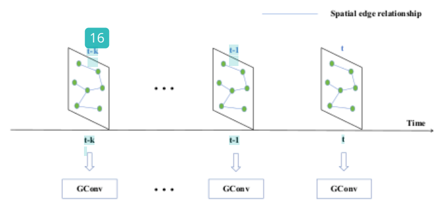
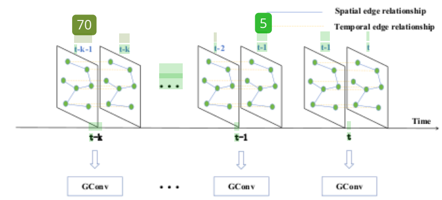
Trip-Aware Spatial-Temporal Graph Convolutional Network for Traffic Flow Forecasting

***Abstract*—Traffic flow forecasting is a necessary technology for realizing intelligent transportation systems (ITS). Accurate modeling of the spatial-temporal correlations inherent in traffic data is crucial for forecasting traffic flow exactly. Although joint spatial-temporal modeling approaches proposed in recent years have been validated to be effective in capturing the spatial-temporal heterogeneity that exists within traffic patterns, their feature aggregation mechanisms still rely on data-driven adaptive weight parameters and lack explicit modeling of traffic flow dynamics. Consequently, models relying on this approach struggle to learn spatial-temporal representations that conform to dynamic traffic flow patterns. To overcome these limitations, we offer an innovative traffic flow prediction methodology utilizing a Trip-Aware Spatial-Temporal Graph Convolutional Network (TASTGCN). This methodology incorporates inter-node distance and vehicle speed to construct a spatial-temporal feature aggregation mechanism that aligns with traffic flow dynamics. Furthermore, we develop two dual-path feature interaction network to capture long-term and short-term temporal dependencies in traffic flow time series data. Finally, we validate our approach through experiments on four real-world datasets, where the results clearly show that our model achieves superior predictive performance.**

***Index Terms*—Spatial-temporal correlations, spatial-temporal features aggregation mechanism, graph convolution networks, traffic flow forecasting.**



(a) GConv extracts spatial features with a single graph at each time step.



(b) GConv extracts spatial-temporal features with multiple graphs at each time step (this figure shows the synchronous processing of two graph data).

Fig. 1. Comparison of GCnov extracts features: (a) Spatial features extraction from single-graph data. (b) Spatial-temporal features extraction from multi-graph data.

I. Introduction

With the rapid socio-economic development and continuous urban expansion, traffic congestion has become a constraint on urban and economic growth. Therefore, achieving accurate predictions of future traffic flow is beneficial to alleviating congestion, optimizing traffic resource allocation, and improving road efficiency.

Various approaches have been developed in traffic flow prediction research to address its challenges during the past few decades. Early approaches mostly employed traditional time series and statistical models to emulate temporal trends in traffic data, such Historical Average (HA) [1] and Autoregressive Integrated Moving Average (ARIMA) [2], [3], [4], [5]. Despite their computational efficiency, these methods fall short in capturing the nonlinear dynamics intrinsic to traffic flow data. To overcome the shortcomings of traditional methods, machine learning techniques have been integrated into the domain of traffic prediction, such as Support Vector Regression (SVR) [6], [7], [8], [6] and Random Forest Regression (RFR) [9], [10], [11], which offering improved generalization and reduced prediction errors. However, these approaches frequently depend on manually designed input features and do not adequately represent the topological dependencies intrinsic to the transportation network topology.

Following advances in deep learning techniques, more sophisticated models have been explored to overcome the shortcomings of traditional machine learning approaches. Some studies have used Recurrent Neural Networks (RNN) [12], Long Short-Term Memory (LSTM) [13], [14], [15] and Gated Recurrent Units (GRU) [16], [17] to enhance the model's capacity to capture long-term temporal dependencies within traffic flow. Nevertheless, these temporal neural network models neglect the topology of the transportation network and fail to explicitly model its spatial structure. Given the effectiveness of Convolutional Neural Networks (CNN) in image-based learning tasks, they are applied to extract spatial dependencies in traffic flow [18], [19]. However, CNN's grid-based architecture is inherently unsuitable for modeling the complex and irregular topological relationships of the transportation network.

In recent years, Graph Convolutional Networks (GCN) [20] have exhibited a high degree of computational efficiency and accuracy in traffic flow prediction because of their capability to process data with non-Euclidean structures. This has led to a new paradigm that integrates GCN with temporal neural networks (e.g., RNN, CNN, LSTM) to accurately simulate the traffic flow's intrinsic spatial and temporal relationships, thereby significantly enhancing predictive accuracy [21], [22],

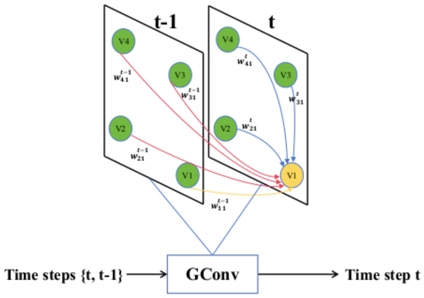


Fig. 2. The spatio-temporal features aggregation mechanism uses the weight parameter to aggregate graph information across time steps.

[23], [24], [25], [26]. Nevertheless, these methods overlook the spatial-temporal heterogeneity in traffic data and lack effective modeling of joint spatial-temporal features [27].

For the purpose of modeling the joint spatial-temporal features, some studies propose a joint spatial-temporal feature extraction method based on GCN, whose core operation is to construct local spatial-temporal graphs by connecting the individual spatial graph of neighboring time steps (as shown in Fig. 1b), and use graph convolution operation to capture the joint spatial-temporal features implied in local spatial-temporal graphs, which effectively avoids the segmentation of the spatial-temporal information [27], [28], [29], [30], [31]. However, the existing mechanism for aggregating spatial-temporal features has inherent flaws. In the process of aggregating the features of adjacent nodes across time steps to learn future features of the target node, the weight parameters used (as shown in Fig. 2) are updated based on data-driven model training and lack a clear constraint mechanism. Specifically, considering the impact of practical factors like the distance between nodes and the vehicle driving speed, the average time spent by vehicles moving from adjacency nodes to the target node is inconsistent, which results in the traffic flow features of each adjacency node under different time steps having different impact weights on the future traffic flow features of the target node. Weight values of existing studies are obtained by data-driven model self-training update. This unconstrained weight learning mechanism leads to the joint spatial-temporal features obtained by the model deviating from the actual traffic flow evolution patterns, and fails to effectively represent the real conditions of traffic flow changes.

Aiming at the problems in existing joint spatial-temporal feature extraction methods, we develop a traffic flow prediction network based on trip-aware spatial-temporal graph convolution. The network innovatively introduces two traffic features with clear physical significance, distances between nodes and vehicle driving speed, which are used to dynamically adjust the weights of adjacent nodes' traffic flow across time steps on the future traffic flow of the target node. This physics-constrained feature aggregation mechanism more accurately captures the propagation patterns of traffic flow across temporal and spatial dimensions, and significantly enhances the interpretability of the joint spatiotemporal feature extraction method. In addition, we design two dual-path spatial-temporal feature interaction networks to achieve more accurate time-dependent modeling and enhance the model prediction performance. In summary, our study makes the following key contributions:

- We propose an innovative trip-aware spatial-temporal graph convolution network (TASTGCN) with the capability to model spatial-temporal dependencies in traffic flow and improve prediction accuracy.
- We develop a novel joint spatial-temporal feature extraction method. The joint spatial-temporal features obtained by this method conform to the traffic flow evolution patterns and are more effective than previous methods.
- Conducting experiments on multiple traffic datasets confirms that TASTGCN achieves superior predictive accuracy.

The remainder of this paper is arranged as follows: Section II discusses existing methods for traffic flow prediction. Section III introduces the fundamental definitions and notations used in this work. Section IV describes the proposed TASTGCN framework along with its major components. Section V shows and discusses the experimental results. Section VI summarizes our work.

II. Related Works

Early traffic forecasting used classical statistical methods, representative of which are the Historical Average (HA) [1], Autoregressive Integrated Moving Average (ARIMA) [2], and Vector Autoregressive (VAR) [32]. These methods tend to capture only the average trends in traffic flow, often overlooking the peaks and rapid fluctuations, and thus have difficulty achieving reliable forecasting results under highly nonlinear and intricate data conditions.

To cope with the highly nonlinear problem of traffic flow data, K-Nearest Neighbor (KNN) [33], Support Vector Machine (SVM) [34], and Bayesian Networks (BN) [35], [36] have been applied to traffic flow forecasting. Although these methods have made some progress in dealing with nonlinear problems, they are still considered shallow learning models with limited feature extraction capabilities, and they also rely on domain expertise for the manual design of input features.

Advancements in deep learning have greatly propelled progress in traffic flow predictions. Early deep learning models for sequence modeling, such as Recurrent Neural Network (RNN) [12], Long Short-Term Memory (LSTM) [13], [14], and Gated Recurrent Unit (GRU) [16], [17] focused on capturing temporal correlations in traffic flow. However, these methods are unable to capture spatial correlations. To address this limitation, several studies have utilized Convolutional Neural Networks (CNN) to capture spatial correlations [18], [19]. In these works, CNN treats traffic data as Euclidean-structured inputs by dividing the traffic network into regular grids and predicting traffic conditions within each grid. However, CNN is suboptimal for learning spatial correlations in transportation networks because it is fundamentally designed to process

Euclidean spatial relationships, typically represented as two-dimensional matrices or raster images. This architectural limitation makes it ineffective for modeling the non-Euclidean structural characteristics of real-world road networks. [37].

