

Urban traffic flow prediction based on Multi-spatio-temporal correlation analysis

Abstract—Predicting future traffic conditions in road traffic networks using historical traffic data has extensive applications in enhancing people’s daily lives, managing road traffic, and facilitating urban planning and design. However, due to the highly dynamic nature of traffic flow, which is spatially and temporally dependent with nonlinear and non-stationary characteristics, predicting traffic flow still encounters significant challenges. To address this issue, this paper proposes a method for predicting traffic flow based on a Multi-Spatial Temporal Correlation Model (MSTCM). Firstly, the temporal features of traffic flow are decomposed into multi-scale representations that capture internal trends and fluctuations caused by external disturbances respectively. Simultaneously, the original spatial features are decomposed into multi-dimensional representations that convey information about road proximity and direction within the road network. Secondly, an attention mechanism combined with a graph convolutional neural network is employed to extract and integrate global as well as local spatial-temporal dependencies for each spatial-temporal combination. Finally, fusion features from each dimension are weighted and recombined before being applied to predict future traffic flow patterns. The proposed method can effectively extract fine-grained temporal-spatial features from multiple perspectives while capturing nonlinear dynamics inherent in the data; it also mitigates noise interference on traffic flow prediction performance. Extensive experiments conducted on four real-world datasets along with comparative analysis against twelve existing algorithm models over a five-year period demonstrate favorable performance of our proposed model.

Index Terms—Traffic flow prediction; Multi-spatio-temporal correlation; Eigen decomposition; Attention mechanism; Graph convolutional neural network; Feature fusion

I. INTRODUCTION

With the ongoing process of urbanization and rapid growth in urban population, traffic congestion has emerged as a pervasive issue that cannot be disregarded by major cities worldwide. To address these challenges comprehensively, nations across the globe are actively exploring strategies to mitigate traffic congestion.

Traffic flow prediction plays a crucial role in forecasting the future trends of traffic by analyzing historical data, thereby facilitating informed travel planning for individuals and aiding traffic management departments in formulating effective control measures and emergency guidance plans. This serves as an important approach to alleviate traffic congestion. The existing methods for traffic flow prediction can be categorized into three groups: point prediction, regional prediction [1]–[3], and network prediction [4]–[8]. These approaches respectively focus on studying traffic data at specific roads, regions or entire road networks to analyze and predict their future trends and traffic flow. In recent years, the widespread deployment

of sensors in the transportation domain has facilitated more convenient access to real-time and comprehensive traffic data, thereby providing robust support for network-level traffic flow prediction research.

However, predicting urban traffic flow remains highly challenging due to two primary reasons. Firstly, the temporal and spatial relationships in traffic flow are not independent but rather coupled, forming complex spatial-temporal dependencies under the joint action of global resonance and local perturbation [9]. Secondly, urban traffic is influenced by various random factors such as people’s travel habits, weather conditions, road construction activities, and traffic accidents. Additionally, the relationship between supply and demand further complicates matters. Consequently, multiple periodic and non-periodic signals coexist within a complex spatial-temporal domain with non-linear and non-stationary characteristics. This presents a significant obstacle for accurate traffic flow prediction.

To capture the intricate and dynamic spatial-temporal correlation relationships in traffic flow, scholars have proposed two categories of temporal-spatial correlation models: hybrid models and fusion models. Hybrid models employ distinct neural networks to independently capture the temporal and spatial dependence relationships in traffic flow, which are then combined using a specific strategy for accurate traffic flow prediction. In contrast, fusion models utilize more sophisticated neural network architectures to analyze and predict traffic flow by integrating spatial-temporal features, enabling a more precise representation of real urban traffic environments and capturing the complex spatial-temporal correlation relationships. However, most existing fusion models primarily focus on local spatial-temporal dependencies while neglecting comprehensive consideration of global dependencies within the road network; moreover, these models often treat urban road traffic networks as undirected graphs without accounting for directional characteristics. To mitigate the impact of nonlinear and non-stationary characteristics on traffic flow prediction, some researchers have employed empirical mode decomposition or wavelet decomposition techniques to decompose traffic flow into multiple time series exhibiting simpler fluctuation patterns. Alternatively, the trend-cycle decomposition method has been utilized to partition the original traffic flow signal into periodic signals and fluctuation signals. Subsequently, time series analysis methods and supervised learning approaches are applied to extract features from each traffic flow signal sequence for accurate prediction. However, these approaches primarily focus on single-point traffic flow prediction with-

out considering spatial dependence relationships within road network traffic flows, rendering them unsuitable for whole network traffic flow prediction.

To tackle these challenges, we conduct a comprehensive analysis of traffic flow characteristics in both spatial and temporal dimensions, exploring the interdependencies between traffic flow across different spatial and temporal scales. Subsequently, we propose a novel urban traffic flow prediction method based on the Multiple Spatio-Temporal Correlation Model (MSTCM). The MSTCM model leverages attention mechanisms and convolutional neural networks to extract global and local spatial-temporal features, integrating diverse traffic flow attributes into features that capture their spatio-temporal dependencies. Ultimately, we aggregate these spatio-temporal traffic flow dependence features with appropriate weights and concatenate them to construct higher-dimensional spatial-temporal representations for predicting future traffic flows. The key contributions of this study are as follows:

- The temporal and spatial characteristics of traffic flow are decomposed into multiple scale temporal representations and multi-dimensional spatial representations, effectively mitigating the impact of noise on prediction results and enhancing the model’s capacity to capture traffic flow features.
- By analyzing and integrating temporal and spatial features of traffic flow from multiple perspectives, the model’s prediction accuracy was significantly enhanced.
- By integrating attention mechanism and convolutional neural networks, the proposed model effectively captures intricate global and local features from traffic flow data, thereby enhancing its capacity for nonlinear feature extraction through increased network depth.
- Extensive experiments were conducted to evaluate the performance of our algorithmic model using four real-world datasets. Comparative analysis against 12 prominent research achievements from the past five years demonstrates outstanding performance.

II. RELATED WORK

This section will succinctly summarize the principal research accomplishments in three domains: deep learning applications in traffic flow prediction, spatial-temporal correlation models, and graph neural networks.

A. Deep Learning-based Traffic Flow Prediction

In recent years, deep learning-based methods for traffic flow prediction have garnered significant attention due to their exceptional feature extraction and nonlinear fitting capabilities, enabling the revelation of intrinsic spatio-temporal correlations in road network traffic flow data and precise modeling of its evolution. To achieve this objective, researchers have dedicated themselves to developing diverse neural network architectures for traffic flow prediction. For instance, RNN [3], [4], [10] or GRU [11] models are employed to explore temporal correlations in time-series traffic data, while GCN [12], [13] structures capture the spatial dependencies of traffic data.

Moreover, neural network structures that integrate spatio-temporal features [4], [5], [7], [14], [15] have been extensively utilized to simultaneously extract both spatial and temporal characteristics of traffic flow. These approaches effectively uncover complex spatio-temporal dependence relationships in urban traffic data and demonstrate notable advantages in predicting traffic flow.

However, the intricate structure of urban traffic networks, characterized by diverse adjacency relationships between nodes, poses challenges in defining a unified convolution operation, thereby impeding the application potential of graph convolutional neural networks in terms of learning capability [16]. Simultaneously, when directly applied to traffic data, graph convolutional neural networks are unable to preserve directional information in traffic flow due to their suitability for processing undirected graph structures. Furthermore, the nonlinear and non-stationary characteristics inherent in traffic flow signals significantly impact the accuracy of the model.

