

Article

Weather Interaction-Aware Spatio-Temporal Attention Networks for Urban Traffic Flow Prediction

Hua Zhong¹ , Jian Wang^{1,*}, Cai Chen¹, Jianlong Wang², Dong Li¹ and Kailin Guo¹

¹ School of Geomatics and Urban Spatial Informatics, Beijing University of Civil Engineering and Architecture, Beijing 102616, China; zhonghua_2022@163.com (H.Z.); 1108130422001@stu.bucea.edu.cn (C.C.); ld130173@gmail.com (D.L.); guokailin117@163.com (K.G.)

² Changjiang Space Information Technology Engineering Co., Ltd. (Wuhan), Wuhan 430010, China; wangjianlong@whu.edu.cn

* Correspondence: wangjian@bucea.edu.cn

Abstract: As the cornerstone of intelligent transportation systems, accurate traffic prediction can reduce the pressure of urban traffic, reduce the cost of residents' travel time, and provide a reference basis for urban construction planning. Existing traffic prediction methods focus on spatio-temporal dependence modeling, ignoring the influence of weather factors on spatio-temporal characteristics, and the prediction task has complexity and an uneven distribution in different spatio-temporal scenarios and weather changes. In view of this, we propose a weather interaction-aware spatio-temporal attention network (WST-ANet), in which we integrate feature models and dynamic graph modules in the encoder and decoder, and use a spatio-temporal weather interaction perception module for prediction. Firstly, the contextual semantics of the traffic flows are fused using a feature embedding module to improve the adaptability to weather drivers; then, an encoder–decoder is constructed by combining the Dynamic Graph Module and the WSTA Block, to extract spatio-temporal aggregated correlations in the roadway network; finally, the feature information of the encoder was weighted and aggregated using the cross-focusing mechanism, and attention was paid to the hidden state of the encoding. Traffic flow was predicted using the PeMS04 and PeMS08 datasets and compared with multiple typical baseline models. It was learned through extensive experiments that the accuracy evaluation result is the smallest in WST-ANet, which demonstrated the superiority of the proposed model. This can more accurately predict future changes in traffic in different weather conditions, providing decision makers with a basis for optimizing scenarios.

Keywords: urban traffic flow prediction; weather interaction perception; graph attention network; dynamic spatio-temporal dependence; attention



Citation: Zhong, H.; Wang, J.; Chen, C.; Wang, J.; Li, D.; Guo, K. Weather Interaction-Aware Spatio-Temporal Attention Networks for Urban Traffic Flow Prediction. *Buildings* **2024**, *14*, 647. <https://doi.org/10.3390/buildings14030647>

Academic Editor: Yung Yau

Received: 31 January 2024

Revised: 25 February 2024

Accepted: 26 February 2024

Published: 29 February 2024



Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Due to rapid urban development and the emergence of the population siphoning effect, urban transportation systems have become more complex. The capacity of the existing road network has gradually become unable to handle such a heavy load, which has triggered more transportation-related problems, including traffic congestion and traffic accidents. Traffic congestion [1–3] is an urbanization problem that co-exists worldwide. In order to promote green and sustainable development, traffic forecasting is a key enabler of Intelligent Transportation Systems (ITS) [4–6]. Through excellent data storage and monitoring systems, ITS seeks to be an important guide in traffic management, congestion alleviation, and other transportation-related issues [7]. The inherent non-Euclidean structure of transportation data makes the dependency and complexity of spatio-temporal data challenging [8]. As shown in Figure 1, there are direct or indirect correlations between the nodes in a road network [9]. In terms of spatial structure, nodes in different spatial locations have different impacts on their surrounding nodes. When a node is congested, it can have an impact on the surrounding upstream and downstream neighboring nodes, and

on the traffic flow to and from the outbound lanes. For example, a localized area with the intersection node, Ptr_2 , as the center of congestion radiates to Ptr_1 , Ptr_3 , and Ptr_5 , such that these three nodes are directly affected by the congestion source from Ptr_2 . In the case of a single road segment node, the nodes associated with that node are sparse and the impact caused by localized congestion is not significant. For example, a localized area with the road segment node, Ptr_4 , as the center of congestion radiates to Ptr_3 , causing that node to be directly affected by the congestion from the Ptr_4 congestion source. In terms of the temporal structure, the traffic characteristics of the different nodes change over time due to the dynamic nature of the time series. Typically, the intersection nodes carrying a heavy traffic volume have a stronger traffic timing than the nodes carrying a light traffic volume; during peak hours, the interaction between the nodes on certain roads is stronger than during the rest time. For example, the change of node Ptr_2 at the previous moment $t - 1$ and the next moment t is more significant compared with Ptr_1 .