Graph Convolutional Networks (GCN) can effectively handle non-Euclidean data. Although GCN was not originally designed for traffic flow forecasting, it demonstrates high efficiency and accuracy in this field because of the natural graph-structured characteristics inherent in traffic data. Spatial-Temporal Graph Convolutional Network (STGCN) [22] and Diffusion Convolutional Recurrent Neural Network (DCRNN) [21] were pioneers in integrating graph convolutional networks with temporal neural networks to model the spatial-temporal dependencies. They represent the traffic network as graphs and separately capture spatial-temporal features by using GCN and temporal neural networks. T-GCN [23] incorporates GRU and GCN to get the spatial topological dependence and the dynamic temporal dependence contained in data. Building on these previous studies, researchers have proposed replacing the conventional static adjacency matrix with a dynamically updated one to more accurately capture time-varying spatial dependencies among nodes, thereby improving the model's capability to adapt to changing patterns within a fixed transportation network structure. ASTGCN [24] utilizes spatial and temporal attention mechanisms to modify the impact of adjacent road nodes (spatial attention) and previous time steps (temporal attention). This design overcomes the limitations of using a fixed adjacency matrix in traditional methods. GWNet [25] captures spatial features by integrating GCN with an adaptive adjacency matrix, and captures temporal features using temporal convolutional networks (TCN). AGCRN [26] integrates two adaptive components: a module for graph generation and a node-specific parameter learning module, aimed at modeling dynamic spatial-temporal dependencies in traffic networks. STJGCN [38] generates dynamic adjacency matrices by merging spatial and temporal features encoded in spatio-temporal embeddings, thereby modeling dynamic temporal correlations in traffic data. STMPNet[39] proposes a multi-graph fusion module that integrates geographical adjacency graphs, dynamic spatial graphs, and static attribute graphs to respectively capture spatially varying temporal relationships, dynamic dependencies, and static attribute interactions.

The above studies [22], [21], [23], [24], [25], [26] use GCN to capture spatial correlations and temporal neural network models (e.g., RNN, CNN, TCN) to extract temporal dependencies. However, they ignore joint spatial-temporal features. In light of this limitation, some studies have proposed methods for synchronously capturing spatial-temporal features based on GCN [27], [28], [29], [30], [31]. [27] proposes a spatial-temporal synchronized graph convolution approach, which incorporates both temporal and spatial dependencies into localized spatiotemporal graphs and performs graph convolution operation, thereby avoiding the loss of joint spatial-temporal information. [28] constructs spatiotemporal graphs across different time periods and performs graph convolution, thus assisting nodes in acquiring a broader spatial-temporal receptive range. [29] proposes a spatiotemporal position relation inference module, which dynamically adjusts weight parameters between nodes by combining their spatial-temporal location embeddings, thus more accurately capturing complex spatial-temporal dependencies. [30] proposes a spatio-temporal heterogeneous graph convolution module, which performs graph convolution on local spatio-temporal graphs, and enhances the model's representation of spatio-temporal heterogeneity by screening key features through stacking multi-layer convolution and pruning operations. [31] proposes a method to reconstruct fusion maps offline via tensor decomposition (Tucker decomposition), which transforms the traditional binary adjacency matrix into tensor form, and captures potential spatio-temporal correlations using the decomposed core tensor.

TABLE I
NOTATIONS

Notation	Meaning
G	Graph-based representation of the transportation network
V	Nodes included in G
E	Edges included in G
$\mathbf{A}$	Adjacency matrix corresponding to the graph G
X_{flow}	Historical traffic observations fed into the model
X_h	Historical flows
X_p	Future flows
$\hat{X}_p$	Predicted future flows
D_{flow}	Dimension of traffic flow X_{flow}
D_{emb}	Dimension of spatial-temporal identify information
T_{in}	Temporal steps of the input data
T_{out}	Temporal steps of the output data
α	Weight mask matrix in time step t
β	Weight mask matrix in time step $t-1$

III. PRELIMNARIES

In this section, the traffic prediction problem is defined, and relevant concepts are presented. Key notations employed in the paper are represented in Table I.

A. Traffic Flow Prediction Problem Description

Traffic flow prediction entails forecasting future traffic flow at each node within the transportation network, utilizing historical traffic flow data from those nodes.

Given a transportation network $G = (V, E, \mathbf{A})$ and its historical traffic flow data detected by the sensors, denoted as $X_h \in R^{N \times H \times C}$. Traffic flow prediction aims to obtain an optimized function $f(\cdot)$ that generates predictions $\hat{X}_p \in R^{N \times P}$ by leveraging G and X_h. The corresponding mathematical formulation is:

$$\hat{X}_p = f(X_h, G; \theta) \tag{1}$$

where V denotes the set of traffic road segments equipped with sensors, E denotes the set of edges that define the connection relationship of traffic road segments. $\mathbf{A}$ is an adjacency matrix that encodes the space structure formed by the road connections. N denotes the quantity of sensor-equipped road regment nodes in a transportation network, H represents the time step of historical data, C represents the number of traffic characteristics monitored by the sensor at each road regment node. P represents the time step for prediction. $\hat{X}_p$ represents the predicted traffic flow over future

P time steps. The function $f(\cdot)$ is parameterized by θ, which includes all learnable weights within the model.

B. Graph Convolutional Networks Description

Graph Convolutional Networks (GCN) are deep learning models specifically designed for processing non-Euclidean data, which aim to learn the target node's features by aggregating its adjacent nodes' features. Single-layer convolution operation of the widely used vanilla GCN[20] that expressed by the following equation:

$$X^{(k+1)} = \mathrm{VGCN}(X^{(k)}) = \mathrm{ReLU}(\hat{\mathbf{A}} X^{(k)} \mathbf{W}) \tag{2}$$

$$\hat{\mathbf{A}} = \hat{\mathbf{D}}^{-\frac{1}{2}} \mathbf{A} \hat{\mathbf{D}}^{-\frac{1}{2}} \tag{3}$$

where $X^{(k)}, X^{(k+1)} \in R^{N \times C}$ respectively denotes the input and output at k-th layer of VGCN. $\mathbf{A}$ denotes an adjacency matrix of the transportation network. $\hat{\mathbf{D}}$ is the degree matrix of the adjacency matrix $\hat{\mathbf{A}}$, with diagonal elements computed as $\mathbf{D}_{ii} = \sum_j \mathbf{A}_{ij}$. $\mathbf{W}$ is a parameter matrix. The function $\mathrm{ReLU}(\cdot)$ is a nonlinear activation function .

C. Spatial-Temporal Features Aggregation Mechanism

Because GCN [20] solely models spatial features at a time step, some studies incorporate the temporal neural network component to capture temporal features. However, these approaches neglect the heterogeneity in spatiotemporal data, as they capture spatial and temporal features independently. To address this limitation, recent studies have proposed novel graph convolution approaches that simultaneously extract spatial and temporal features [27], [28], [29], [30]. These models develop the spatial-temporal feature aggregation mechanism by incorporating temporal edges into the spatial aggregation process of GCN. Aiming to facilitate the subsequent explanation of our work, we take STSGCN [27] to exemplify the mechanism of spatial-temporal features aggregation.

STSGCN constructs a spatiotemporal local graph spanning multiple time steps (presented here as an example of two time steps), which is formed by connecting each node at the present time step t with its equivalent node at the prior time step $t-1$. New adjacency matrix corresponding to the constructed spatiotemporal graph can be expressed as below:

$$\mathbf{A}^{*} = \begin{bmatrix} \mathbf{A}^{t} & \mathbf{I} \\ \mathbf{I} & \mathbf{A}^{t-1} \end{bmatrix} \in R^{2N \times 2N} \tag{4}$$

where $\mathbf{A}^{t-1}$, $\mathbf{A}^{t}$ respectively represent spatial connection relationships of the road segment nodes that correspond to the time steps $t-1$ and t. $\mathbf{I}$ denotes the identity matrix. Since edge weights in $\mathbf{A}^{*}$ are fixed at 1.0, the matrix fails to distinguish the varying influence of different adjacent nodes. To address the limitation, an adaptable weight-masking matrix $\mathbf{M} \in R^{2N \times 2N}$ is introduced to enable adaptive edge weighting. Accordingly, the new adjacency matrix is modified as follows:

$$\mathbf{A}_{sts} = \mathbf{M} \odot \mathbf{A}^{*} \tag{5}$$

where $\odot$ denotes the Hadamard Product, referring to multiplication performed element-wise. Then the operation of spatial-temporal synchronous graph convolution is represented as:

$$\hat{X}^{(t+1)} = \mathrm{STSGC}(\hat{X}) = \sigma\left(\mathbf{A}_{\text{sts}} \hat{X} \mathbf{W}\right) \in R^{2N \times C} \tag{6}$$

where $\hat{X} = \left[X^{(\mathbf{t})}, X^{(\mathbf{t-1})}\right] \in R^{2N \times C}$ is the tensor that amalgamates the traffic flow features from two consecutive time steps t and $t-1$. $\mathbf{W}$ serves as a learnable parameter matrix within the model. Due to the node representations at time step t having already integrated features from other time steps, STSGCN applies a cropping operation to discard those redundant features and preserve only the representations at time step t. This cropping operation can be written as:

$$X^{(t+1)} = \mathrm{Crop}(\hat{X}^{(t+1)}) \in R^{N \times C} \tag{7}$$

Although the current synchronous spatial-temporal features extraction network represented by STSGCN has achieved promising results in traffic flow forecasting, it still suffers from two critical limitations. First, the existing spatio-temporal features aggregation mechanism lacks principled explanations and is not sufficiently comprehensive in modeling, which results in the extracted joint spatio-temporal features being difficult to accurately reflect the actual evolutionary laws of traffic flow accurately. Second, the weight mask matrix $\mathbf{M}$ is completely dependent on the self-training of the model, which means the learned weights do not conform to the actual traffic constraints and cannot effectively simulate the dynamic spatio-temporal relationship.