B. Spatio-temporal correlation models

The temporal and spatial correlation of urban traffic flow is highly intricate [9]. Firstly, urban traffic exhibits dynamic and interdependent changes in both time and space. Secondly, different regions possess distinct types of temporal and spatial correlation, which evolve over time. Consequently, recent research efforts have been devoted to developing models capable of extracting the underlying temporal and spatial dependence relationships within traffic flow.

Currently, two primary categories of spatio-temporal correlation models exist: composite spatio-temporal models [1], [4]–[6] and fusion spatio-temporal models [17]–[20]. Composite spatio-temporal models capture temporal and spatial dependence relationships separately within traffic flow and subsequently combine them using a specific strategy for traffic flow prediction. For instance, Yu et al. [5] utilized two temporal convolutional blocks along with one spatial convolutional block to learn the dynamic behavior over time as well as the spatial patterns and features of traffic flow. Guo et al. [21] employed spatial attention mechanisms to establish correlations between different locations in terms of space while utilizing temporal attention mechanisms to capture dynamic correlations across various time periods. It is evident that the modeling of composite spatio-temporal relationships has not fully considered the joint interaction between space and time, leading to a disconnection between these two dimensions. While modeling traffic flow in terms of temporal or spatial correlation, the fusion of temporal-spatial models integrates the spatial-temporal characteristics of traffic flow to capture the interplay between space and time. For instance, Li et al. [18] employed a data-driven approach to construct a traffic flow time graph and combined it with a set of spatial graphs to form a spatio-temporal fusion graph, thereby uncovering hidden similarities and temporal-spatial dependence relationships within the traffic flow. Wang et al. [19] captured the spatial dependencies within the traffic flow using a bidirectional message passing mechanism and

explored high-order temporal dimensions by employing a gate recurrent unit (GRU) to capture complex nonlinear correlations within the traffic flow. Jin et al. [20] introduced an automatic expansion module called Auto-DSTSG that combines spatial graph neural networks with extended temporal convolutional networks, enabling capturing various types of temporal-spatial dependence relationships at different levels through stacking multiple modules.

Despite this, most current spatio-temporal correlation models fail to consider the directionality of the traffic network when simulating spatial dependencies, and inadequately address the trade-off between global and local spatio-temporal dependencies.

C. Graph neural network

Graph Neural Networks (GNNs) is a deep learning framework that enables direct learning from graph structured data. The fundamental concept underlying GNNs involves aggregating node features with those of their neighboring nodes. Previous research has demonstrated the remarkable capabilities of graph neural networks in learning from graph structured data [3]. Consequently, they have found extensive applications in traffic prediction to capture the latent spatial dependencies within traffic data [22], [23].

In recent years, prominent graph neural networks include spectral graph convolutional networks (SGCN), spatial graph convolutional networks (SGCN), and graph attention networks (GAT). Spectral graph convolutional networks primarily transform spatial graph signals into the spectral domain through Fourier transformation and perform convolution calculations in this domain [24]. However, their main limitation lies in their reliance on the Laplacian matrix of the graph, which restricts their applicability to scenarios with fixed structures. To address this dependency issue, Kipf et al. [22] simplified the graph convolution operation by employing message passing in the spatial domain and enhanced the performance of spatial graph convolutional networks by utilizing Chebyshev expansion to reduce Laplacian calculation complexity. However, due to the constraint imposed by the Laplacian matrix, spatial graph convolutional networks are only suitable for processing undirected graph-structured data. To consider the significance of learning spatial dependencies among neighboring nodes, Velickovic et al. [25] introduced a pioneering approach known as graph attention neural network (GAT). This network effectively incorporates feature information from neighboring nodes by calculating attention scores between nodes and their neighbors to determine edge weights between them. However, when dealing with sparse graph structures akin to traffic data, the feature learning capability of the GAT network is comparatively inferior to that of spatial graph convolutional neural networks.

Given the limitations inherent in graph convolutional neural networks and the deficiencies observed in graph attention neural networks, there is still a need to explore network model structures that are adept at efficiently extracting spatio-temporal features from traffic data.

III. PRELIMINARY

This section will provide a concise examination of potential temporal and spatial correlations in urban road traffic flow, while formally presenting and elucidating the research problem.

A. Spatio-temporal correlation analysis of traffic flow

In the urban traffic network, vehicles traverse interconnected roads, giving rise to intricate interdependencies between time and space. As depicted in Figure 1, a schematic representation of a local spatial unit within the traffic network, the red nodes denote the current intersection (referred to as v_0), while the gray and blue nodes represent downstream intersections that are directly or indirectly linked to v_0 . We will employ this diagram as an illustrative example for investigating potential temporal and spatial correlations in urban traffic flow.

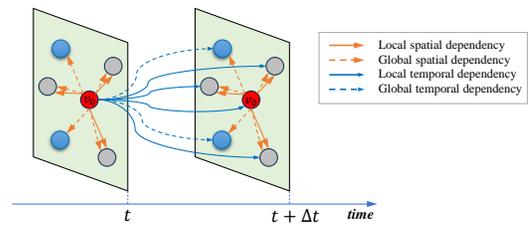


Fig. 1. Schematic diagram of traffic flow's spatio-temporal correlation.

Firstly, due to the topological structure of the traffic network, the flow of vehicles at time t from the current intersection v_0 is constrained to a limited range of downstream intersections that are directly connected to it. Assuming that after Δt_i time units, a portion of vehicle flow enters a directly connected downstream intersection v_i ; among these flows, both the amount entering intersection v_i and the corresponding travel time Δt_i are intricately linked with traffic conditions and network structure near v_0 . We refer to this phenomenon as the local spatial-temporal relationship of traffic flow, which arises from being confined by local spatial structure. Furthermore, within the entire transportation network, although certain intersections may not have direct connections, they can serve as concentrated areas encompassing traffic hubs, businesses, or office buildings. As a result, a spatial-temporal correlation relationship exists among these intersections that is not constrained by temporal duration or spatial extent. We refer to this type of correlation relationship as global spatial-temporal correlation.

B. Problem formulation

We utilize a directed graph $G = (V, E)$ to depict the topological structure of the urban road network. In this model, $V = \{v_1, v_2, \dots, v_N\}$ represents the set of intersections, and $E = \{e_1, e_2, \dots, e_M\}$ represents the set of road segments. Matrix A is employed to denote the adjacency relationships between intersections in G . If there exists an edge $e = \langle v_i, v_j \rangle \in E$, then $A_{ij} = 1$; otherwise, $A_{ij} = 0$. Assuming $X^t \in R^N$ signifies the traffic flow at each intersection in

G at time t , $X = \{X^{t-F+1}, X^{t-F+2}, \dots, X^t\}$ denotes the collection of traffic flow observations obtained by continuous monitoring of road network G for F consecutive time units. The objective of this paper is to establish a temporal and spatial dependency extraction model capable of learning traffic flow evolution from historical traffic flow set X and acquiring a mapping function $\psi(\cdot)$ (as depicted in Equation 1) that predicts traffic flow at each intersection in G for H subsequent time steps based on X .

$$[X^{t-F+1}, X^{t-F+2}, \dots, X^t] \xrightarrow{\psi(\cdot)} [Y^{t+1}, Y^{t+2}, \dots, Y^{t+H}] \quad (1)$$

Specifically, $X^{t-i} \in R^N$ represents the observed traffic flow at each intersection in G at the time $t - i$, and $Y^{t+j} \in R^N$ represents the predicted traffic flow at each intersection in G for the future time step j .