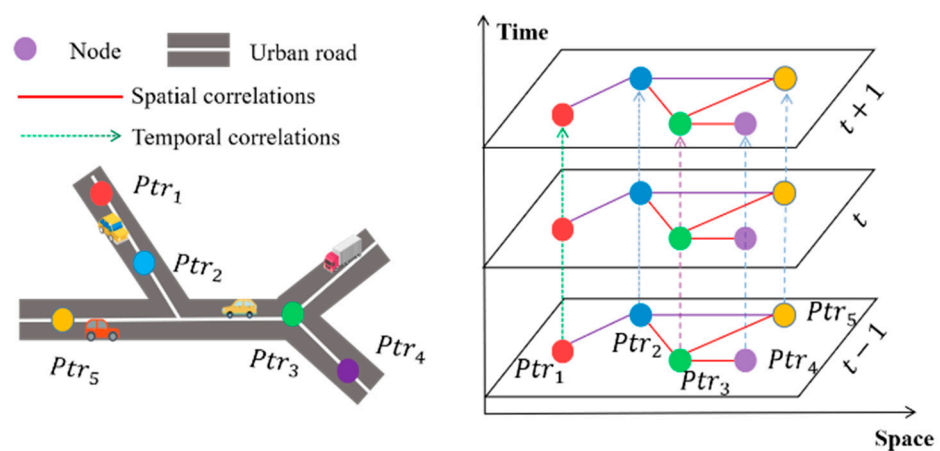


Figure 1. Spatial and temporal relationships of road network nodes. (i) Correlation in space. The flow of a node (e.g., Ptr_2) at the moment is influenced by the flow of its upstream node (Ptr_1) before the moment, and also affects the flow of the downstream nodes (Ptr_3 , Ptr_5) after the moment. (ii) Temporal correlation. The flow of a node at the moment is correlated with its own state before the moment.

From a macro perspective, early methods used traditional time series [10,11] or machine learning models [12] to perform forecasting tasks. These techniques have poor prediction accuracy and are unable to capture detailed, unpredictable spatio-temporal connections [13]. Convolutional Neural Networks (CNNs) [14,15] and Recurrent Neural Networks (RNNs) [16] are two illustrations of deep learning models that have successfully predicted spatio-temporal correlations and have surpassed typical methods in recent years. CNNs cannot be applied to graph-based structured data, which includes traffic network flow prediction; therefore, this is a shortcoming of these means of communication. Recently published studies have targeted traffic prediction as a graph modeling problem, exploiting graphs that represent spatial correlations between nodes and capturing spatio-temporal interactions in traffic networks. The successful application of Graph Convolution Networks (GCNs) in graph processing illustrates how to utilize GCNs in constructing prediction models that perform very well in terms of prediction accuracy. GCNs work mainly by modeling the correlation of the traffic nodes to extract the traffic flow of spatio-temporal features [17]. The adjacency matrix is used in traditional GCNs to simulate the spatial cooperation between multiple nodes. Rather than using the generalized adjacency matrix, a number of investigators have found that the physical characteristic of distance between nodes more accurately reflects the physical distribution of the nodes [18,19]. Despite being extensively used to construct GCNs and their variations, the two referred-to matrices that take node connections into account still have two weaknesses: static feature extraction and short-term feature temporal extraction. Although traffic nodes reside across the road network, there is a stressful logical mapping between the data stored on different nodes.

It is possible to lose the dynamic connection between the nodes when dealing with confusing traffic data if the adjacency matrix or the distance matrix serve to symbolize the complex relationship. In addition, the complex spatial relationship between traffic nodes in different regions not only depends on the distance between the nodes, but also on the external multi-factors that have the same influence on the spatial relationship, such as weather [20,21], POI [22], and social events [23]. Moreover, the time-dependent relationship may not only be associated with a fixed periodicity; the weather factors also exist in the temporal sequence, and this fitted sequence of features has the same effect on the time-dependence of traffic. It becomes difficult to represent the interactions involving nodes as outside forces in the prediction model using purely the traditional adjacency matrix [19]. Therefore, the deep spatio-temporal relationship characteristics embedded in the external factors can be used to dissect the complexity of the transportation road network at a fine-grained level, thus improving the comprehensiveness of traffic forecasts.

Accurate traffic forecasting allows city managers to better plan resource allocation for road development and safety, saving commuters valuable time. To overcome these limitations and accomplish a dynamic and thorough spatio-temporal dependency extraction, we propose a weather interaction-aware spatio-temporal attention network (WST-ANet), in which we integrate feature models and dynamic graph modules in the encoder and decoder, and use a spatio-temporal weather interaction perception module for prediction. In particular, the model can learn the spatio-temporal correlation between the meteorological and transportation nodes, build a spatio-temporal feature and spatio-temporal attention multi-graph network model based on meteorological data sensing, and extract the relevant features using the channel attention mechanism. Meanwhile, we built a spatio-temporal dynamic spatial network that can effectively extract deep dynamic spatial dependencies by repeatedly updating the relational connections. Furthermore, by employing spatio-temporal attention to capture spatio-temporal dependencies, we can apply the attention mechanism to enhance the significance of spatio-temporal information and improve prediction outcomes. The main contributions of this paper are summarized as follows:

- We utilized the proposed the WST-ANet model, which consists of a spatio-temporal attention module for weather interaction sensing and a cross-attention mechanism. The model adaptively simulates the complex spatio-temporal interdependencies between weather and traffic using an encoder–decoder architecture, with an attention mechanism to achieve dynamic spatio-temporal and complex-dependent traffic prediction incorporating weather factors. This allows the model to adaptively focus on the spatial and temporal characteristics of the region, realizing the capture of the dynamic spatial and temporal characteristics of the road network, thus improving the accuracy of the prediction.
- We designed a new DGM, a dynamic graph module that adaptively captures the connectivity relationships of road networks on a spatial level, mines the adaptive graph for hidden information, and extracts the spatial correlations among the nodes step by step in depth. This method updates the adjacency matrix and iterates the aggregation of features to better fit the dynamic scenarios of urban road networks, thus improving the robustness of the prediction.
- We constructed an interactive perception fusion method of weather features and spatio-temporal features. The vector embedding was utilized to fuse temporal features, spatial features, and weather attributes to generate contextual semantics on exogenous weather-driven multimodal feature embedding. This spatio-temporal scenario synergizes weather changes to learn the multimodal urban road network characteristics and comprehensively grasp the weather and spatio-temporal interactive fusion of the traffic conditions in the city.
- In order to validate the effectiveness of the proposed model, we conducted comprehensive comparison experiments and prediction validation with 14 baseline models on two datasets.