IV. Methodology

Fig. 3 presents the framework of TASTGCN. First, the input data undergoes enhancement through the TemporalPreprocess (TPP) Module and is passed into the Dual-Path Long-Term Feature Interaction Network (DLFIN), which captures long-term cyclical patterns in traffic flow. The traffic data augmented with spatial-temporal identity information is input into the Dual-Path Short-Term Feature Interaction Network (DSFIN) to model short-term fluctuation patterns in traffic flow. Next, long-term temporal features from DLFIN and short-term temporal features from DSFIN are concatenated and then passed to the Trip-Aware Spatial-Temporal Graph Convolution (TASTGC), which adaptively adjusts spatial-temporal weights to extract physical information-enhanced joint spatial-temporal features. Finally, the traffic flow predictions are generated through a gated linear unit (GLU) followed by a fully connected layer. The components are described in detail below.

A. Spatial-Temporal Identity Information Embedding

Traffic flow data are multivariate time series (MTS), which suffer from the difficulty of distinguishing between temporal and spatial dimensions. Specifically, temporal indistinguishability refers to the fact that in the same traffic road segment node, the traffic flow change patterns in different periods may be highly similar, but their subsequent change trends are significantly different; spatial indistinguishability refers to the fact that in the same period, the traffic flow change patterns of multiple traffic road segment nodes may show similarity, but their subsequent evolutions vary greatly. The spatial and temporal indistinguishability challenges the model's accuracy in predicting future traffic flow based on past patterns. In view of the previous work [40], we address the problem of indistinguishable spatiotemporal dimensions in multivariate temporal

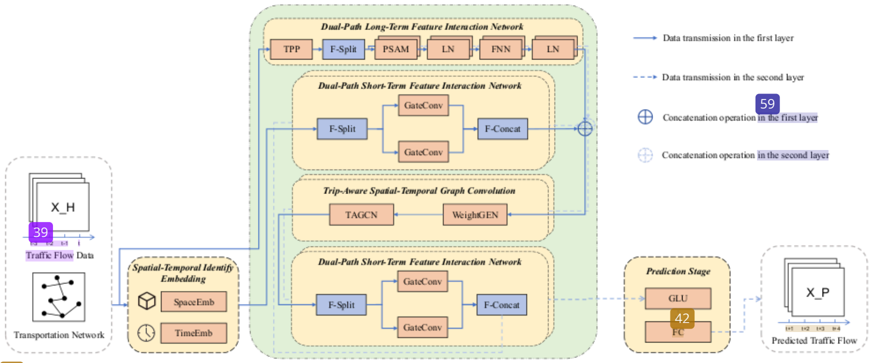


Fig. 3. The model framework of TASTGCN. It is composed of four main modules: spatial-temporal identity information embedding layer, dual-path long-term feature interaction network, dual-path short-term feature interaction network, and trip-aware spatial-temporal graph convolution.

data by embedding spatial-temporal identity information in traffic flow data.

Traffic flow data has periodic patterns, such as peak periods in a day, weekdays and weekends within a week. Therefore, we define two temporal identity information $T_{day} \in R^{T_d \times D_{emb}}$ and $T_{week} \in R^{T_w \times D_{emb}}$ to denote the "day" and "week" positions respectively, where T_d is the number of data collected by sensors at intervals throughout the day, T_w denotes the total amount of days in a week, and D_{emb} denotes dimensions of the each identity information.

To embed temporal identity information for T_{in} time points in the traffic flow data $X_{flow} \in R^{B \times N \times T_{in} \times D_{flow}}$ that is input into the model. We first embed two temporal indexes to the entire time series data of the traffic flow, which are used to find the corresponding day identity information and week identity information for each time point. Time index of "day" $I_{day} \in \{1, 2, ..., T_{in}\}$, time index of "week" $I_{week} \in \{1, 2, ..., 7\}$. The embedded time-indexed traffic flow data is $\hat{X}_{flow} \in R^{B \times N \times T_{in} \times (D_{flow}+2)}$. Subsequently, we extract temporal indices from each traffic data instance to obtain the corresponding embeddings, represented as $E_{day} \in R^{T_{in} \times D_{emb}}$ and $E_{week} \in R^{T_{in} \times D_{emb}}$. In addition, we define a spatial identity information $E_{sapce} \in R^{N \times D_{emb}}$ that is used to distinguish the spatial position of road segment nodes, where N is the total count of road segment nodes. Finally, concatenating E_{day}, E_{week}, E_{space} with the traffic flow data X_{flow} yields the output $X_{emb} \in R^{B \times N \times T_{in} \times (D_{flow}+3D_{emb})}$ of the spatial-temporal identity information embedding layer:

$$X_{emb} = X_{flow} || E_{day} || E_{week} || E_{space} \tag{8}$$

where T_{in} denotes the time length of each traffic data instance, D_{flow} denotes the dimensions of traffic flow feature.

B. Dual-Path Long-Term Feature Interaction Network

Traffic flow exhibits significant long-term temporal dependencies, which are reflected in the complex associations between future traffic conditions and historical states from hours, days, or even weeks earlier. These dependencies stem from patterns of human movement, such as the typical rush hour intervals in mornings and evenings, as well as weekday–weekend cycles. Effectively capturing such long-term correlations enables more accurate modeling of traffic dynamics and improves forecasting performance.To this end, we propose the Dual-Path Long-Term Feature Interaction Network (DLFIN), which adopts an encoder-based structure to model long-term temporal dependencies in traffic flow.

The following is a description of the DLFIN calculation process:

$$\begin{aligned} H^{(l)} &= \mathrm{DLFIN}(\mathrm{X}_{flow}) \\ &= \mathrm{LayerNorm}(\mathrm{FNN}(H_{mid}^{(l)}) + H_{mid}^{(l)}) \end{aligned} \tag{9}$$

here, $H^{(l)}$ denotes the output from the l-th parallel encoder in DLFIN, with $l \in \{1, 2\}$. $\mathrm{LayerNorm}(\cdot)$ denotes the layer normalization operation for stabilizing the hidden state distribution. $\mathrm{FFN}(\cdot)$ denotes the feed-forward neural network used for nonlinear transformation. $H_{mid}^{(l)}$ represents the intermediate features of the l-th parallel encoder, calculated as:

$$H_{mid}^{(l)} = \mathrm{LayerNorm}(\mathrm{Dropout}(\mathrm{PSAM}(X^{(l)})) + X^{(l)}) \tag{10}$$

where $\mathrm{PSAM}(\cdot)$ is the prob sparse attention mechanism [41], which reduces computational complexity while preserving key information capture capabilities. $X^{(i)}$ denotes the features of traffic flow data split into two parts, which can be written as follows:

$$X^{(i)} = \text{F-Split}(\mathrm{Conv}_{1d}(X_{eh})), i \in \{1, 2\} \tag{11}$$

where $\mathrm{Conv}_{1d}(\cdot)$ is a one-dimensional convolutional operation, $\text{F-Split}(\cdot)$ is the feature split operation, which splits the two halves along the feature dimensions of the tensor. X_{eh} is a time series of traffic flow that has been enhanced by a module, which can be expressed as:

$$\begin{aligned}X_{eh} &= \mathrm{TemporalPreprocess}(X_{flow}) \\ &= \mathrm{LayerNorm}(\mathrm{DSConv}(X_{flow}))\end{aligned} \tag{12}$$

where $\mathrm{TemporalPreprocess}(\cdot)$ enhances features by using Depthwise Separable Convolution (DSC) to perform a channel-by-channel convolution operation on the input data. This design allows each feature dimension to independently capture its temporal evolution pattern, followed by layer normalization to stabilize the distribution. In contrast to the global mapping mechanism of linear layers, this approach retains the convolutional advantage in modeling temporal locality while eliminating redundant inter-channel interactions through channel separation. The output is a dimension-preserved feature map with improved representational efficiency.

After DLFIN processes the traffic flow sequence, two sets of long-term dependent features, $H^{(1)}$ and $H^{(2)}$, are produced. Feeding these features into the subsequent network module facilitates the model in extracting long-term temporal dependencies, thereby enhancing prediction accuracy.

C. Dual-Path Short-Term Feature Interaction Network

In traffic flow prediction tasks, short-term temporal feature extraction is crucial for capturing rapid traffic fluctuations. For the purpose of extracting short-term temporal dependencies, we develop the Dual-Path Short-Term Feature Interaction Network (DSFIN). The process of DSFIN can be expressed as:

$$S^{(l)} = \mathrm{DSFIN}(S^{(l-1)}) \tag{13}$$

where $S^{(l-1)}$ is the traffic flow data to the l-th DSFIN and $S^{(0)} = X_{emb} \in R^{B\times N\times T_{\mathrm{in}}\times(D_{flow}+3D)}$. Convolutional neural networks capture short-term dramatic fluctuations in traffic data. Hence, we first split the input features into two independent paths, which are then subjected to extended convolution operations separately. We also introduce residual connections to enhance gradient propagation.

$$(Z_1, Z_2) = \text{F-Split}(\mathrm{Conv}_{1d}(Z^{(l-1)})) \tag{14}$$

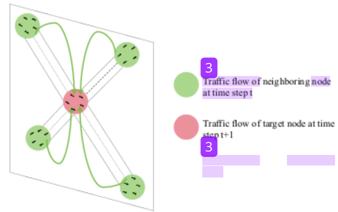
$$\begin{aligned}Z^{(l)} = \text{F-Concat}\Big(&(Z_1 \odot \mathrm{Sigmoid}(\mathrm{Conv}_{1d}(Z_1))), \\ &(Z_2 \odot \mathrm{Sigmoid}(\mathrm{Conv}_{1d}(Z_2)))\Big)\end{aligned} \tag{15}$$

$$S^{(l)} = \mathrm{Linear}(\mathrm{Relu}(Z^{(l)} + \mathrm{Conv}_{1d}(Z^{(l-1)}))) \tag{16}$$

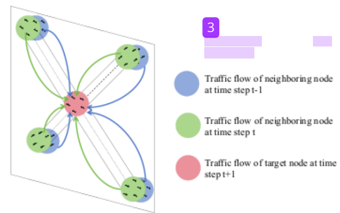
where $Z_1, Z_2 \in R^{B\times N\times T\times(D/2)}$. $\text{F-Concat}(\cdot)$ denotes the operation of concatenating along the feature dimension. $\mathrm{Conv}_{1d}(\cdot)$ denotes a 1D convolution operation along the time dimension of the traffic flow. $\mathrm{Sigmoid}(\cdot)$ and $\mathrm{Relu}(\cdot)$ are nonlinear activation functions. $\odot$ is the Hadamard product.