IV. TRAFFIC FLOW PREDICTION BASED ON MSTCM

This section will investigate the intricate temporal and spatial correlations in urban traffic flows, examine multi-dimensional temporal and spatial representation methods for traffic flow, develop algorithms for extracting temporal and spatial dependency relationships, and construct a traffic flow prediction model based on multi-perspective temporal and spatial correlation models.

A. Model Framework

Considering the intricate temporal and spatial dependencies in traffic flow within urban road network G , we commence by examining the spatial and temporal characteristics of traffic networks to comprehensively analyze and describe the multidimensional spatio-temporal features of traffic flow. Building upon this analysis, we propose a novel Multiple Spatio-Temporal Correlation Model (MSTCM) to effectively capture and predict future evolution trends by incorporating both spatial and temporal aspects.

The overall framework of the traffic flow prediction model based on multi-temporal and spatial correlation model is illustrated in Figure 2, comprising three main modules: (a) temporal and spatial representation, (b) temporal and spatial feature extraction, and (c) feature fusion and prediction. Given the temporal and spatial features of urban traffic flow (X, A) , where $X \in R^{N \times F}$ represents the traffic tensor of the road network that describes the temporal features of traffic flow; $A \in R^{N \times N}$ represents the connectivity relationship between intersections in the traffic network G that describes the spatial features of urban traffic flow. Firstly, within the temporal and spatial representation module, tensor X is decomposed into tensors X_u and X to respectively represent trend features and fluctuation features. Additionally, a first-order adjacent matrix A_f is generated based on spatial feature A to express connectivity relationships between intersections in G . Furthermore, second-order in-degree matrix A_{sin} as well as second-order out-degree matrix A_{sout} are generated to depict road direction characteristics within G . Subsequently, within the temporal and spatial feature extraction module, internal

dependence relationships among different time-space combinations are extracted resulting in a comprehensive representation of traffic flow features encompassing various temporal-spatial dependencies. Finally, within the feature fusion and prediction module, traffic flow features from different time-space dimensions are fused together for future traffic flow prediction.

B. Spatio-temporal representation

(1) **Temporal feature representation.** In urban road traffic, the behavior of traffic participants plays a crucial role in the evolution of traffic flow. However, external factors such as weather and traffic events interfere with the behavior of traffic participants, resulting in nonlinear and non-stationary characteristics of traffic flow. This increases the complexity of learning the temporal evolution features of traffic flow. To accurately capture both subjective behavior and external disturbance features inherent in traffic flow, we employ the damage-free seasonal trend decomposition algorithm [26] to decompose the traffic flow signal into a trend signal representing subjective behavioral features and a fluctuation signal representing external event disturbance features.

Considering the road network's traffic flow signal $X \in R^{N \times F}$ and the sliding window size w , this study employs the sliding average strategy described in Equation 2 to calculate the trend signal in traffic flow.

$$X_{u,t}^i = \frac{\xi_{t-\lceil w/2 \rceil}^i + \dots + \xi_{t+\lceil w/2 \rceil}^i}{w}, \text{ where } \xi_s^i = \begin{cases} X_s^i & \text{if } 1 \leq s \leq F \\ 0 & \text{otherwise} \end{cases} \quad (2)$$

Where $X_{u,t}^i$ represents the trend feature of intersection i at time t in graph G , which denotes the average traffic flow within a time window of size w centered at t . If the time window range exceeds the input traffic flow time interval, then zeros are padded to obtain $X_u \in R^{N \times F}$, representing the trend signal of traffic flow in graph G . Let $X = X - X_u$; hence, $X \in R^{N \times F}$ signifies the fluctuation feature signal of traffic flow.

(2) **Spatial feature representation.** Spatial graph convolutional neural networks (GCNs) are widely employed for analyzing graph-structured data due to their robust feature aggregation capabilities. However, the current approach is limited to processing undirected graph structures with semi-definite Laplacian matrices. In urban road traffic networks, which exhibit a directed spatial graph topology, direct application of spatial GCNs for related data analysis is not feasible. Existing traffic flow prediction methods based on graph convolutional neural networks often consider traffic networks as undirected graphs, leading to the loss of directional features in traffic flow and reduced prediction accuracy. To tackle this issue, we employ the approach proposed in [27], which decomposes the directed graph structure of the urban road network into multiple spatial structures exhibiting adjacency relationships consistent with semi-definite Laplacian matrices. These structures are utilized to depict proximity relationships and directional information between nodes.

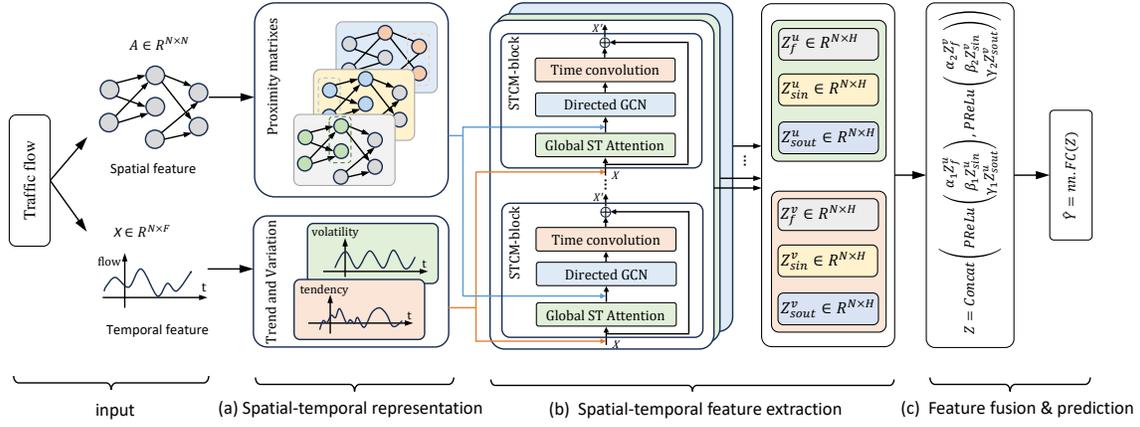


Fig. 2. Schematic diagram of traffic flow's spatio-temporal correlation.

Based on the adjacency matrix A of the given urban road traffic network G , utilizing formula 3, we can derive the first-order proximity matrix A_f that characterizes the neighboring relationships among intersections in G , as well as the second-order in-degree proximity matrix A_{sin} and out-degree proximity matrix A_{sout} that describe the upstream and downstream directions of intersections, respectively.

$$\begin{cases} A_f(i, j) &= A^{sym}(i, j) \\ A_{sin}(i, j) &= \sum_k \frac{A_{ki} A_{kj}}{\sum_v A_{kv}} \\ A_{sout}(i, j) &= \sum_k \frac{A_{ik} A_{jk}}{\sum_v A_{vk}} \end{cases} \quad (3)$$

Specifically, A^{sym} represents the symmetric matrix derived from the adjacency matrix A , which characterizes the direct spatial correlation relationships among intersections in the road network. As demonstrated in [27], all matrices A_f , A_{sin} , and A_{sout} exhibit positive semidefinite properties.