The rest of the paper is organized as follows: Section 2 describes the progress made during the different phases of traffic prediction research; Section 3 details the proposed model, as well as the components; Sections 4 and 5 conduct experiments using real traffic data to evaluate the performance of the model and the analysis of results; Section 6 concludes the paper and summarizes the plans for further work in the future.

2. Related Work

Traffic prediction studies are important and instructive in the broader field of transportation and urban planning. The methods used for traffic prediction are endless, from the initial use of statistical methods for time-series prediction, to the later use of neural network models for spatial feature prediction, and now, the integration of deep learning techniques for spatio-temporal prediction, which is more mature.

2.1. Time-Series Traffic Forecasting

The emergence of data streams in the past century has led to rich and extensive research work by researchers on time-series prediction [24]. This temporal characterization of data is applicable to several domains, such as finance, healthcare, and stocks. In particular, the application of such time-series characterization research methods is maturing in the field of traffic prediction [25]. Traditional traffic prediction methods are model-driven, based on the use of linear pure control system theoretical approaches, such as vector auto-regression (VAR) [10], historical averaging models (HAs) [26], autoregressive integrated moving average models (ARIMAs) [27], and Kalman filtering techniques [28]. Usually, these modeling assumptions require the stability of the target data; therefore, this stable structured state cannot effectively manage nonlinear data. In the traffic prediction task, the traffic state of the road network had both temporal and spatial dependencies; moreover, there were complex dynamic spatio-temporal dependencies in the traffic data. Traditional model-driven methods cannot describe the spatio-temporal nature of traffic scenarios.

In addition, deep learning techniques de-constructed through neural network structures, such as recurrent neural networks (RNNs) [16], long short-term memory (LSTM) [29], and gated neural units (GRUs) [30], have shown superior performance in capturing the correlation of temporal units. However, the common problem of these research works is that they only consider the time series and ignore the spatial information on traffic features. Therefore, the single consideration of the temporal sequence of traffic has significant limitations in predicting the wholeness and completeness of the traffic characteristics of a road network. This limitation fundamentally stifles the application of global road networks in transportation.

2.2. Space-Time Traffic Forecasting

With the concept of spatio-temporal features, graph topology modeling for transportation research has become a research point of growing interest. The non-Euclidean structure of urban road networks is modeled through edges and nodes, and these upstream and downstream node relationships enable the effective representation of convolution operations on unstructured nonlinear spatial data. Taking the graph convolution network, GCN [17,31], as an example, this spatial convolutional layer utilizes Fourier transform or Laplace transform for feature aggregation and extraction, and the attention mechanism [32] is added on top of the GCN to generate a constrainable graph attention network GAT [33]. In addition, Guo et al. [34] combined the attention mechanism with spatio-temporal graph convolution using a spatio-temporal graph convolutional network of attention, to construct a convolutional model capable of capturing spatio-temporal features and spatio-temporal dynamic correlations. Song et al. [35], Luo et al. [36], and Li et al. [37] performed research work based on the ASTGCN model. Cirstea et al. [38] proposed a spatio-temporal perceptual attention network (ST-WA), which randomly encoded time series to generate location-specific and time-varying model parameters to better capture spatio-temporal dynamics. Bai et al. [39] proposed an adaptive graph convolutional recursive network (AGCRN) to capture spatio-temporal dynamics through adaptive graph generation and node-adaptive

parameter learning, to augment traditional graph convolution and incorporate it into a neural network with recurrent operation to capture more complex spatio-temporal correlations. Wu et al. [40] proposed a Graph WaveNet combining diffusion convolution and extended stochastic convolution. Zhao et al. [41] considered spatio-temporal depth relations mining at the location level. Bao et al. [42] characterized dynamic spatio-temporal features by constructing complex correlation matrices through multi-feature and attention mechanisms, based on spatio-temporal complex graph convolutional networks. However, these investigations only assessed distance and disregarded the spatio-temporal structure of the graph, as well as the impact of the surroundings components on traffic flow, resulting in the inability to make accurate traffic predictions.

