feng, and L. Pei, "Prediction Model for Short-Term Traffic Flow Based on A K-Means-Gated Recurrent Unit Combination," *IET Intelligent Transport Systems*, vol. 16, no. 5, pp. 675-690, 2022.
- [8] W. Li, T. Zou, H. Wang, and H. Huang, "Traffic Accident Quantity Prediction Model Based on Dual-scale Long Short-Term Memory Network," *Journal of Zhejiang University (Engineering Science)*, vol. 54, no. 8, pp. 1613-1619, 2020.
- [9] Q. Ouyang, T. Sun, Y. Xue, and Z. Liu, "Long Short-Term Memory and Graph Convolution Network for Forecasting the Crude Oil Traffic Flow," *IEEE Access*, vol. 10, pp. 18922-18932, 2022.
- [10] Q. Wu, Z. Jiang, K. Hong, H. Liu, and J. Ding, "Tensor-Based Recurrent Neural Network and Multi-Modal Prediction With Its Applications in Traffic Network Management," *IEEE Transactions on Network and Service Management*, vol. 18, no. 1, pp. 780-792, 2021.
- [11] G. Urbarri and G. B. Mindlin, "Dynamical Time Series Embeddings in Recurrent Neural Networks," *Chaos, Solitons & Fractals*, vol. 154, pp. 111612, 2022.
- [12] N. Hu, D. Zhang, K. Xie, W. Liang, C. Diao, and K. C. Li, "Multi-range Bidirectional Mask Graph Convolution Based GRU Networks for Traffic Prediction," *Journal of Systems Architecture*, vol. 133, pp. 102775, 2022.
- [13] W. Zhang, R. Yao, X. Du, Y. Liu, R. Wang, and L. Wang, "Traffic Flow Prediction Under Multiple Adverse Weather Based on Self-Attention Mechanism and Deep Learning Models," *Physica, A. Statistical mechanics and its applications*, vol. 625, pp. 128988, 2023.
- [14] W. Zhang, Y. Yu, Y. Qi, F. Shu, and Y. Wang, "Short-term traffic flow prediction based on spatio-temporal analysis and CNN deep learning," *Transportmetrica*, vol. 15, no. 2, pp. 1688-1711, 2019.
- [15] T. Mallick, P. Balaprakash, E. Rask, and J. Macfarlane, "Graph-Partitioning-Based Diffusion Convolutional Recurrent Neural Network for Large-Scale Traffic Forecasting," *Transportation Research Record*, vol. 2674, no. 2, pp. 473-488, 2020.
- [16] Z. Wu, S. Pan, G. Long, J. Jiang, and C. Zhang, "Graph WaveNet for Deep Spatial-Temporal Graph Modeling," in *Proc. 28th International-Joint Conference on Artificial Intelligence*, 10-16 August, 2019, Macao, China, pp. 1907-1913.
- [17] B. Yu, H. Yin, and Z. Zhu, "Spatio-Temporal Graph Convolutional Networks: A Deep Learning Framework for Traffic Forecasting," in *Proc. 27th International-Joint Conference on Artificial Intelligence*, 13-19 July, 2018, Stockholm, Sweden, pp. 3634-3640.
- [18] Z. Liu, F. Ding, Y. Dai, L. Li, T. Chen, and H. Tan, "Spatial-Temporal Graph Convolution Network Model With Traffic Fundamental Diagram Information Informed for Network Traffic Flow Prediction," *Expert Systems with Applications*, vol. 249, pp. 123543, 2024.
- [19] Y. Gao, Y. Feng, S. Ji, and R. Ji, "HGNN+: General Hypergraph Neural Networks," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 45, no. 3, pp. 3181-3199, 2023.
- [20] Y. Zhang, Z. Wu, Y. Lin, and Y. Zhao, "Spatio-temporal hyper-relationship graph convolutional network for traffic flow forecasting," *Journal of Computer Applications*, vol. 41, no. 12, pp. 3578-3584, 2021.

- [21] X. Luo, J. Peng, and J. Liang, "Directed Hypergraph Attention Network for Traffic Forecasting," *IET Intelligent Transport Systems*, vol. 16, no. 1, pp. 85-98, 2021.
- [22] J. Tully, R. Haight, B. Hutchinson, S. Huang, J. Y. Lee, and S. Katipamula, "Dilated Causal Convolutional Neural Networks for Forecasting Zone Airflow to Estimate Short-Term Energy Consumption," *Energy and buildings*, vol. 286, pp. 112890, 2023.
- [23] X. Zhang and J. You, "A Gated Dilated Causal Convolution Based Encoder-Decoder for Network Traffic Forecasting," *IEEE Access*, vol. 8, pp. 6087-6097, 2020.
- [24] T. Qi, G. Li, L. Chen, and Y. Xue, "ADGCN: An Asynchronous Dilation Graph Convolutional Network for Traffic Flow Prediction," *IEEE Internet of Things Journal*, vol. 9, no. 5, pp. 4001-4014, 2022.