C. Spatio-temporal correlation extraction and feature fusion

Given a pair of temporal-spatial feature of traffic flow (X, A) , where $X \in \{X_u, X\}$ and $A \in \{A_f, A_{sin}, A_{sout}\}$, we extract both global and local temporal-spatial correlation features from the traffic flow data within the temporal-spatial framework (X, A) . Subsequently, these correlation features are sequentially integrated with the original traffic flow features

(1) Global spatio-temporal correlation extraction and feature fusion.

The global spatio-temporal correlation relationship between traffic flows in the road network describes the correlation between traffic flows on different roads in terms of both time and space, regardless of the road network topology. In order to learn this relationship, we employ an attention mechanism. To facilitate explanation, we initially extend the temporal feature of traffic flow to $X \in R^{N \times C \times F}$, where N represents the number of intersections in the road network, C represents the number of observed feature channels, and F represents the length of the observed traffic flow sequence. Initially, there is only one channel for the traffic flow temporal feature ($C = 1$).

By utilizing Equation 4, we can compute the global temporal attention coefficient AT for traffic flow.

$$AT = LeakyReLU((XW_1W_2)(XW_1W_2)^T + b_t) \quad (4)$$

Specifically, $W_1 \in R^C$, $W_2 \in R^{N \times N'}$, and $b_t \in R^{F \times F}$ are the learnable parameters, while $LeakyReLU(\cdot)$ serves as the activation function, N' represents the length of the expanded space feature dimension. To mitigate the differential impact of time feature scales on our model, we normalize the time attention coefficient AT using Equation 5, resulting in standardized time attention $AT' \in R^{F \times F}$.

$$AT'_{ij} = \text{softmax}_j(AT_{ij}) = \frac{\exp(AT_{ij})}{\sum_{k=1}^N \exp(AT_{ik})} \quad (5)$$

Subsequently, through the transformation of $X' = XAT'$, we integrate the global time-correlated features with the feature representation of traffic flow, thereby acquiring the fused representation X' for traffic flow.

Building upon this, we further extract the global spatial correlation relationships within traffic flow. Subsequently, the spatial attention coefficient AS in traffic flow can be computed using formula 6.

$$AS = LeakyReLU((X'U_1U_2)(X'U_1U_2)^T + b_s) \quad (6)$$

Specifically, $U_1 \in R^C$, $U_2 \in R^{F \times F'}$, and $b_s \in R^{N \times N}$ represent the learnable parameters. The activation function $LeakyReLU(\cdot)$ is utilized, while F' denotes the length of the traffic flow temporal feature dimension after expansion. Subsequently, Equation 7 is employed to normalize the spatial attention coefficient AS in order to mitigate the differential impact caused by variations in spatial feature scales within road traffic flow.

$$AS'_{ij} = \text{softmax}_j(AS_{ij}) = \frac{\exp(AS_{ij})}{\sum_{k=1}^N \exp(AS_{ik})} \quad (7)$$

The global spatio-temporal correlation features are finally integrated with the traffic flow feature representation X' to

obtain X'' , denoted as $X'' = AS'X' = AS'XAT'$, thereby yielding a comprehensive traffic flow representation $X'' \in R^{N \times C \times F}$ that encompasses both global temporal and spatial correlation features of traffic flow.

(2) Local spatio-temporal correlation extraction and feature fusion.

The local temporal correlation relationship quantifies the proportion of vehicle flows exiting one intersection and entering the directly connected downstream intersection over time. After obtaining the traffic flow temporal feature representation X'' , which integrates global temporal correlation features, we further investigate the local temporal correlation relationship of road network traffic flows by incorporating the spatial feature representation A of the road network.

Given a spatial representation $A \in \{A_f, A_{sin}, A_{sout}\}$ of the road network G , it is known that A is a positive semi-definite matrix. Therefore, we can obtain the Laplacian matrix $L_A = I - D^{-1/2}AD^{-1/2}$, where $D \in R^{N \times N}$ represents the degree matrix of matrix A . The eigenvalue decomposition of matrix L_A can be expressed as $L_A = U\Lambda U^T$, where $\Lambda = \text{diag}([\Lambda_0, \dots, \Lambda_{N-1}]) \in R^{N \times N}$ denotes the diagonal matrix containing the eigenvalues of L_A . Let $Y = X''$, then the graph convolution operation on the road network G with respect to the spatial representation A can be expressed as follows in equation 8:

$$g_{\theta} *_{G_A} Y = g_{\theta}(L_A)Y = g_{\theta}(U\Lambda U^T)Y = U g_{\theta}(\Lambda)U^T Y \quad (8)$$

Here, $*_{G_A}$ denotes the spatial feature A of the graph convolution operation on road network G , g represents a learnable parameter and $g(\Lambda)$ signifies the polynomial of eigenvalues derived from the Laplacian matrix L_A . To mitigate computational complexity, this study approximates $g(\Lambda)$ using Chebyshev polynomial coefficients up to degree K , as illustrated in Equation 9.

$$g_{\theta} *_{G_A} Y = U \sum_{k=0}^K \theta_k T_k(\Lambda)U^T Y = \sum_{k=0}^K \theta_k T_k(L_A)Y \quad (9)$$

Where the recurrence relation $T_k(x) = 2xT_{k-1}(x) - T_{k-2}(x)$ is employed, with initial conditions $T_0(x) = 1$ and $T_1(x) = x$. To enhance the model's learning and expressive capabilities, we introduce the active function $LeakyReLU(\cdot)$ for nonlinear processing of the output from the convolutional neural network. We define $Y' = LeakyReLU(g_{\theta} *_{G_A} Y)$, which represents the integrated traffic flow feature representation incorporating local spatial correlation relationship features.

Subsequently, we will conduct further investigation into the local temporal correlation relationships in traffic flow. By acquiring a time weight coefficient vector $E \in R^{F \times m}$, we will aggregate the traffic flow features at any given time with the features within a window of size m centered around that specific time point. Here, F represents the length of the traffic flow time feature. Ultimately, by leveraging Equation 10, we will integrate these local temporal correlation features into the road network traffic flow Y' , thereby obtaining a fused

representation Y'' that incorporates both global and local spatio-temporal correlations."

$$Y''_{ij} = \sum_{k=0}^{m-1} \tilde{Y}_{ij} E_{jk}, \text{ where } \tilde{Y}_{ij} = \begin{cases} Y'_{ij} & \text{if } \lceil \frac{m}{2} \rceil - t \leq j \leq F + \lceil \frac{m}{2} \rceil - t \\ 0 & \text{otherwise} \end{cases} \quad (10)$$

Therefore, we can represent the multi-temporal and spatial correlation features of traffic flow as a function of the temporal feature X , spatial feature A , and the learnable parameters $\Theta = \{g, W_1, W_2, U_1, U_2, E, b_t, b_s\}$, as depicted in Equation 11.

$$Z_{X,A} = \Phi(X, A, \Theta) \quad (11)$$

Here, $Z_{X,A}$ denotes a comprehensive traffic flow representation that integrates both global and local spatio-temporal correlation features within the spatial and temporal context (X, A) . For clarity purposes, we will henceforth refer to $Z_{X,A}$ as Z .