2.3. Spatio-Temporal Traffic Prediction with Embedded Factors

Traffic situations face different variations along with the presence of multi-source heterogeneous environmental factors, such as points of interest (POIs) [22], social events [23], and weather [21,43,44]. Embedding these factors into the traffic scenarios, the spatio-temporal model driven by environmental factors is more relevant to urban traffic, and at the same time, there is an improvement in stability. For example, Geng et al. [45] used POI information to predict spatial features. Zheng et al. [46] constructed a Graph Multi-Man Attention Network (GMAN) using an encoder–decoder structure, which consisted of a module with multiple spatio-temporal attention blocks that simulated dynamic spatial and temporal connections by utilizing attention in the procedures. Zou et al. [47] combined GMAN to conduct a large number of prediction studies. Zhang et al. [48] used one-hot coding and combined temporal information with weather for prediction. Wang et al. [49] constructed a graph attention network that effectively extracted weather-driven spatio-temporal features by convolving weather with spatio-temporal feature modules. However, exploring the dynamic spatio-temporal dependency between traffic and weather, the model of utilizing this complex dependency in traffic prediction remains unresolved effectively.

3. Methodology

3.1. Problem Description

The basic case of traffic forecasting is to make scientifically sound predictions of future traffic conditions based on historical data. Traffic flow can be quantified to reflect the complexity between roads, relying on the basic theory of graphs, and traffic prediction, by considering the city as a network structure. Therefore, the traffic road network can be symbolized as an undirected graph, $G = (V, E, A)$, where $V = \{v_1, v_2, \dots, v_N\}$ denotes the collection of road nodes in the road network structure, and N denotes the number of road nodes. E denotes a collection of edges connecting the different sensors, which can reflect the association between the road network sections. All the connectivity information is stored in the abstract primitive adjacency matrix, $A \in R^{N \times N}$, which is used to measure the spatial correlation dependency between several nodes. The corresponding element in the adjacency matrix is 1 when nodes i and j are adjacent to each other, and the corresponding element in the adjacency matrix is 0 when nodes i and j are not adjacent to each other. The traffic condition at time step, t , is represented as the graph signal, $X_t \in \mathfrak{R}^{N \times C}$, on the graph, G , where C is the number of characteristics of the road condition to be reviewed (e.g., traffic flow, traffic speed). In the present research, only the traffic flow was investigated.

Given the traffic network, G , the historical traffic flow, $\chi_h = [X_{H+1}, \dots, X_{H+P}] \in \mathfrak{R}^{P \times N \times C}$, predicts the future traffic flow, $\hat{\chi}_p = [X_{H+P+1}, \dots, X_{H+P+Q}] \in \mathfrak{R}^{Q \times N \times C}$, with time step T_p on the basis of the time step, T_H , of the given traffic network, G . This prediction process was reflected by a certain mapping function relation, f . The mapping relation was expressed by the following equation:

$$\hat{\chi}_p = f(\chi_h, G) \quad (1)$$

3.2. Framework Overview

Figure 2 shows the structure of our proposed WST-ANet model, which is considered to take an encoder–decoder structure. Both the encoder and decoder were constructed from the L-layer Weather Spatio-Temporal Attention Block (WSTA Block) and the Dynamic Graph Module (DGM). Since different methods are needed to extract the spatio-temporal features perceived by weather data, we designed the embedding pattern to introduce the weather data and realize the fusion of weather and spatio-temporal features. For each Attention block, there is a spatial weather attention, a temporal weather attention, and a dynamic graph module (DGM). To diminish the implications of error propagation, Cross-Attention was divided between the encoder and decoder for traffic speed and conditions at preceding and future time steps. The characteristics of these blocks are described below.