D. Trip-Aware Spatial-Temporal Graph Convolution

The traffic network is inherently non-Euclidean, and the resulting traffic data exhibits complex spatial and temporal dynamics. Due to the presence of implicit high-order correlations among different road segments, traditional grid-based models are often inadequate at modeling the underlying structure. GCN has shown strong potential in traffic flow forecasting, as it leverages feature propagation and aggregation mechanisms over graph structures to uncover latent spatial dependencies.



(a) Aggregate traffic flow features from one time step in neighboring nodes to learn the traffic flow features of the next time step in the target node.



(b) Aggregate traffic flow features from two time steps in neighboring nodes to learn the traffic flow features of the next time step in the target node.

Fig. 4. Comparison of traffic flow feature aggregation strategies: (a) Aggregate features under a single time step. (b)Aggregate features under two time steps.

A representative example is the vanilla GCN [20], which aggregates the traffic flow of adjacent nodes at the present time step to forecast the traffic flow of the target node at the subsequent time step (as shown in Fig. 4a). This feature aggregation mechanism implicitly contains an unreasonable assumption that vehicles at adjacent nodes can all reach the target node within a certain time interval. However, due to the limitations of road distance and vehicle speed in reality, vehicles may need multiple time intervals to complete the movement between nodes, so it is inaccurate to learn to predict the target node's traffic flow at the next time step by exclusively relying on the traffic flow of neighboring nodes at the current time step. Therefore, we develop a trip-aware spatial-temporal graph convolution to expand the aggregation range of graph convolution in the time dimension (as shown in Fig. 4b). This convolution method dynamically adjusts the weight parameters used to aggregate the traffic flow features of neighboring nodes under two historical time steps according to the two real-world physical factors: the inter-node distance and the vehicle speed, to achieve a more accurate prediction of the target node's traffic flow for the subsequent time step. The trip-aware convolution operation can be formulated as:

$$\begin{aligned}S^{(t+1)} &= \mathrm{TASTGC}(\hat{S}, \mathbf{A}^*) \\ &= \mathrm{ReLU}(\mathbf{A}^* \odot ((\mathbf{K}^* + \mathbf{M}) \odot \mathbf{Mask})\hat{S}\mathbf{W})\end{aligned} \tag{17}$$

where $\mathrm{ReLU}(\cdot)$ denotes the activation function. $\hat{S} = \left[S^{(t)}, S^{(t-1)}\right] \in R^{2N\times T_{in}\times D}$ denotes the tensor of traffic flow features spliced traffic flow features at time step t and t-1. $\mathbf{A}^* = \left[\hat{\mathbf{A}}, \hat{\mathbf{A}}\right] \in R^{N\times 2N}$, $\hat{\mathbf{A}}$ obtained by Eq. (3). $\mathbf{S}$ denotes the normalized distance matrix of the transportation network. $\mathbf{Mask}$ is a mask matrix of $\hat{\mathbf{A}}$, used to lower the number of trainable parameters. $\mathbf{M} \in R^{N\times 2N}$ is the trainable parameter matrix that is initially a zero matrix, $\mathbf{W} \in R^{D_{in}\times D_{out}}$ is a matrix of trainable parameters initialized based on a uniform distribution. $\mathbf{K}^* = [\alpha, \beta] \in R^{N\times 2N}$ is a weight parameter matrix that determines how features from time steps t and $t-1$ are weighted when aggregating traffic flow features for the target node's next time step. Namely, the values in $\mathbf{K}^*$ directly control the relative contributions of two time steps' features during the aggregation process, with these values being computed by Eq. (18) and Eq. (19).

$$\alpha = \mathrm{Clip}(1 - \frac{\mathbf{S}}{(\mathbf{V_{mps}} \times \Delta T)}, 0, 1) \tag{18}$$

$$\beta = 1 - \alpha \tag{19}$$

where $\mathrm{Clip}(\cdot)$ indicates that the weights are restricted to be between 0 and 1, $\mathbf{S}$ is the distance matrix between the nodes of each traffic segment, ΔT is the detection time interval of the sensor (in seconds). $\mathbf{V}_{mps}$ (in meters per second) is the average speed of vehicle movement between the nodes, due to the lack of actual average speed of vehicular traffic between nodes, we use a traffic flow based speed estimation method to learn a traffic congestion coefficient $\mathbf{M}_{rate}$ through a neural network, whose value is negatively correlated to the speed of traffic, and then obtain $\mathbf{V}_{mps}$ from Eq. (20).

$$\mathbf{V}_{mps} = (\mathbf{V}_{kph} \odot (1 - \ln(1 + \mathbf{M}_{rate}))) \times k \tag{20}$$

where $\mathbf{V}_{kph}$ (in kilometers per hour) is a pre-defined speed matrix, representing the maximum speed that vehicles can drive through between adjacent road segments. $k = \frac{1}{3.6}$ is unit conversion coefficient. $\mathbf{M}_{rate}$ is the road congestion coefficient, which is obtained by Eq. (21).

$$\mathbf{M}_{rate} = \mathrm{Sigmoid}(\mathrm{ReLU}(\mathrm{Linear}(X_{flow}))) \tag{21}$$

where $\mathrm{Sigmoid}(\cdot)$ is the activation function, and $\mathrm{Linear}(\cdot)$ denotes the linear layer.

The core mechanism of the trip-aware spatial-temporal graph convolution lies in dynamically adjusting the weights of features from adjacent nodes at time steps t and $t-1$ based on two realistic factors. When the distance between nodes $\mathbf{s}$ increases or the average vehicle driving speed $\mathbf{v}$ decreases, the time required for a vehicle to reach the target node from its neighbors also increases. As a result, the weight assigned to the traffic features of neighboring nodes at time step t decreases, while the weight assigned to those at time step $t-1$ increases. Conversely, when $\mathbf{s}$ decreases or $\mathbf{v}$ increases, the opposite adjustment occurs.

The model can more precisely capture the spatiotemporal evolution patterns in traffic flow by utilizing the aforementioned weight parameters update mechanism, which is based on trip features. In the region where the road segment nodes are farther away or the vehicle speed is lower, the model will rely more on the feature values of the historical time step $t-1$. While in the region where the road segment nodes are closer or the vehicle speed is higher, the model will give more weight to the feature values at the time step t. This mechanism enables the model to extract spatial-temporal features that are more in line with the actual traffic operation patterns and improves the prediction accuracy.

E. Prediction Layer

In the final stage, predictions are produced by employing a GLU and a fully connected layer. GLU refines global spatio-temporal features by enhancing relevant patterns and suppressing redundancy. A fully connected regression layer then processes these refined features to produce the final output. The prediction phase operations are outlined as follows:

$$S_l, S_r = \text{F-Split}(S) \tag{22}$$

$$\hat{X}_p = \mathrm{FC}(S_l * \mathrm{Sigmoid}(S_r)) \tag{23}$$

where $\text{F-Split}(\cdot)$ denotes the split of the input tensor S in half along the feature dimensions into S_l and S_r, $\hat{X}_p \in R^{N\times T_{out}}$ is the traffic flow predicted. T_{out} is the predicted time steps. $\hat{X}_p$ denotes the predicted traffic flow.

F. Loss Function

To quantitatively evaluate the accuracy of the model's predictions, we employ MAE as the primary loss function during training. MAE is a widely used metric in regression tasks that captures the average magnitude of prediction errors, providing a scale-consistent and interpretable measure of model performance. Formally, given the observed values $X_{i,t} \in R^{N\times P}$ and the predicted values $\hat{X}_{i,t} \in R^{N\times P}$, the MAE loss is defined as:

$$\mathcal{L}_{\mathrm{MAE}} = \frac{1}{N \times P} \sum_{i=1}^{N} \sum_{t=1}^{P} \left| \hat{X}_{i,t} - X_{i,t} \right| \tag{24}$$

where N denotes the quantity of sensor nodes, P is the forecasting horizon. The loss formulation ensures that each prediction contributes equally to the final optimization objective. Algorithm 1 outlines the training procedure of TASTGCN.

V. Experiments

A. Datasets and Evaluation Metrics

For the purpose of evaluating the predictive capability of the TASTGCN, we carried out experiments on four datasets, namely PEMS03, PEMS04, PEMS07, and PEMS08. [42]. Detailed information regarding these datasets is presented in Table II.

The original data is captured every thirty seconds and subsequently consolidated into five-minute intervals. Consequently, 12 samples are collected per hour. The spatial adjacency matrix of each dataset is determined by the topology of the corresponding actual transportation network. The traffic data from the previous hour is employed to forecast the next hour's data in experiments. The predictive performance of models is quantitatively assessed using three metrics: MAE, RMSE, and MAPE. MAE captures the average magnitude of prediction deviations by computing the mean absolute difference between forecasted and observed values, offering a scale-preserving and

Algorithm 1 The Process of TASTGCN for Forecasting Traffic Flow.