D. Deep spatio-temporal correlation extraction

According to the dynamic nature of traffic flow, the temporal and spatial correlation of traffic flow will also propagate across time and space domains. Therefore, we employ stacked modules for extracting higher-order, non-linear temporal and spatial correlations. Additionally, in order to mitigate the issue of gradient vanishing during multi-layer network propagation, we introduce a residual network as depicted in Equation 12. We denote the obtained traffic flow feature representation after the l^{th} layer of feature extraction and fusion as $Z^{(l)}$.

$$Z^{(l)} = \Phi(Z^{(l-1)}, A, \Theta^{(l)}) + Z^{(l-1)} \quad (12)$$

Let $Z^{(0)} = X$, by substituting $Z^{(l-1)} = \Phi(Z^{(l-2)}, A, \Theta^{(l-1)})$ into equation 12, we obtain

$$\begin{aligned} Z^{(l)} &= \Phi(\Phi(Z^{(l-2)}, A, \Theta^{(l-1)}), A, \Theta^{(l)}) + \Phi(Z^{(l-2)}, A, \Theta^{(l-1)}) \\ &= \underbrace{\Phi(\dots \Phi(\Phi(X, A, \Theta^{(1)}), \dots), A, \Theta^{(l)})}_{l} + \dots + \Phi(X, A, \Theta^{(1)}) + X \end{aligned}$$

Clearly, when $Z^{(0)} = X$ is evident. The equation 12 can be simplified to the subsequent equation 13 if we briefly denote $\Phi(\dots \Phi(\Phi(X, A, \Theta^{(1)}), \dots), A, \Theta^{(l)})$ as $\Theta^{(l)}(X, A, \Theta^{(0)} \dots \Theta^{(l)})$.

$$Z^{(l)} = \sum_{i=0}^l \Phi^{(i)}(X, A, \Theta^{(0)} \dots \Theta^{(i)}) \quad (13)$$

E. Traffic flow prediction based on MSTCM

By considering the two types of temporal features $\{X_u, X_v\}$ and the three types of spatial features $\{A_f, A_{sin}, A_{sout}\}$, we can individually select one temporal feature and one spatial feature to form six distinct combinations of space-time features. Subsequently, by leveraging these six combinations, we are able to learn and derive fusion features for traffic flow as expressed in Equation 14.

TABLE I
DESCRIPTION OF THE EXPERIMENTAL DATASET

Datasets	Vertices	Edges	Timestamps	Time Range
PEMS03	358	547	26208	9/1/2018-11/30/2018
PEMS04	307	340	16992	1/1/2018-2/28/2018
PEMS07	883	866	28224	5/1/2017-8/31/2017
PEMS08	170	295	17856	7/1/2016-8/31/2016

$$\begin{cases}
 Z_{u,f}^{(l)} = \sum_{i=0}^l \Phi_{u,f}^{(i)}(X_u, A_f, \Theta_{u,f}^{(0)} \dots \Theta_{u,f}^{(i)}) \\
 Z_{u,sin}^{(l)} = \sum_{i=0}^l \Phi_{u,sin}^{(i)}(X_u, A_{sin}, \Theta_{u,sin}^{(0)} \dots \Theta_{u,sin}^{(i)}) \\
 Z_{u,sout}^{(l)} = \sum_{i=0}^l \Phi_{u,sout}^{(i)}(X_u, A_{sout}, \Theta_{u,sout}^{(0)} \dots \Theta_{u,sout}^{(i)}) \\
 Z_{v,f}^{(l)} = \sum_{i=0}^l \Phi_{v,f}^{(i)}(X_v, A_f, \Theta_{v,f}^{(0)} \dots \Theta_{v,f}^{(i)}) \\
 Z_{v,sin}^{(l)} = \sum_{i=0}^l \Phi_{v,sin}^{(i)}(X_v, A_{sin}, \Theta_{v,sin}^{(0)} \dots \Theta_{v,sin}^{(i)}) \\
 Z_{v,sout}^{(l)} = \sum_{i=0}^l \Phi_{v,sout}^{(i)}(X_v, A_{sout}, \Theta_{v,sout}^{(0)} \dots \Theta_{v,sout}^{(i)})
 \end{cases} \quad (14)$$

In order to accurately depict the disparities in the impact of temporal-spatial traffic flow fusion features on the future evolution trend of traffic flow, we employ learnable weight coefficients to assign weights to each temporal-spatial combination's fusion features and subsequently concatenate them along the time dimension. Ultimately, we acquire the feature representation $Z^{(l)} \in R^{N \times 6F}$ for the multi-temporal and spatially correlated traffic flow through activation function $\text{PReLU}(\cdot)$, as demonstrated in Equation 15.

$$Z^{(l)} = \text{PReLU} \left(\text{Concat} \left(\begin{matrix} \alpha_f^1 Z_{u,f}^{(l)}, & \beta_{sin}^1 Z_{u,sin}^{(l)}, & \gamma_{sout}^1 Z_{u,sout}^{(l)} \\ \alpha_f^2 Z_{v,f}^{(l)}, & \beta_{sin}^2 Z_{v,sin}^{(l)}, & \gamma_{sout}^2 Z_{v,sout}^{(l)} \end{matrix} \right) \right) \quad (15)$$

Ultimately, a fully connected network is employed to learn the weight matrix $W \in R^{6F \times H}$, as depicted in Equation 16, which facilitates mapping of the fused features $Z^{(l)}$ from multiple temporal and spatial domains to predict traffic flow $\hat{X} \in R^{N \times H}$ for the subsequent H time steps.

$$\hat{X} = Z^{(l)} W \quad (16)$$

By utilizing the loss function $f_L(\cdot)$, we accomplish optimization objective as demonstrated in Equation 17.

$$\min_{\Theta} L_p(\Theta) = \frac{1}{N} \frac{1}{H} \sum_{n=1}^N \sum_{t=1}^H f_L(X_n^t, \hat{X}_n^t) \quad (17)$$

Where Θ represents the set of trainable parameters within the network model, Θ denotes the number of intersections in the road network, and H signifies the predicted traffic flow length.

V. EXPERIMENTS

This section will conduct comprehensive experiments on four real-world datasets to assess the efficacy of the proposed model in this paper and compare and analyze it with several existing baseline algorithms.

A. Experimental Environment and Data

The experimental procedures were conducted using Python 3.10 and the deep learning framework Pytorch 2.1.0, on a Linux server cluster equipped with an NVIDIA-A800 GPU. Model performance was evaluated using four real-world datasets and compared against the latest research findings. All experimental datasets were obtained from the California Performance Measurement System (PEMS) [28], with detailed descriptions provided in Table I.

B. Evaluation metrics and baseline models

In the analysis of the experimental results, we employed three indicator functions, namely Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and Mean Absolute Percentage Error (MAPE), to assess the model's accuracy. We compared our method with twelve significant baseline methods from the past five years. These baseline methods primarily encompass four categories: graph neural network-based approaches such as STSGCN [17], LSGCN [29], AutoSTG [30], STFGNN [18], STGODE [28], AutoDSTSGN [20], STG-NCDE [31], and ASTGCN [6]; attention mechanism-based techniques like PDFormer [32]; methods that combine graph neural networks and attention mechanisms, such as STJGCN [33] and ST-Ware [34]; as well as spatio-temporal data pre-modeling and processing methods like STD-MAE [35].