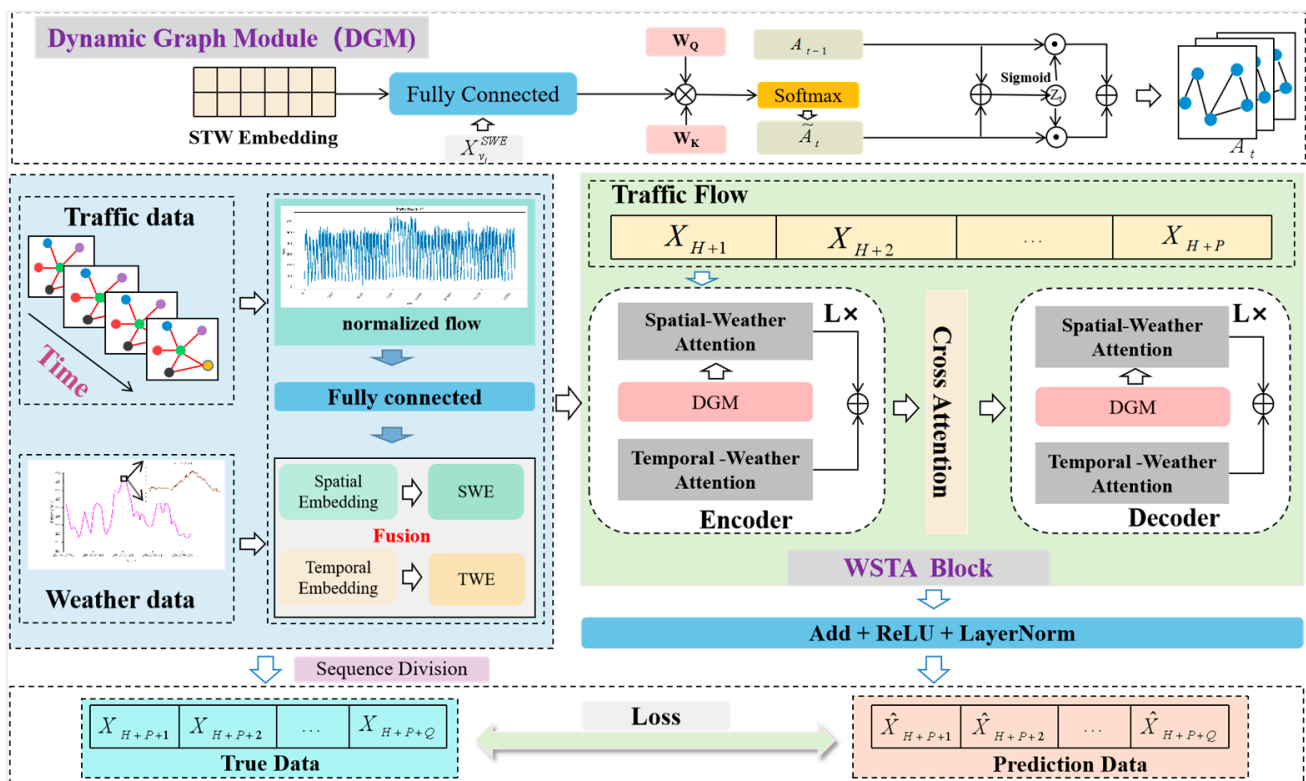


Figure 2. WST-ANet model architecture.

3.3. Feature Embedding Module

Since changes in traffic road conditions are constrained by the original road network and likewise by the weather, it is of great importance to consider road spatio-temporal information and weather drivers in forecasting.

In order to capture the features of the road network, the node2vec method [50] was used to learn the vertex representation, convert the graph nodes (i.e., traffic sensors) into feature embedding, and introduce a spatial embedding module to convert the spatial sensor data into vectors. These vectors then provided information in a two-layer fully connected neural network (FC) to generate spatial embedding, $e_{v_i}^{SE} \in \mathbb{R}^D$, where the sensor nodes $v_i \in V$. To facilitate connectivity, all layers produced D-dimensional outputs. The embedding stage goes through the Fully-Connected neural network to achieve a learnable spatial embedding and allow data to flow more easily across the model.

The current state of traffic fluctuates across time, and a single spatial embedding is static and insufficient to reflect the dynamic changes in traffic conditions. Therefore, we designed a time embedding module to construct the time information into vectors using solo thermal coding. The encoding principle divides a day into T time steps, and uses

independent heat encoding to encode every single day of the entire week and every other step of the day as \mathfrak{R}^7 and \mathfrak{R}^T , respectively. The two were connected into a whole vector, \mathfrak{R}^{T+7} , and the vector, \mathfrak{R}^{T+7} , was input into the FC to obtain the time domain embedding $e_{t_j}^{TE} \in \mathfrak{R}^D$, where $t_j \in \{t_1, \dots, t_P\}$ was used as a floating window for each time series, containing the past time steps in length.

Influenced by weather perception, weather factors cause the subjective traveling and driving behavior of people, which can have different impacts on traffic. For example, drivers tend to increase their speed on sunny days, while they tend to decrease their speed in heavy rainstorms, snow, and other weather. Combining this property of weather-driven transportation, we present a weather-embedding module that uses weather data from Weather Spark and encodes the weather information into vectors that adapt to the precise location of each sensor using solo thermal coding. Finally, we applied FC to these vectors to obtain the weather embedding, where $e_{v_i, t_j}^{WE} \in \mathfrak{R}^D$, $v_i \in V, t_j \in \{t_1, \dots, t_P\}$. Traffic characteristics in spatial scenarios are affected by weather factors. To capture the impacts of changes in the weather on space transportation, we constructed Spatial Weather Embedding (SWE), in which the weather–space module fuses spatial traffic flow features with weather characteristics to generate a spatial traffic flow feature with weather awareness. For example, the spatial weather embedding feature of the spatial node, v_i , where sensor i is located at time step t_j can be expressed as e^{SWE} , as follows.

$$e_{v_i, t_j}^{SWE} = e_{v_i}^{SE} + e_{v_i, t_j}^{WE} \quad (2)$$

where, $v_i \in V, t_j \in \{t_1, \dots, t_P, \dots, t_{P+Q}\}$, and $e^{SWE} \in \mathfrak{R}^{(P+Q) \times N \times D}$ denote the SWE of traffic road network nodes in the process of time step $P + Q$. To accurately replicate the geographical link between weather and traffic, the SWE features were input into the weather space attention for adaptive feature mining.

Due to the existence of explicit temporal features between both time and weather, the embedded traffic conditions and weather conditions vary over time. Therefore, we designed the Time-Weather Embedding (TWE), where the time–weather module fuses the temporal features of the traffic flow with the weather features to form a time-series traffic feature structure that includes weather awareness. At time step, t_j , the temporal weather embedding feature of the spatial node v where sensor i is located was expressed as e^{TWE} , as follows.

$$e_{v_i, t_j}^{TWE} = e_{t_j}^{TE} + e_{v_i, t_j}^{WE} \quad (3)$$

where, $v_i \in V, t_j \in \{t_1, \dots, t_P, \dots, t_{P+Q}\}$, $e^{TWE} \in \mathfrak{R}^{(P+Q) \times N \times D}$ denotes the TWE of the N traffic road network nodes in the process of time step $P + Q$. In order to effectively simulate the spatial relationship between weather and traffic, the TWE features were input into the Temporal Weather Attention for adaptive feature mining.