Input: Traffic flow data $X_{flow} \in R^{N \times T_{\text{in}} \times D_{flow}}$; Traffic road network G;

Output: The forecasted traffic flow $\hat{X}_p$

Obtain X_{eh} through X_{flow} and Eq. (12);
Obtain $X^{(1)}$, $X^{(2)}$ through X_{eh} and Eq. (11);
Obtain $H^{(1)}$, $H^{(2)}$ through $X^{(1)}$, $X^{(2)}$, and Eq. (9);
Integrate the spatial-temporal identity information E_{day}, E_{week}, E_{space} and the traffic flow data X_{flow} to obtain $X_{emb} \in R^{N \times T_{\text{in}} \times (D_{flow}+3D_{emb})}$;
Infer $\mathbf{M}_{rate}$ through Eq. (21);
Obtain $\mathbf{V}_{mps}$ through Eq. (20);
for $i \in \{1, 2\}$ **do**
    Obtain $S^{(1)}$ through Eq. (13);
    Construct P by splicing $S^{(1)}$ and $X^{(i)}$;
    Aggregate spatio-temporal features from neighboring nodes through P and Eq. (15) yields F;
    Obtain $S^{(2)}$ through F and Eq. (13);
end for
Obtain $\mathbf{S_l}$ and $\mathbf{S_r}$ through Eq. (22);
Obtain the predicted traffic flow $\hat{X}_p$ through Eq. (23).
return $\hat{X}_p$;

robust evaluation that treats all errors equally. RMSE emphasizes larger errors by squaring the residuals before averaging and then taking the square root, making it more sensitive to outliers and suitable for scenarios where significant deviations are especially undesirable. MAPE provides an interpretable, relative measure of prediction accuracy by expressing absolute errors as a percentage of the ground truth, thereby enabling error comparison across variables of different scales. These metrics are defined as:

$$\text{MAE} = \frac{1}{n}\sum_{i=1}^{n} |\hat{y}_i - y_i| \tag{25}$$

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n} (\hat{y}_i - y_i)^2} \tag{26}$$

$$\text{MAPE} = \frac{100\%}{n}\sum_{i=1}^{n} \left|\frac{\hat{y}_i - y_i}{y_i + \epsilon}\right| \tag{27}$$

where $\hat{y}_i$ represents the observed traffic flow, and y_i is the result predicted by the model. n is the total number of data samples. ϵ is a minor positive constant included to prevent division by zero in the calculation of MAPE. In this study, it is manually set to 1×10^{-5}. To ensure a uniform scale and stable distribution, the raw traffic data was normalized during preprocessing.

B. Baseline Methods

The TASTGCN is benchmarked against night representative models, providing a comprehensive evaluation of its performance.

- STGCN [22]: Spatial-Temporal Graph Convolutional Network merges graph convolution and 1D convolution to efficiently acquire spatial and temporal correlations.

TABLE II
PEMS SERIES DATASET DETAILS

Dataset	Nodes	Edges	Samples	Time Span
PEMS03	358	547	26,208	2018/09/01-2018/11/30
PEMS04	307	340	16,992	2018/01/01-2018/02/28
PEMS07	883	866	28,224	2017/05/01-2017/08/31
PEMS08	170	295	17,856	2016/07/01-2016/08/31

- DCRNN [21]: Diffusion Convolutional Recurrent Neural Network models spatial dependencies by leveraging diffusion graph convolution and captures temporal dynamics utilizing Gated Recurrent Units (GRU).
- ASTGCN [24]: Attention-based Spatial-Temporal Graph Convolutional Network incorporates attention mechanisms into both time and space graph convolution to enhance sensitivity to critical spatio-temporal features.
- GWNet [25]: GraphWaveNet constructs an adaptive adjacency matrix for the purpose of learning spatial relationships and utilizes 1-D dilated causal convolutions for modeling temporal correlations.
- STSGCN [27]: Spatial-temporal Synchronous Graph Convolutional Network introduces a new convolution operation to model joint spatiotemporal dependencies.
- STFGNN [43]: Spatial-Temporal Fusion Graph Neural Network generates graphs based on time series similarity and integrates fused graphs for effectively capturing long-term relationships.
- AGCRN [26]: Adaptive Graph Convolutional Recurrent Network enhances graph convolution with two adaptive modules and integrates it with a GRU to record dynamic spatiotemporal patterns.
- SCINet [44]: Sample Convolution and Interaction Network adopts a recursive downsampling-convolution-interaction framework, extracting essential temporal features from downsampled sub-sequences through multiple convolutions to represent complex temporal dynamics.
- STPGCN [29]: Spatial-Temporal Position-Aware Graph Convolution Network proposes a position-aware relational inference module that leverages spatial and temporal embeddings of nodes to dynamically model spatio-temporal dependencies.

C. Experimental Setup

The datasets used in our experiments are split into training, validation, and testing subsets with a ratio of 6:2:2. Each prediction utilizes historical data from the previous hour, represented by 12 sequential time steps, to estimate the traffic flow for the upcoming hour, also comprising 12 time steps.

All experimental procedures are carried out on a machine featuring an NVIDIA RTX 4060 GPU and 32 GB of RAM. The TASTGCN model is implemented utilizing PyTorch version 1.8.2 and Python 3.8.18. During the training phase, the model is optimized using the Adam optimizer, with a learning rate of 0.001, a batch size of 16, and the number of training epochs set to 50.

D. Results and Analysis of Predictive Performance

Table III presents the performance comparison of ten models

TABLE III
COMPARISON OF MODEL PERFORMANCE

Model	PEMS03 (60min)			PEMS04 (60min)			PEMS07 (60min)			PEMS08 (60min)		
	MAE	RMSE	MAPE	MAE	RMSE	MAPE	MAE	RMSE	MAPE	MAE	RMSE	MAPE
STGCN	17.49	30.12	17.15	22.70	35.55	14.59	25.38	38.78	11.08	18.02	27.83	11.40
DCRNN	18.18	30.31	18.91	24.70	38.12	17.12	25.30	38.58	11.66	17.86	27.83	11.45
ASTGCN	17.69	29.66	19.40	22.93	35.22	16.56	28.05	42.57	13.92	18.61	28.16	18.61
GWNet	19.85	32.94	19.31	25.45	39.70	17.29	26.85	42.78	12.12	19.13	31.05	12.68
STSGCN	17.48	39.21	16.78	21.19	33.65	13.90	24.26	39.03	10.21	17.13	26.80	10.96
STFGNN	16.77	28.34	16.30	19.83	31.88	13.02	22.07	35.80	9.21	16.64	26.22	10.60
AGCRN	15.82	27.65	15.06	19.70	32.48	12.99	21.00	35.02	8.84	16.13	25.53	10.33
SCINET	15.58	**25.11**	14.72	19.63	31.60	12.18	21.49	34.27	8.92	16.12	25.15	9.99
STPGCN	15.48	27.00	16.02	18.70	**30.39**	12.20	20.01	33.37	8.58	14.38	23.47	9.37
TASTGCN	**14.99**	25.48	**14.41**	**18.61**	30.54	**11.72**	**19.62**	**32.96**	**8.14**	**13.77**	**23.40**	**8.72**

across four datasets. The results demonstrate that TASTGCN almost achieves the best performance on four real-world datasets. In the traffic flow prediction task, although most baseline models are capable of leveraging spatial and temporal information, they all have certain limitations: DCRNN has difficulty modeling long-term temporal dependencies based on the GRU structure, which leads to a decline in the long-term prediction performance. STGCN processes spatio-temporal data in a purely convolutional manner, it cannot capture global contextual information features due to the limitation of convolutional kernels. Despite ASTGCN introducing an attention mechanism to enhance the modeling capability, its computational complexity increases significantly, and there may be a matching bias between the attention weights and the sequence representation, which affects the robustness of the model. GWNet and AGCRN capture spatial correlations through the adaptive matrix, but ignore the time delay of traffic propagation, which leads to reduced accuracy in modeling spatio-temporal relations. STSGCN captures spatial-temporal heterogeneity through the local spatial-temporal graph but has limited capability to model global temporal dependencies. STFGNN uses the dynamic time warping (DTW) algorithm to construct the temporal graph, and its computational complexity restricts its practicality in large-scale scenarios. SCINet improves the prediction performance through complex temporal dynamics, but it does not utilize the topological structure information of the transportation network, which makes the spatial relationship modeling a notable limitation. STPGCN infers spatial-temporal dependence through the embedded spatial-temporal information, which does not take into account the realistic laws of traffic propagation flow.

E. Ablation Experiments

In order to deeply investigate the principles and effectiveness of the different modules in TASTGCN, we performed ablation tests on the PEMS04 and PEMS08. The experiment utilizes the prior hour's data to forecast the following hour's data, with both consisting of 12 time steps. Several variants of TASTGCN are developed as follows:

- noTASTGC: Replaces weight mask matrices with trip feature constraints by trainable weight mask matrices.
- noSTIE: Removing the spatial-temporal information embedding layer.
- noDLFIN: Removing the dual-path long-term feature interaction network.
- noDSFIN: Removing the dual-path short-term feature interaction network.

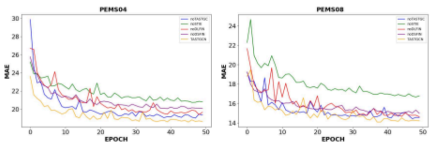


Fig. 5. Change in MAE with training rounds for model variants with different modules removed.

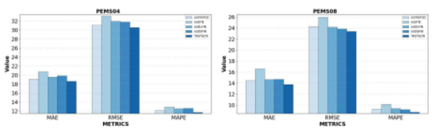


Fig. 6. MAE, RMSE, and MAPE results for model variants with different modules removed.