C. Model Performance Analysis

To comprehensively investigate the impact of various parameters on the model's accuracy, we conducted a series of experiments on the PEMS08 dataset to explore and validate the accuracy associated with important parameters. Throughout the experiment, we randomly partitioned all data into three distinct datasets - training, testing, and validation - in a ratio of 6:2:2. Simultaneously, several crucial model parameters were configured as follows: setting input and output lengths to 12 hours; utilizing a sliding window size of 15 for traffic flow time feature decomposition; employing a temporal and spatial convolutional kernel size of 64; selecting Chebyshev polynomial order as 3; constructing a two-layer spatial-temporal relationship extraction network. Furthermore, to enhance adaptability, we randomly divided the training data into multiple batches and trained the model using batch processing with a default batch size of 32 samples.

(1) In the MSTCM model, we decompose non-linear traffic flow into relatively simple trend and fluctuation signals and independently learn their spatial and temporal features to enhance prediction accuracy. Firstly, we investigate the influence of the sliding time window size (Win) on model accuracy through a series of experiments. The experimental results are depicted in Figure3(a), where initially there is a significant increase in model accuracy as Win increases, followed by a plateau at $Win=15$ which represents the peak performance. Subsequently, with further increase in Win , there is a slight decrease in model accuracy. As discussed in Section 4.2 analysis, the decomposition of traffic flow temporal features aims to separate stable characteristics determined by traffic

demand from random fluctuations caused by sudden factors. Among these components, the trend signal captures stable traffic flow features while the fluctuation signal represents changes induced by sudden factors. Increasing the sliding window size gradually aligns the trend signal obtained during decomposition with stable characteristics determined by traffic demand leading to improved model accuracy. However, excessive enlargement of window size may mistakenly identify small-scale changes in traffic demand as random fluctuations thereby reducing overall prediction accuracy.

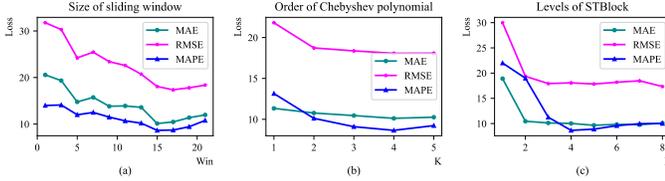


Fig. 3. Model's accuracy influenced by the parameters Win , K and L .

(2) To reduce computational complexity, we employ the K th-order Chebyshev polynomial approximation to represent the characteristic polynomial of the Laplacian matrix when investigating local spatial correlation relationships in traffic flow analysis. Generally, a higher order Chebyshev polynomial enables more accurate capture of spatial correlation relationships, thereby enhancing model accuracy. Consequently, we conducted a series of experiments to examine the impact of Chebyshev polynomial order on model accuracy. The experimental results are depicted in Figure 3(b), illustrating that as K increases, model accuracy initially improves rapidly and then stabilizes; beyond $K = 4$, there is negligible change in model accuracy. The main reason for this is that as K increases, the additional high-order coefficients that are added bring an increasingly small increase in expressive ability.

(3) To enhance the MSTCM's capacity for learning temporal-spatial features of traffic flow, we employ a multi-layer temporal-spatial correlation extraction module to acquire higher-dimensional non-linear features. In order to investigate the impact of the number of layers in the temporal-spatial relationship extraction network on model accuracy, a series of experiments were conducted. The experimental results are depicted in Figure 3(c), where initially, as the number of network layers L increases, there is a rapid improvement in model accuracy; subsequently, it stabilizes and slightly decreases with further increase in network layers. This behavior can be attributed to an expanded receptive field resulting from increased network layers that aids in capturing higher-dimensional non-linear features; however, excessive layering may lead to overfitting issues. Consequently, augmenting the number of network layers within a narrower range yields benefits that outweigh costs and enhances model accuracy; nevertheless, beyond this range, diminishing returns occur along with escalating costs leading to stable or even reduced model accuracy.

(4) Subsequently, a series of experiments will be conducted

to analyze and validate the impact of traffic flow time feature decomposition on model accuracy. Initially, a sliding window ($Win=15$) is employed to decompose traffic flow into two features: *Trend* and *Variation*. One or both of these features are then selected to learn the evolution patterns of traffic flow and predict future conditions. Simultaneously, a comparison will be made between experimental results and predictions using the original traffic flow data (*Origin*). As depicted in Figure 4(a), combining the trend and variation features yields the highest accuracy in traffic flow prediction; utilizing only the trend feature achieves second-best accuracy but significantly outperforms predictions based on original traffic flow data; employing solely the variation feature yields inferior results, much lower than those obtained from predicting with original traffic signals. These experimental findings further confirm two key points: firstly, in representing the evolution process of traffic flow, trend serves as the most stable and crucial feature; secondly, incorporating disturbance information renders future traffic flow prediction more challenging. This study has significant implications for enhancing predictive accuracy regarding future traffic conditions by extracting trend and variation features from traffic flows while separately learning their evolutionary laws.

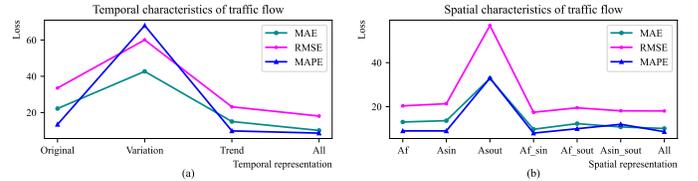


Fig. 4. Model's accuracy influenced by spatio-temporal representation of traffic flow.

(5) In the MSTCM model, to capture more intricate spatial correlations, we decompose the spatial features of traffic flow into a first-order adjacency matrix A_f representing proximity relationships, and second-order in-degree adjacency matrix A_{sin} and out-degree adjacency matrix A_{sout} representing directional information. Consequently, in our experiments, we utilize one or more of these aforementioned three spatial features to learn the correlation patterns of traffic flow and forecast future traffic conditions. Simultaneously, we analyze prediction outcomes to evaluate the impact of decomposing traffic flow's spatial features on model accuracy. As illustrated in Figure 4(b), when employing only one spatial feature, both A_f and A_{sin} demonstrate significantly higher accuracy compared to A_{sout} ; indicating that A_f and A_{sin} encapsulate richer information regarding spatial characteristics than does A_{sout} . When incorporating two different types of spatial features simultaneously, all combinations yield higher accuracy than using a single type alone; among all tested combinations, the combination of A_f and A_{sin} slightly outperforms others. Finally, by integrating all three distinct types of spatial features simultaneously, our model achieves its highest level of accuracy. These experimental findings substantiate that these