In order to obtain dynamic adaptive spatio-temporal weather changes, the above embedding module was subjected to spatio-temporal weather embedding to form the Spatio-temporal Weather Embedding Module (STWE), which fused the spatio-temporal traffic features of weather. When the time step was t_j , the temporal weather embedding feature of the spatial node v_i where sensor i is located was expressed as e^{STWE} .

$$e_{v_i, t_j}^{STWE} = e_{t_j}^{TE} + e_{v_i}^{SE} + e_{v_i, t_j}^{WE} \quad (4)$$

where, $v_i \in V, t_j \in \{t_1, \dots, t_P, \dots, t_{P+Q}\}$, $e^{STWE} \in \mathfrak{R}^{(P+Q) \times N \times D}$.

Combining the input features, χ_h and e_h^{STWE} , the final input embedding feature E was as follows.

$$E = \text{Concat}(\chi_h, e_h^{STWE}) \in \mathfrak{R}^{P \times N \times D} \quad (5)$$

3.4. Dynamic Graph Module

In this part, we created a dynamic graph process module. This module aims to generate an adaptive adjacency matrix, representing the iterative update of the dynamic spatial correlation of the traffic road network by fusing the dynamic features extracted from the road attributes by means of a fusion mechanism. This is used to symbolize the weather-driven spatial characteristics of the road network in an immediate response in the dynamic process. Because the dynamic spatio-temporal correlation depends mainly on the real-time traffic state, it is necessary to simulate the dynamic spatial correlation by entering the real-time traffic state. The goal of generating a dynamic feature matrix was to be certain that each of the input dynamic, latent spatial, and temporal information was written properly. To accomplish this, the nodes were first represented as embeddings, merging the current traffic state, X^{SWE} , at each time step, including time-of-day and time-of-week embeddings. To extract the latent spaces of weather interaction properties between the nodes even further, the features of the traffic state, X_t , were extracted using two nonlinear fully connected layers and dimensionally transformed. At time step, t , two dynamic feature matrices were generated and absorbed.

$$E_{DF} = FC(X^{SWE}) \quad (6)$$

where $E_{DF} \in \mathfrak{R}^{P \times N \times D}$, $FC(\cdot)$ denotes the network of the two nonlinear fully connected layers. Then, the parameters K , Q , and V of the self-attention were computed.

$$Q = E_{DF}W_Q, K = E_{DF}W_K, V = E_{DF}W_V \quad (7)$$

where the learnable parameter $W_Q, W_K, W_V \in \mathfrak{R}^{D \times D}$. Then, using the self-attention mechanism, the dynamic feature matrix for the current time step was computed and publicized as follows.

$$\tilde{A}_t = \text{Softmax}\left(\frac{QK^T}{\sqrt{d_h}}\right). \quad (8)$$

where $\tilde{A}_t \in \mathfrak{R}^{N \times N}$ denotes the correlation of the road nodes at the time step, thus allowing each dynamic feature matrix to learn the neighboring features that reflect the traffic topology after the input time step. For the continuous time intervals, a gating mechanism was used for feature extraction, where spatial topology information was extracted using the current dynamic feature matrix, \tilde{A}_t , and the dynamic adjacency matrix, A_{t-1} , of the previous time step.

$$z_t = \text{sigmoid}(\tilde{A}_t W_{\tilde{A}_t} + A_{t-1} W_{A_{t-1}}). \quad (9)$$

where $W_{\tilde{A}_t}, W_{A_{t-1}} \in \mathfrak{R}^{N \times N}$ denotes the two learnable transformation matrices, which ultimately results in a dynamic adjacency matrix, A_{SW} . The metric was as follows.

$$A_{SW} = \begin{cases} \tilde{A}_0, & t = 0 \\ z_t \circ \tilde{A}_t + (1 - z_t) \circ A_{t-1}, & t > 0 \end{cases} \quad (10)$$

where \circ denotes the Hadamard product, i.e., the multiplication of the elements corresponding to two matrices of the same dimension.

3.5. Temporal-Weather Interactive Module

In order to efficiently capture the spatio-temporal correlations of weather interaction perception from traffic sequences, this paper proposes a WSTA block structure as shown in Figure 3. Specifically, the coder–decoder for spatio-temporal weather interaction perception includes a spatial weather transformer and a temporal weather transformer, as well as a dynamic graph module for processing dynamic adaptive correlations. This process features the inputs of spatial weather embedding and temporal embedding, respectively, and ultimately iteratively updates to obtain dynamic spatio-temporal features, including

weather-driven features. The spatial weather transformer was used to capture the weather-driven spatiality of the transportation road network. Within a transportation spatial region, the traffic condition for a single road segment is influenced by the traffic patterns of other road segments and operates with the dynamic structure of the road network. Correspondingly, the weather in a region is also affected by conditions outside the region, e.g., weather conditions between regions affect the traffic conditions. In order to utilize the interactive features of space and weather, we provided the Spatial-Weather Attention (SWA) framework to capture the correlation information between road sensors in a variable way.