TABLE IV
PERFORMANCE RESULTS FOR EACH MODEL VARIANT

Models	PEMS04 (60min)			PEMS08 (60min)		
	MAE	RMSE	MAPE	MAE	RMSE	MAPE
noTASTGC	19.11	31.07	12.17	14.47	24.27	9.29
noSTIE	20.73	33.09	12.94	16.58	25.94	10.15
noDLFIN	19.58	31.92	12.55	14.66	24.17	9.40
noDSFIN	19.85	31.79	12.64	14.68	23.85	9.21
TASTGCN	18.61	30.54	11.72	13.77	23.40	8.72

Curve of MAE with number of epoch for each variant model is presented in Fig. 5, the final performance comparison is shown in Fig. 6 and Table IV. TASTGCN shows the most excellent prediction performance, which is significantly better

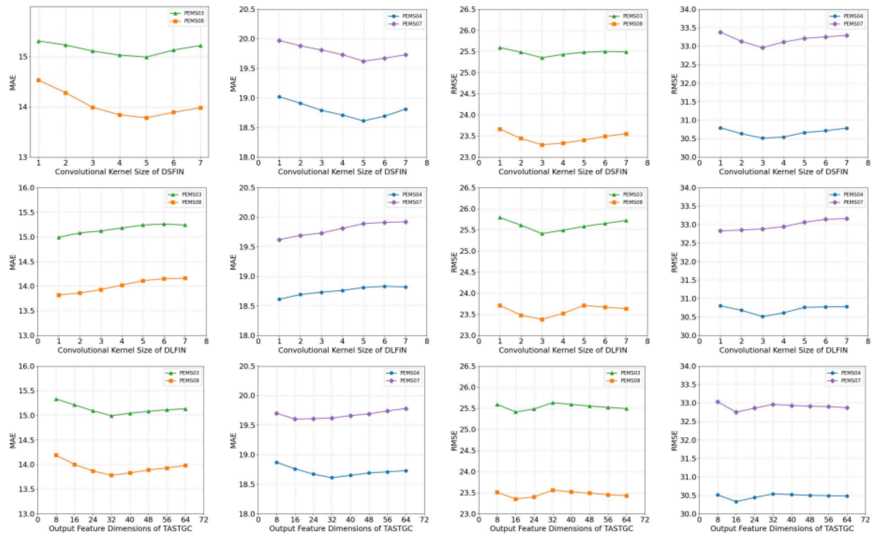


Fig. 7. Hyperparameter study on the PEMS series of datasets.

than the noTASTGC, which verifies that the trip-aware spatial-temporal graph convolution module proposed in this paper is able to efficiently model complex spatial-temporal dependencies. Removing the spatial-temporal information embedding layer (noSTIE) leads to a significant degradation of the model performance, which proves the key role of spatial-temporal location embedding for accurate inference of spatiotemporal relations among data. In addition, both removing the dual-path long-term feature interaction network (noDLFIN) and the dual-path short-term feature interaction network (noDSFIN) cause a reduction in the model prediction performance, which demonstrates the importance of capturing the long- and short-term temporal dependence features in data to increase the prediction accuracy.

F. Hyperparameter Analysis

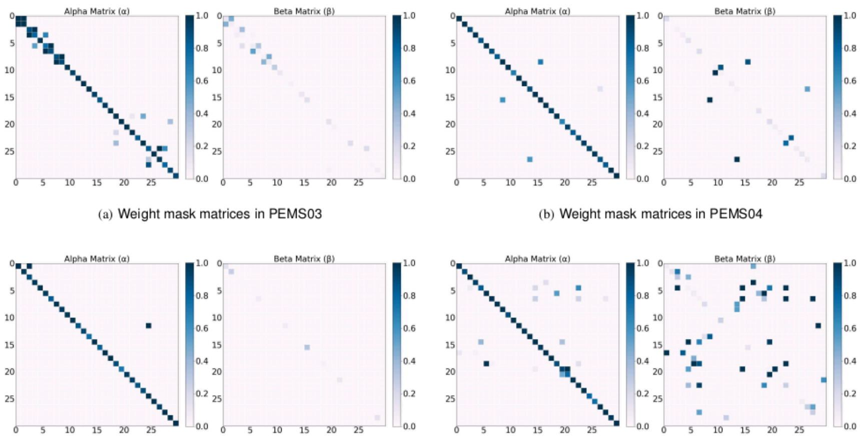
To investigate the sensitivity of the model to hyperparameters, we conduct studies on the PEMS03, PEMS04, PEMS07, and PEMS08. Specifically, we examine the impact of three hyperparameters on the model's performance: the convolutional kernel size of DSFIN, the convolutional kernel size of DLFIN, and the output feature dimensions of TASTGCN.

The hyperparameter experiments' outcomes are presented in Figure. in Fig. 7. To begin with, the model demonstrates optimal performance in terms of the lowest MAE or RMSE when the convolutional kernel size of DSFIN is set to 3 or 5. An excessively large or small kernel size results in degradation of the model's performance. Secondly, increasing the kernel size in DLFIN's dimensionally extended convolution operation negatively affects the model's MAE accuracy, but the RMSE metric performs best when DLFIN's convolution kernel is taken to be 3. Lastly, the MAE metric of the model is best when the output feature dimension of TASTGC is 32, but the RMSE metric of the model is best when the output dimension is 16. Since these three hyperparameters have similar effects on the model on the PEMS series of datasets, TASTGCN uses the same hyperparameters on the PEMS series of datasets.

G. Visualization Analysysis

To demonstrate the role of weight mask matrices in capturing traffic flow features across different time steps in the trip-aware graph convolution, we visualize the weight matrices α and β as heat maps. Specifically, we select α and β from the last training batch at the point when the model achieves its best performance on each dataset. For plotting purposes, we select the first 30 nodes in the matrix. The visualization results are shown in Fig. 8.

To better analyze the heat map results, we counted the number of connectivity relations for the first 30 nodes in the datasets. The counts were 50, 38, 34, and 72 for PEMS03, PEMS04, PEMS07, and PEMS08, respectively. Based on the heatmaps, the following observations can be made. First, there is a clear positive correlation between the number of elements in the weight matrix and the number of edges between nodes. This is particularly evident in the PEMS07 and PEMS08, which have the fewest and the most edges, respectively, and



(a) Weight mask matrices in PEMS03

(b) Weight mask matrices in PEMS04

(c) Weight mask matrices in PEMS07

(d) Weight mask matrices in PEMS08

Fig. 8. Weight mask matrix alpha (α) and beta (β) for the first thirty nodes at the end of training on the PEMS series of datasets.

whose heatmaps accordingly exhibit the sparsest and densest weight distributions. Second, since the color intensity in the heatmaps reflects the contribution of the features of the time step corresponding to the weight matrix to the convolution process, the analysis indicates that nodes across all four datasets primarily rely on their own features at the time step t when learning and aggregating information from the nodes themselves. Finally, after excluding the diagonal weights, notable differences emerge in the degree of dependence on features from the time step $t-1$: the PEMS08 dataset shows the strongest reliance on traffic features from $t-1$, followed by PEMS04, PEMS03, and PEMS07 in decreasing order. This finding reveals that trip-aware spatial-temporal graph convolution exhibits varying feature learning patterns across different traffic networks.

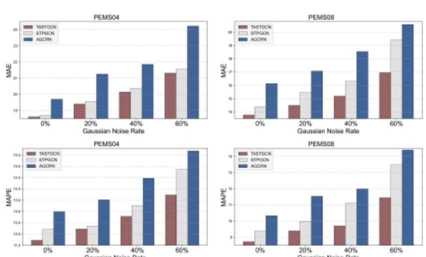


Fig. 9. Robustness analysis on PEMS04 and PEMS08 datasets.

H. Robustness Evaluation

In practical traffic prediction contexts, the gathered data may exhibit noise. Therefore, testing the robustness of models is crucial for applying traffic prediction models in practice. We compare TASTGCN with AGCRN, SCINET, and STPGCN.

To emulate real-world noise, we injected Gaussian noise (mean = 10, standard deviation = 500) into the first 60% of the data in PEMS04 and PEMS08, which corresponds to the training set. The noise was added at three levels: 20%, 40%, and 60%. The test and validation sets were kept noise-free.

Since SCINET is too sensitive to Gaussian noise, resulting in large error variations, in order to visualize the experimental results, the experimental result plots of the robustness test only show TASTGCN, STPGCN, and AGCRN. The results of robustness evaluation are shown in Fig. 9. TASTGCN achieves the best performance in datasets with different data quality, demonstrating stronger robustness compared to AGCRN, and STPGCN. In addition, SCINET's poorer robustness to noise may be due to the fact that its graph convolution smoothes out the noise through the neighboring nodes, while SCINET only relies on a single-point time-series convolution, where the noise is directly transmitted, leading to an increase in prediction error.

VI. Conclusion

We propose a novel trip-aware spatial-temporal graph convolutional network (TASTGCN) for traffic flow prediction. TASTGCN introduces learnable spatial-temporal identity information, enabling it to explicitly distinguish between different nodes and periods. Two dual-path feature interaction networks are designed, where the long-term network captures periodic patterns, while the short-term network focuses on transient fluctuations. Developing a trip-aware graph convolution to ensure that the extracted spatial-temporal features align with the real-world evolution of traffic flow by incorporating constraints such as inter-node distance and vehicle speed. Experiments on four datasets show the effectiveness of TASTGCN in predicting traffic flow.

References

[1] J. Liu and W. Guan, "A summary of traffic flow forecasting methods," *Journal of highway and transportation research and development*, vol. 21, no. 3, pp. 82–85, 2004.

[2] B. M. Williams and L. A. Hoel, "Modeling and forecasting vehicular traffic flow as a seasonal arima process: Theoretical basis and empirical results," *Journal of transportation engineering*, vol. 129, no. 6, pp. 664–672, 2003.

[3] G. E. Box and D. A. Pierce, "Distribution of residual autocorrelations in autoregressive-integrated moving average time series models," *Journal of the American statistical Association*, vol. 65, no. 332, pp. 1509–1526, 1970.

[4] M. S. Ahmed and A. R. Cook, *Analysis of freeway traffic time-series data by using Box-Jenkins techniques*, 1979, no. 722.

[5] S. Makridakis and M. Hibon, "Arma models and the box–jenkins methodology," *Journal of forecasting*, vol. 16, no. 3, pp. 147–163, 1997.

[6] H. Drucker, C. J. Burges, L. Kaufman, A. Smola, and V. Vapnik, "Support vector regression machines," *Advances in neural information processing systems*, vol. 9, 1996.

[7] C.-H. Wu, J.-M. Ho, and D.-T. Lee, "Travel-time prediction with support vector regression," *IEEE transactions on intelligent transportation systems*, vol. 5, no. 4, pp. 276–281, 2004.