TABLE II
PERFORMANCE COMPARISON OF DIFFERENT APPROACHES FOR TRAFFIC FLOW FORECASTING

	PEMS03			PEMS04			PEMS07			PEMS08		
	RMSE	MAE	MAPE(%)	RMSE	MAE	MAPE(%)	RMSE	MAE	MAPE(%)	RMSE	MAE	MAPE(%)
STSGCN (2020)	29.21	17.48	16.78	33.65	21.19	13.90	39.03	24.26	10.21	26.80	17.13	10.96
LSGCN (2020)	29.85	17.94	16.98	33.86	21.53	13.18	41.46	27.31	11.98	26.76	17.73	11.20
AutoSTG (2021)	27.63	16.27	16.10	32.51	20.38	14.12	36.47	23.22	9.95	25.46	16.37	10.36
STFGNN (2021)	28.34	16.77	16.30	32.51	20.48	16.77	36.60	23.46	9.21	26.25	16.94	10.60
STGODE (2021)	27.84	16.50	16.69	32.82	20.84	13.77	37.54	22.99	10.14	25.97	16.81	10.62
Auto-DSTSGN (2022)	25.17	14.59	14.22	30.48	18.85	13.21	33.02	20.08	8.57	23.76	14.74	9.45
STG-NCDE (2022)	27.09	15.57	15.06	31.09	19.21	12.76	33.84	20.53	8.80	24.81	15.45	9.92
ST-aware (2022)	16.63	15.17	15.83	31.02	19.06	12.52	34.05	20.74	8.77	24.62	15.41	9.94
STGCN (2022)	29.56	17.34	17.21	35.22	22.93	16.56	37.87	21.01	10.73	28.06	18.25	11.64
PDFormer (2023)	25.39	14.94	15.82	29.97	18.32	12.10	32.87	19.83	8.53	23.51	13.58	9.05
STJGCN (2023)	25.70	14.92	14.81	30.35	18.81	11.92	33.01	19.95	8.31	23.74	14.53	9.15
STD-MAE (2024)	24.43	13.80	13.96	29.25	17.80	11.97	31.44	18.65	7.84	22.47	13.44	8.76
MSTCM (Ours)	14.79	11.94	14.17	22.81	12.48	12.08	28.57	17.68	8.27	18.07	10.12	8.65

*The reported results of the utilized baseline method are exclusively sourced from primary literature.

three distinct types of spatial features complement each other in extracting underlying correlations within traffic flow.

D. Performance comparison analysis

We conducted a series of experiments to compare the proposed MSTCM model with twelve baseline methods that have been used in the past five years. In these experiments, all methods were configured with input and output lengths set to 12, indicating the utilization of historical traffic data from the previous hour for predicting traffic flow in the subsequent hour. The prediction accuracy of various methods on four datasets is presented in Table 2. The experimental results demonstrate that the MSTCM model has achieved remarkable performance, significantly outperforming other indicators when compared to the baseline methods, except for MAPE values on the PEMS03 and PEMS07 datasets.

By comparing the key technologies and performance of each baseline method, we observed that the eight graph neural network-based baselines demonstrated similar overall performance on the four datasets, albeit with some variations. Notably, the PDFomer method incorporating an attention mechanism outperformed most graph neural network-based baselines and only slightly lagged behind the Auto-DSTHCN method. The two approaches for integrating attention mechanisms with graph neural networks exhibited comparable performance across all metrics, except for a significant disparity in the RMSE metric on the PEMS03 dataset. These methods outperformed the baseline graph neural network approach and yielded results comparable to those of the PDFomer method. The STD-MAE method presented a distinctive approach wherein a pre-trained model was designed to enhance traffic flow data quality before feeding it into downstream prediction models, which is fundamentally different from existing algorithms. It is worth mentioning that as a downstream prediction model for STD-MAE, GWNnet model [36] achieved superior accuracy compared to six baseline algorithms.

In response to the aforementioned experimental phenomena, we conducted an analysis of the underlying logic behind these

observations. Our findings suggest that graph neural networks and attention mechanisms emphasize different aspects when extracting spatial-temporal features of traffic flow. Specifically, graph neural networks primarily focus on aggregating local network’s spatial-temporal features; although they can extend feature aggregation through deep network layers, discovering global spatial-temporal features of the road network remains challenging. On the other hand, attention mechanisms excel at capturing global spatial-temporal features of the road network without being constrained by space and time but exhibit limitations in extracting local area’s spatial-temporal features. By combining both approaches in our MSTCM model, we leverage their respective strengths and compensate for their weaknesses to achieve superior performance. In this model, we not only employ attention mechanisms and graph neural networks to extract global and local spatial-temporal features of traffic flow but also decompose its spatiotemporal characteristics while extracting more refined and nonlinear perspectives from various spatiotemporal dimensions.

VI. CONCLUSION

Due to the combined influence of intrinsic traffic demand and external random factors, urban road traffic flow exhibits highly dynamic nonlinear and non-stationary characteristics, which pose a significant challenge for traffic flow prediction. To mitigate the complexity arising from temporal and spatial features of traffic flow and their adverse impact on prediction accuracy, this study employs a feature decomposition strategy. By decomposing traffic flow into multiple representations in different temporal and spatial domains, as well as utilizing multi-layer temporal relationship extraction modules to capture non-linear dependencies within each domain, we effectively predict the future evolution trend of road network traffic by weighting and fusing the temporal and spatial dependence features. We conducted extensive experiments on four real-world datasets to thoroughly analyze and evaluate our proposed model’s performance. Furthermore, we compared our method with twelve state-of-the-art approaches achieved over the past

five years. The comprehensive evaluation results demonstrate outstanding performance of our proposed method.