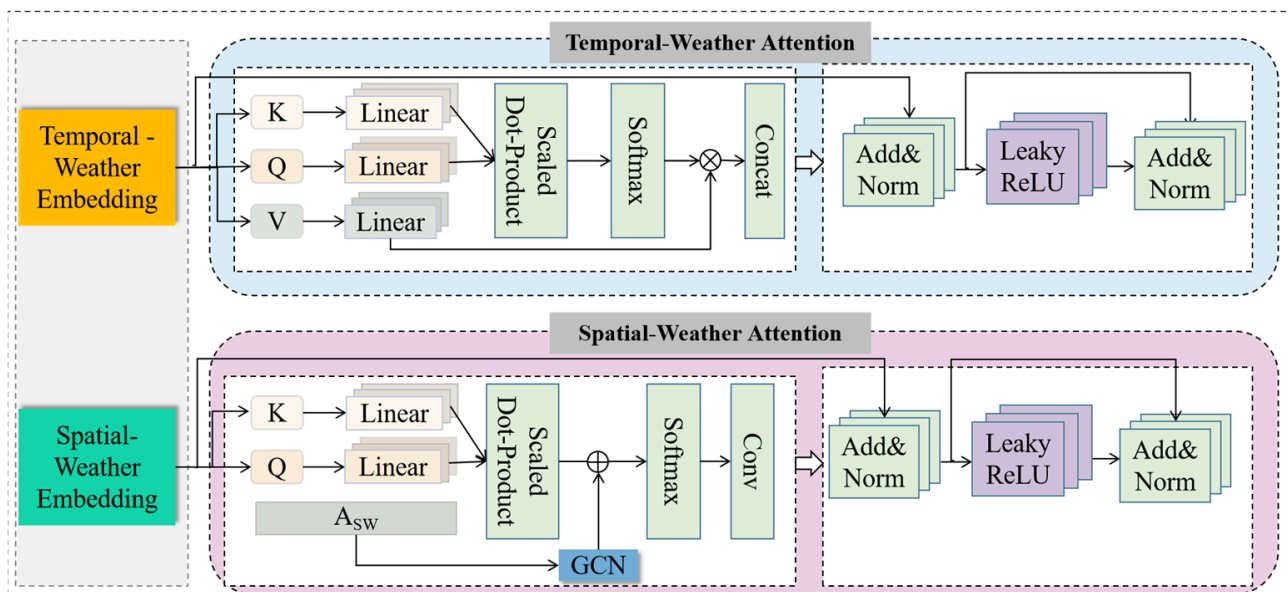


Figure 3. The WSTA block structure.

After obtaining the dynamic neighbor matrix features, the graph dynamic neighbor matrix A_{SW} was formed using distance construction and then input to the graph convolutional neural network. The l th layer of the input was defined as $H^{(l-1)}$, where the hidden state of vertex v_i at time step t_j was $h_{t_j}^{l-1}$. By capturing the spatial features between the nodes through their first-order neighborhoods, the GCN model was then built by stacking many convolutional layers, each of which can be shown through the symbol as:

$$\hat{H}_i^l = \Phi_{GCN}(A_{SW}, H^{(l-1)}) = \sigma(\tilde{\Lambda}^{-\frac{1}{2}}(A_{SW} + I_N)\tilde{\Lambda}^{-\frac{1}{2}}H^{(l-1)}\theta^{(l-1)}). \quad (11)$$

where I_N denotes the unit matrix, $\tilde{\Lambda}$ denotes the degree moment of the transportation road network, $H^{(l-1)}$ denotes the output of layer $l-1$, $\theta^{(l)}$ denotes the hyperparameters of layer $l-1$, and $\sigma(\cdot)$ denotes the softmax activation function model.

Using the adjacency matrix of the dynamic graph module, the dynamic weights of the spatial convolutional layer were expressed. These weights were then utilized to build the weight summation function, and the fine-grained edge elements were expressed as:

$$hs_{v_i,t}^l = \sum_{v \in V} \phi_{v_i,v} \cdot h_{v,t}^{l-1}, \quad \sum_{v \in V} \phi_{v_i,v} = 1 \quad (12)$$

To capture the relationships between the surrounding vertices, we chose the Multihead Attention score to assign ϕ . The correlation between vertices v_i and v was computed using the scaled dot product method and normalized using the activation function:

$$s_{v_i,v}^k = \frac{\langle f_{s,1}^k(h_{v_i,t}^{l-1} \| \phi_{v_i,t}^{SWE}), f_{s,2}^k(h_{v_i,t}^{l-1} \| \phi_{v_i,t}^{SWE}) \rangle}{\sqrt{d}}, \quad (13)$$

$$\phi_{v_i,v}^k = \frac{\exp(\text{LeakyReLU}(s_{v_i,v}^k))}{\sum_{v_i \in V} \exp(\text{LeakyReLU}(s_{v_i,v}^k))} \quad (14)$$

where $\phi_{v_i,v}$ denotes that V is the attention score that represents the correlation between the vertex v_i and v , \parallel denotes the join operation, and $\langle \cdot \rangle$ denotes the inner product operator. The hidden state of each vertex was then updated at each time step as:

$$hs_{v_i,t}^l = \parallel_{k=1}^K \left\{ \sum_{v \in V} \phi_{v_i,v}^k \cdot f_{s,3}^k(h_{v,t}^{l-1}) \right\} \quad (15)$$

where $f_{s,1}^k(\cdot)$, $f_{s,2}^k(\cdot)$, and $f_{s,3}^k(\cdot)$ denote the three different nonlinear projections:

$$f(x) = \text{ReLU}(xW + b) \quad (16)$$

where W and b are the learnable parameters and ReLU is the activation function. Using K -parallel computing, the learnable parameters for all vertices and time steps were shared. Each head attention generated a $d = D/K$ dimensional output.