[8] N. Mitrovic, M. T. Asif, J. Dauwels, and P. Jaillet, "Low-dimensional models for compressed sensing and prediction of large-scale traffic data," *IEEE Transactions on Intelligent Transportation Systems*, vol. 16, no. 5, pp. 2949–2954, 2015.

[9] R. Chen, C.-Y. Liang, W.-C. Hong, and D.-X. Gu, "Forecasting holiday daily tourist flow based on seasonal support vector regression with adaptive genetic algorithm," *Applied Soft Computing*, vol. 26, pp. 435–443, 2015.

[10] G. Leshem and Y. Ritov, "Traffic flow prediction using adaboost algorithm with random forests as a weak learner," in *Proceedings of world academy of science, engineering and technology*, vol. 19. Citeseer, 2007, pp. 193–198.

[11] U. Johansson, H. Boström, T. Löfström, and H. Linusson, "Regression conformal prediction with random forests," *Machine learning*, vol. 97, pp. 155–176, 2014.

[12] N. Ramakrishnan and T. Soni, "Network traffic prediction using recurrent neural networks," in *2018 17th IEEE international conference on machine learning and applications (ICMLA)*. IEEE, 2018, pp. 187–193.

[13] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural computation*, vol. 9, no. 8, pp. 1735–1780, 1997.

[14] Z. Zhao, W. Chen, X. Wu, P. C. Chen, and J. Liu, "Lstm network: a deep learning approach for short-term traffic forecast," *IET intelligent transport systems*, vol. 11, no. 2, pp. 68–75, 2017.

[15] H. Yu, Z. Wu, S. Wang, Y. Wang, and X. Ma, "Spatiotemporal recurrent convolutional networks for traffic prediction in transportation networks," *Sensors*, vol. 17, no. 7, p. 1501, 2017.

[16] K. Cho, B. Van Merriënboer, C. Gulcehre, D. Bahdanau, F. Bougares, H. Schwenk, and Y. Bengio, "Learning phrase representations using rnn encoder-decoder for statistical machine translation," *arXiv preprint arXiv:1406.1078*, 2014.

[17] R. Fu, Z. Zhang, and L. Li, "Using lstm and gru neural network methods for traffic flow prediction," in *2016 31st Youth academic annual conference of Chinese association of automation (YAC)*. IEEE, 2016, pp. 324–328.

[18] J. Zhang, Y. Zheng, and D. Qi, "Deep spatio-temporal residual networks for citywide crowd flows prediction," in *Proceedings of the AAAI conference on artificial intelligence*, vol. 31, no. 1, 2017.

[19] S. Guo, Y. Lin, S. Li, Z. Chen, and H. Wan, "Deep spatial–temporal 3d convolutional neural networks for traffic data forecasting," *IEEE Transactions on Intelligent Transportation Systems*, vol. 20, no. 10, pp. 3913–3926, 2019.

[20] T. N. Kipf and M. Welling, "Semi-supervised classification with graph convolutional networks," *arXiv preprint arXiv:1609.02907*, 2016.

[21] Y. Li, R. Yu, C. Shahabi, and Y. Liu, "Diffusion convolutional recurrent neural network: Data-driven traffic forecasting," *arXiv preprint arXiv:1707.01926*, 2017.

[22] B. Yu, H. Yin, and Z. Zhu, "Spatio-temporal graph convolutional networks: A deep learning framework for traffic forecasting," *arXiv preprint arXiv:1709.04875*, 2017.

[23] 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, 2019.

[24] S. Guo, Y. Lin, N. Feng, C. Song, and H. Wan, "Attention based spatial-temporal graph convolutional networks for traffic flow forecasting," in *Proceedings of the AAAI conference on artificial intelligence*, vol. 33, no. 01, 2019, pp. 922–929.

[25] Z. Wu, S. Pan, G. Long, J. Jiang, and C. Zhang, "Graph wavenet for deep spatial-temporal graph modeling," *arXiv preprint arXiv:1906.00121*, 2019.

[26] L. Bai, L. Yao, C. Li, X. Wang, and C. Wang, "Adaptive graph convolutional recurrent network for traffic forecasting," *Advances in neural information processing systems*, vol. 33, pp. 17 804–17 815, 2020.

[27] 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.

[28] S. Li, L. Ge, Y. Lin, and B. Zeng, "Adaptive spatial-temporal fusion graph convolutional networks for traffic flow forecasting," in *2022 International Joint Conference on Neural Networks (IJCNN)*. IEEE, 2022, pp. 1–8.

[29] Y. Zhao, Y. Lin, H. Wen, T. Wei, X. Jin, and H. Wan, "Spatial-temporal position-aware graph convolution networks for traffic flow forecasting," *IEEE Transactions on Intelligent Transportation Systems*, vol. 24, no. 8, pp. 8650–8666, 2022.

[30] W. Zhu, B. Long, and J. Xiao, "Spatial-temporal retentive heterogeneous graph convolutional network for traffic flow prediction," in *2024 International Joint Conference on Neural Networks (IJCNN)*. IEEE, 2024, pp. 1–8.

[31] Q. Li, X. Yang, Y. Wang, Y. Wu, and D. He, "Spatial–temporal traffic modeling with a fusion graph reconstructed by tensor decomposition," *IEEE Transactions on Intelligent Transportation Systems*, vol. 25, no. 2, pp. 1749–1760, 2023.

[32] Z. Lu, C. Zhou, J. Wu, H. Jiang, and S. Cui, "Integrating granger causality and vector auto-regression for traffic prediction of large-scale wlans." *KSII Transactions on Internet & Information Systems*, vol. 10, no. 1, 2016.

[33] P. Cai, Y. Wang, G. Lu, P. Chen, C. Ding, and J. Sun, "A spatiotemporal correlative k-nearest neighbor model for short-term traffic multistep forecasting," *Transportation Research Part C: Emerging Technologies*, vol. 62, pp. 21–34, 2016.

[34] C. J. Burges, "A tutorial on support vector machine for pattern recognition," *Data mining and knowledge discovery*, vol. 2, no. 2, pp. 955–974, 1998.

[35] J. Van Lint, S. Hoogendoorn, and H. J. van Zuylen, "Accurate freeway travel time prediction with state-space neural networks under missing data," *Transportation Research Part C: Emerging Technologies*, vol. 13, no. 5-6, pp. 347–369, 2005.

[36] G. Fusco, C. Colombaroni, and N. Isaenko, "Comparative analysis of implicit models for real-time short-term traffic predictions," *IET Intelligent Transport Systems*, vol. 10, no. 4, pp. 270–278, 2016.

[37] Z. Wu, S. Pan, F. Chen, G. Long, C. Zhang, and S. Y. Philip, "A comprehensive survey on graph neural networks," *IEEE transactions on neural networks and learning systems*, vol. 32, no. 1, pp. 4–24, 2020.

[38] 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, 2023.

[39] Y. An, Z. Li, X. Li, W. Liu, X. Yang, H. Sun, M. Chen, Y. Zheng, and Y. Gong, "Spatio-temporal multivariate probabilistic modeling for traffic prediction," *IEEE Transactions on Knowledge and Data Engineering*, 2025.

[40] Z. Shao, Z. Zhang, F. Wang, W. Wei, and Y. Xu, "Spatial-temporal identity: A simple yet effective baseline for multivariate time series forecasting," in *Proceedings of the 31st ACM international conference on information & knowledge management*, 2022, pp. 4454–4458.

[41] H. Zhou, S. Zhang, J. Peng, S. Zhang, J. Li, H. Xiong, and W. Zhang, "Informer: Beyond efficient transformer for long sequence time-series forecasting," in *Proceedings of the AAAI conference on artificial intelligence*, vol. 35, no. 12, 2021, pp. 11 106–11 115.

[42] C. Chen, K. Petty, A. Skabardonis, P. Varaiya, and Z. Jia, “Freeway performance measurement system: mining loop detector data,” *Transportation research record*, vol. 1748, no. 1, pp. 96–102, 2001.

[43] 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.

[44] M. Liu, A. Zeng, M. Chen, Z. Xu, Q. Lai, L. Ma, and Q. Xu, “Scinet: Time series modeling and forecasting with sample convolution and interaction,” *Advances in Neural Information Processing Systems*, vol. 35, pp. 5816–5828, 2022.

Trip-Aware Spatial-Temporal Graph Convolutional Network for Traffic Flow Forecasting.pdf

ORIGINALITY REPORT

10%	7%	9%	1%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	www.mdpi.com Internet Source	1%
2	export.arxiv.org Internet Source	<1%
3	Jia Hu, Xianghong Lin, Chu Wang. "DSTGCN: Dynamic Spatial-Temporal Graph Convolutional Network for Traffic Prediction", IEEE Sensors Journal, 2022 Publication	<1%
4	vbn.aau.dk Internet Source	<1%
5	"Advanced Intelligent Computing Technology and Applications", Springer Science and Business Media LLC, 2023 Publication	<1%
6	www.hindawi.com Internet Source	<1%
7	link.springer.com Internet Source	<1%
8	Submitted to University of New South Wales Student Paper	<1%
9	Qianqian Ren, Yang Li, Yong Liu. "Transformer-enhanced periodic temporal convolution network for long short-term traffic flow forecasting", Expert Systems with Applications, 2023 Publication	<1%
10	unsworks.unsw.edu.au Internet Source	