REFERENCES

- [1] C. Huang, W. Kong, G. Dai, and Y. Liu, "Ltphm: Long-term traffic prediction based on hybrid model," in *Proceedings of the 30th ACM International Conference on Information & Knowledge Management*, 2021, pp. 3093–3097.
- [2] J. Sun, J. Zhang, Q. Li, X. Yi, Y. Liang, and Y. Zheng, "Predicting citywide crowd flows in irregular regions using multi-view graph convolutional networks," *IEEE Transactions on Knowledge and Data Engineering*, vol. 34, no. 5, pp. 2348–2359, 2022.
- [3] X. Zhang, C. Huang, Y. Xu, and L. Xia, "Spatial-temporal convolutional graph attention networks for citywide traffic flow forecasting," in *Proceedings of the 29th ACM International Conference on Information & Knowledge Management*, ser. CIKM '20, New York, NY, USA, 2020, pp. 1853C–1862.
- [4] Y. Li, R. Yu, C. Shahabi, and Y. Liu, "Diffusion convolutional recurrent neural network: Data-driven traffic forecasting," in *International Conference on Learning Representations*, 2018, pp. 1–16.
- [5] B. Yu, H. Yin, and Z. Zhu, "Spatio-temporal graph convolutional networks: A deep learning framework for traffic forecasting," in *Proceedings of the Twenty-Seventh International Joint Conference on Artificial Intelligence*, 2018, pp. 3634C–3640.
- [6] 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, 2022.
- [7] L. Zhao, Y. Song, C. Zhang, Y. Liu, P. Wang, T. Lin, M. Deng, and H. Li, "T-gcn: A temporal graph convolutional network for traffic prediction," *IEEE Transactions on Intelligent Transportation Systems*, vol. 21, no. 9, pp. 3848–3858, 2020.
- [8] J. Ji, J. Wang, Z. Jiang, J. Jiang, and H. Zhang, "Stden: Towards physics-guided neural networks for traffic flow prediction," in *Proceedings of the 36th AAAI Conference on Artificial Intelligence*, ser. AAAI '22, 2022, pp. 4048–4056.
- [9] Z. Pan, W. Zhang, Y. Liang, W. Zhang, Y. Yu, J. Zhang, and Y. Zheng, "Spatio-temporal meta learning for urban traffic prediction," *IEEE Transactions on Knowledge and Data Engineering*, vol. 34, no. 3, pp. 1462–1476, 2022.
- [10] Y. Yu, X. Si, C. Hu, and J. Zhang, "A review of recurrent neural networks: Lstm cells and network architectures," *Neural Computation*, vol. 31, no. 7, pp. 1235–1270, 2019.
- [11] R. Dey and F. M. Salem, "Gate-variants of gated recurrent unit (gru) neural networks," in *2017 IEEE 60th International Midwest Symposium on Circuits and Systems (MWSCAS)*, 2017, pp. 1597–1600.
- [12] H. Yao, F. Wu, J. Ke, X. Tang, Y. Jia, S. Lu, P. Gong, J. Ye, D. Chuxing, and Z. Li, "Deep multi-view spatial-temporal network for taxi demand prediction," in *Proceedings of the 32th AAAI Conference on Artificial Intelligence and Thirtieth Innovative Applications of Artificial Intelligence Conference and Eighth AAAI Symposium on Educational Advances in Artificial Intelligence*, ser. AAAI'18. AAAI Press, 2018.
- [13] J. Gu, Z. Wang, J. Kuen, L. Ma, A. Shahroudy, B. Shuai, T. Liu, X. Wang, G. Wang, J. Cai, and T. Chen, "Recent advances in convolutional neural networks," *Pattern Recognition*, vol. 77, pp. 354–377, 2018.
- [14] B. Wang, Y. Lin, S. Guo, and H. Wan, "Gsnet: Learning spatial-temporal correlations from geographical and semantic aspects for traffic accident risk forecasting," 2021, pp. 4402–4409.
- [15] ZhengChuanpan, X. Fan, C. Wang, Qi, and Jianzhong, "Gman: A graph multi-attention network for traffic prediction," 2020, pp. 1234–1241.
- [16] G. Jin, Y. Liang, Y. Fang, Z. Shao, J. Huang, J. Zhang, and Y. Zheng, "Spatio-temporal graph neural networks for predictive learning in urban computing: A survey," *IEEE Transactions on Knowledge and Data Engineering*, pp. 1–20, 2023.
- [17] C. Song, Y. Lin, S. Guo, and H. Wan, "Spatial-temporal synchronous graph convolutional networks: A new framework for spatial-temporal network data forecasting," in *Proceedings of the AAAI conference on artificial intelligence*, vol. 34, no. 01, 2020, pp. 914–921.
- [18] M. Li and Z. Zhu, "Spatial-temporal fusion graph neural networks for traffic flow forecasting," in *Proceedings of the AAAI conference on artificial intelligence*, vol. 35, no. 5, 2021, pp. 4189–4196.
- [19] Y. Wang, J. Zheng, Y. Du, C. Huang, and P. Li, "Traffic-ggnn: Predicting traffic flow via attentional spatial-temporal gated graph neural networks," *IEEE Transactions on Intelligent Transportation Systems*, vol. 23, no. 10, pp. 18423–18432, 2022.
- [20] G. Jin, F. Li, J. Zhang, M. Wang, and J. Huang, "Automated dilated spatio-temporal synchronous graph modeling for traffic prediction," *IEEE Transactions on Intelligent Transportation Systems*, vol. 24, no. 8, pp. 8820–8830, 2023.
- [21] 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, 2022.
- [22] H. Zhang, G. Lu, M. Zhan, and B. Zhang, "Semi-supervised classification of graph convolutional networks with laplacian rank constraints," *Neural Process. Lett.*, vol. 54, no. 4, pp. 2645–2656, 2022.
- [23] K.-H. N. Bui, J. Cho, and H. Yi, "Spatial-temporal graph neural network for traffic forecasting: An overview and open research issues," *Applied Intelligence*, vol. 52, no. 8, pp. 2763–2774, 2021.
- [24] J. Zhou, G. Cui, S. Hu, Z. Zhang, C. Yang, Z. Liu, L. Wang, C. Li, and M. Sun, "Graph neural networks: A review of methods and applications," *AI open*, vol. 1, pp. 57–81, 2020.
- [25] P. Velickovic, G. Cucurull, A. Casanova, A. Romero, P. Liò, and Y. Bengio, "Graph attention networks," in *6th International Conference on Learning Representations, ICLR 2018, Vancouver, BC, Canada, April 30 - May 3, 2018, Conference Track Proceedings*, 2018, pp. 1–12.
- [26] X. He, Y. Li, J. Tan, B. Wu, and F. Li, "Oneshotstl: One-shot seasonal-trend decomposition for online time series anomaly detection and forecasting," *Proc. VLDB Endow.*, vol. 16, no. 6, pp. 1399–1412, feb 2023.
- [27] Z. Tong, Y. Liang, C. Sun, D. S. Rosenblum, and A. Lim, "Directed graph convolutional network," *ArXiv*, vol. abs/2004.13970, 2020. [Online]. Available: <https://api.semanticscholar.org/CorpusID:216642198>
- [28] 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*, ser. KDD '21, New York, NY, USA, 2021, pp. 364–373.
- [29] R. Huang, C. Huang, Y. Liu, G. Dai, and W. Kong, "Lsgcn: long short-term traffic prediction with graph convolutional networks," in *Proceedings of the Twenty-Ninth International Joint Conference on Artificial Intelligence*, ser. IJCAI'20, 2021, pp. 2355–2361.
- [30] Z. Pan, S. Ke, X. Yang, Y. Liang, Y. Yu, J. Zhang, and Y. Zheng, "Autostg: Neural architecture search for predictions of spatio-temporal graph?" in *Proceedings of the Web Conference 2021*, ser. WWW '21, New York, NY, USA, 2021, pp. 1846–1855.
- [31] J. Choi, H. Choi, J. Hwang, and N. Park, "Graph neural controlled differential equations for traffic forecasting," in *Thirty-Sixth AAAI Conference on Artificial Intelligence, AAAI 2022, February 22 - March 1, 2022*, 2022, pp. 6367–6374.
- [32] J. Jiang, C. Han, W. X. Zhao, and J. Wang, "Pdformer: Propagation delay-aware dynamic long-range transformer for traffic flow prediction," in *Thirty-Seventh AAAI Conference on Artificial Intelligence, AAAI 2023, Washington, DC, USA, February 7-14, 2023*. AAAI Press, 2023, pp. 4365–4373.
- [33] C. Zheng, X. Fan, S. Pan, H. Jin, Z. Peng, Z. Wu, C. Wang, and P. S. Yu, "Spatio-temporal joint graph convolutional networks for traffic forecasting," *IEEE Transactions on Knowledge and Data Engineering*, vol. 36, no. 1, pp. 372–385, 2024.
- [34] R.-G. Cirstea, B. Yang, C. Guo, T. Kieu, and S. Pan, "Towards spatio-temporal aware traffic time series forecasting," in *2022 IEEE 38th International Conference on Data Engineering (ICDE)*, 2022, pp. 2900–2913.
- [35] H. Gao, R. Jiang, Z. Dong, J. Deng, Y. Ma, and X. Song, "Spatial-temporal-decoupled masked pre-training for spatiotemporal forecasting," in *Proceedings of International Joint Conference on Artificial Intelligence, Jeju, August 03-09, 2024*, 2024.
- [36] Z. Wu, S. Pan, G. Long, J. Jiang, and C. Zhang, "Graph wavenet for deep spatial-temporal graph modeling," in *Proceedings of the 28th International Joint Conference on Artificial Intelligence*, ser. IJCAI'19, 2019, pp. 1907–1913.