<1 %

11	Chuang Cai, Zhijian Qu, Liqun Ma, Lianfei Yu, Wenbo Liu, Chongguang Ren. "Spatial-Temporal Traffic Flow Prediction with Fusion Graph Convolution Network and Enhanced Gated Recurrent Units", IEEE Access, 2024 Publication	<1 %
12	waseda.repo.nii.ac.jp Internet Source	<1 %
13	dokumen.pub Internet Source	<1 %
14	Xing Wang, Feifei Chen, Biao Jin, Mingwei Lin, Fumin Zou, Ruihao Zeng. "DSTSPYN: a dynamic spatial-temporal similarity pyramid network for traffic flow prediction", Applied Intelligence, 2024 Publication	<1 %
15	"Web Information Systems Engineering – WISE 2020", Springer Science and Business Media LLC, 2020 Publication	<1 %
16	deepai.org Internet Source	<1 %
17	pure.tue.nl Internet Source	<1 %
18	R. N. V. Jagan Mohan, B. H. V. S. Rama Krishnam Raju, V. Chandra Sekhar, T. V. K. P. Prasad. "Algorithms in Advanced Artificial Intelligence - Proceedings of International Conference on Algorithms in Advanced Artificial Intelligence (ICAAAI-2024)", CRC Press, 2025 Publication	<1 %
19	www.springerprofessional.de Internet Source	<1 %

20	d197for5662m48.cloudfront.net Internet Source	<1%
21	library.oapen.org Internet Source	<1%
22	www.nature.com Internet Source	<1%
23	www.researchgate.net Internet Source	<1%
24	Aoyu Liu, Yaying Zhang. "An efficient spatial-temporal transformer with temporal aggregation and spatial memory for traffic forecasting", Expert Systems with Applications, 2024 Publication	<1%
25	Yanfeng Sun, Xiangheng Jiang, Yongli Hu, Fuqing Duan, Kan Guo, Boyue Wang, Junbin Gao, Baocai Yin. "Dual Dynamic Spatial-Temporal Graph Convolution Network for Traffic Prediction", IEEE Transactions on Intelligent Transportation Systems, 2022 Publication	<1%
26	fysiikka.uku.fi Internet Source	<1%
27	jisem-journal.com Internet Source	<1%
28	"Rough Sets", Springer Science and Business Media LLC, 2025 Publication	<1%
29	Aoyu Liu, Yaying Zhang. "Spatial–Temporal Dynamic Graph Convolutional Network With Interactive Learning for Traffic Forecasting", IEEE Transactions on Intelligent Transportation Systems, 2024 Publication	<1%

30	Bolong Zheng, Qi Hu, Lingfeng Ming, Jilin Hu, Lu Chen, Kai Zheng, Christian S. Jensen. "SOUP: Spatial-Temporal Demand Forecasting and Competitive Supply", IEEE Transactions on Knowledge and Data Engineering, 2021 Publication	<1 %
31	Xian Yu, Yinxin Bao, Quan Shi. "Spatial-temporal synchronous graphsage for traffic prediction", Applied Intelligence, 2024 Publication	<1 %
32	Yuxuan Zhang, Senzhang Wang, Bing Chen, Jiannong Cao, Zhiqiu Huang. "TrafficGAN: Network-Scale Deep Traffic Prediction With Generative Adversarial Nets", IEEE Transactions on Intelligent Transportation Systems, 2019 Publication	<1 %
33	assets-eu.researchsquare.com Internet Source	<1 %
34	ebuah.uah.es Internet Source	<1 %
35	"Neural Information Processing", Springer Science and Business Media LLC, 2024 Publication	<1 %
36	Adnan Zeb, Jianying Zheng, Yongchao Ye, Junde Chen, Shiyao Zhang, Xuetao Wei, James Jianqiao Yu. "Traffic forecasting with meta attentive graph convolutional recurrent network", Expert Systems with Applications, 2025 Publication	<1 %
37	Armando Fandango, R. Paul Wiegand. "Towards investigation of iterative strategy for data mining of short-term traffic flow with Recurrent Neural Networks", Proceedings of the 2nd International Conference on	<1 %

Information System and Data Mining - ICISDM '18, 2018
Publication

38	Chang Lu, Rui Wang, Beibei Huang, Yuan Li, Zunkai Huang, Yicong Zhou, Aiwen Luo. "ALResNet: Attention-Driven Lightweight Residual Network for Fast and Accurate Image Recognition", 2021 The 4th International Conference on Machine Learning and Machine Intelligence, 2021 Publication	<1 %
39	Liu Bin-sheng, Xing Zhan-wen, Yang Hai-tao, Hou Yu-peng. "A Combination Predicted Model of Short Term Traffic Flow", 2006 International Conference on Management Science and Engineering, 2006 Publication	<1 %
40	Siyang Luo. "TSGAT: Spatial graph attention-based spatio-temporal flow prediction for big data retailing", IEEE Access, 2022 Publication	<1 %
41	api.repository.cam.ac.uk Internet Source	<1 %
42	hal.science Internet Source	<1 %
43	web.archive.org Internet Source	<1 %
44	www2.mdpi.com Internet Source	<1 %
45	zheng-kai.com Internet Source	<1 %
46	"Database Systems for Advanced Applications", Springer Science and Business Media LLC, 2019 Publication	<1 %

47	"Intelligent Computing Methodologies", Springer Science and Business Media LLC, 2017 Publication	<1 %
48	Castro-Neto, M.. "Online-SVR for short-term traffic flow prediction under typical and atypical traffic conditions", Expert Systems With Applications, 200904 Publication	<1 %
49	Guiyang Luo, Hui Zhang, Quan Yuan, Jinglin Li, Wendong Wang, Fei-Yue Wang. "One Size Fits All: A Unified Traffic Predictor for Capturing the Essential Spatial–Temporal Dependency", IEEE Transactions on Neural Networks and Learning Systems, 2024 Publication	<1 %
50	J.W.C. van Lint. "Incremental and online learning through extended kalman filtering with constraint weights for freeway travel time prediction", 2006 IEEE Intelligent Transportation Systems Conference, 2006 Publication	<1 %
51	Junhao Ming, Dongmei Zhang, Wei Han. "Multi-Scale Spatial–Temporal Transformer: A Novel Framework for Spatial–Temporal Edge Data Prediction", Applied Sciences, 2023 Publication	<1 %
52	Lugosch, Loren Peter. "Deep Neural Networks for Voice Control", McGill University (Canada), 2024 Publication	<1 %
53	Mao Chen, Yi Xu, Liangzhe Han, Leilei Sun. "Chapter 11 Sampling Spatial-Temporal Attention Network forTraffic Forecasting", Springer Science and Business Media LLC, 2023 Publication	<1 %

54	Qingyuan Zhan, Guixing Wu, Chuang Gan. "MAGCN: A Multi-Adaptive Graph Convolutional Network for Traffic Forecasting", 2021 International Joint Conference on Neural Networks (IJCNN), 2021 Publication	<1%
55	Shun Wang, Yong Zhang, Yongli Hu, Baocai Yin. "Knowledge fusion enhanced graph neural network for traffic flow prediction", Physica A: Statistical Mechanics and its Applications, 2023 Publication	<1%
56	Xianglong Luo, Danyang Li, Yu Yang, Shengrui Zhang. "Spatiotemporal Traffic Flow Prediction with KNN and LSTM", Journal of Advanced Transportation, 2019 Publication	<1%
57	Yuchao Wang, Zixiang Tian, Huixuan Fu. "Multivariate USV Motion Prediction Method Based on a Temporal Attention Weighted TCN-Bi-LSTM Model", Journal of Marine Science and Engineering, 2024 Publication	<1%
58	arxiv.org Internet Source	<1%
59	ebin.pub Internet Source	<1%
60	mdpi-res.com Internet Source	<1%
61	mospace.umsystem.edu Internet Source	<1%
62	neurohive.io Internet Source	<1%
63	orca.cardiff.ac.uk Internet Source	<1%

64	Yuxin Chen, Jingyi Huo, Fangru Lin, Hui Yan. "Beyond homophily in spatial–temporal traffic flow forecasting", Neural Networks, 2025 Publication	<1 %
65	"Advances in Knowledge Discovery and Data Mining", Springer Science and Business Media LLC, 2020 Publication	<1 %
66	Fouzi Harrou, Abdelhafid Zeroual, Farid Kadri, Ying Sun. "Enhancing road traffic flow prediction with improved deep learning using wavelet transforms", Results in Engineering, 2024 Publication	<1 %
67	Heng Zhang, Zhizhe Lin, Hai Xie, Jinglin Zhou, Youyi Song, Teng Zhou. "Two-way heterogeneity model for dynamic spatiotemporal traffic flow prediction", Knowledge-Based Systems, 2025 Publication	<1 %
68	Kyungeun Lee, Moonjung Eo, Euna Jung, Yoonjin Yoon, Wonjong Rhee. "Short-term Traffic Prediction with Deep Neural Networks: A Survey", IEEE Access, 2021 Publication	<1 %
69	Senwen Li, Liang Ge, Yongquan Lin, Bo Zeng. "Adaptive Spatial-Temporal Fusion Graph Convolutional Networks for Traffic Flow Forecasting", 2022 International Joint Conference on Neural Networks (IJCNN), 2022 Publication	<1 %
70	Xueyan Yin, Genze Wu, Jinze Wei, Yanming Shen, Heng Qi, Baocai Yin. "Multi-stage attention spatial-temporal graph networks for traffic prediction", Neurocomputing, 2021 Publication	<1 %

71	Ying Ma, Haijie Lou, Ming Yan, Fanghui Sun, Guoqi Li. "Spatio-temporal fusion graph convolutional network for traffic flow forecasting", Information Fusion, 2024 Publication	<1 %
72	Zou Zhene, Peng Hao, Liu Lin, Xiong Guixi, Bowen Du, Md Zakirul Alam Bhuiyan, Yuntao Long, Da Li. "Deep Convolutional Mesh RNN for Urban Traffic Passenger Flows Prediction", 2018 IEEE SmartWorld, Ubiquitous Intelligence & Computing, Advanced & Trusted Computing, Scalable Computing & Communications, Cloud & Big Data Computing, Internet of People and Smart City Innovation (SmartWorld/SCALCOM/UIC/ATC/CBDCom/IOP/SCI), 2018 Publication	<1 %

Exclude quotes	On	Exclude matches	Off
Exclude bibliography	On		