The temporality of the weather-driven traffic road networks was captured using temporal weather transformers. Traffic and weather patterns alter throughout time. We proposed Time-Weather Attention (TWA) to adaptively simulate weather and time-traffic correlations between distinct time steps in order to capture these connections. We calculated the attention score of vertex, v_i , between time steps t_j and t , and then normalized it using softmax, which was expressed using the following equation:

$$u_{t_j,t}^k = \frac{\langle f_{s,1}^k(h_{v_i,t_j}^{l-1} \| e_{v_i,t_j}^{\text{TWE}}), f_{s,2}^k(h_{v_i,t}^{l-1} \| e_{v_i,t}^{\text{TWE}}) \rangle}{\sqrt{d}}, \quad (17)$$

$$\gamma_{t_j,t}^k = \frac{\exp(\text{LeakyReLU}(u_{t_j,t}^k))}{\sum_{t_j \in N_{t_j}} \exp(\text{LeakyReLU}(u_{t_j,t}^k))} \quad (18)$$

where $u_{t_j,t}$ denotes the correlation between time step t_j and t , γ denotes the attention score of the k th head of the importance of time step t for t_j , and N_{t_j} denotes the set of time steps before t_j . Given the obtained attention score γ^k , we computed the hidden state as:

$$ht_{v_i,t}^l = \parallel_{k=1}^K \left\{ \sum_{v \in V} \gamma_{t_j,t}^k \cdot f_{s,3}^k(h_{v,t}^{l-1}) \right\} \quad (19)$$

where $f(\cdot)$ has the same meaning as above.

The temporal weather features and spatial weather features were obtained by calculating the spatial and temporal effects of the weather, which were fused adaptively according to each vertex and time step. Denoting the outputs of SWA and TWA as H_{SW}^l and H_{TW}^l , respectively, the entanglement, g , as well as the fused feature result, H , were obtained using a linear model with S-type activation:

$$g = \delta(H_{SW}W_{SW} + H_{TW}W_{TW} + b_g). \quad (20)$$

$$H = g \circ H_{SW} + (1 - g) \circ H_{TW}. \quad (21)$$

where $W_{SW} \in \mathfrak{R}^{D \times D}$, $W_{TW} \in \mathfrak{R}^{D \times D}$ denotes the learning parameters, δ denotes the sigmoid activation function, and \circ denotes the Hadamard product, i.e., the multiplication of the elements corresponding to two matrices of the same dimension.

3.6. Cross-Attention

The cross-attention mechanism [51] is a method of weighting different positions in the input sequence to extract key information using a self-attention mechanism. In contrast to recurrent neural networks, which process the input sequence sequentially and step-by-step, Transformer instead pays attention to the information at all positions simultaneously using the self-attention mechanism, which greatly reduces the time complexity of processing long sequences. A frequent method in sequence prediction challenges is to predict sequences one step at a time, using the result of the previous step as the input for the following prediction. However, this approach can cause errors in accumulation between different prediction steps in long-term prediction [39,46,52]. We proposed a cross-self-attention module between the encoder and the prediction decoder to solve this limitation. The cross-attention module symbolizes the link between each historical time step and all future predicted time steps directly.

Formally, given the history time step $t_j^P \in \{t_1, \dots, t_P\}$, the future prediction time step $t_j^F \in \{t_{P+1}, \dots, t_{P+Q}\}$, and the sensor v , the history step P and the future step Q are denoted as $E_h = e^{SWE}[v_i, :] + e_h^{TWE}$ and $E_F = e^{SWE}[v_i, :] + e_F^{TWE}$, respectively. The dimensional reconstruction was carried out, and the reconstructed data of the two features were lost to the cross-attention mechanism for computation to obtain the following:

$$\text{CrossAttention}(E_h, E_F) = \text{softmax}\left(\frac{\langle E_h W_1, (E_F W_2)^T \rangle}{\sqrt{D}}\right) E_F W_3 \quad (22)$$

where $W_1 \in \mathbb{R}^{P \times D}$, $W_2 = W_3 \in \mathbb{R}^{Q \times D}$ is the learned projection matrix.

3.7. Loss Function

In this study, we utilized the Huber loss function [53] to train the model, such that the prediction results were as near to the actual traffic situation as feasible. As a result, the goal of the loss function was to reduce the prediction error. Considering the prediction process, the long process prediction had the risk of multi-source error accumulation; therefore, in this paper, the length of the prediction window was set to P . The prediction result was denoted as $\hat{Y}_{pre} = [\hat{X}_{H+P+1}, \dots, \hat{X}_{H+P+Q}]$ and the real result was denoted as $Y_{true} = [X_{H+P+1}, \dots, X_{H+P+Q}]$. The loss function was denoted as:

$$\text{loss}(\hat{Y}_{pre}, Y_{true}) = \begin{cases} \frac{1}{2}(\hat{Y}_{pre} - Y_{true})^2 & |\hat{Y}_{pre} - Y_{true}| \leq 1 \\ |\hat{Y}_{pre} - Y_{true}| - \frac{1}{2} & \text{otherwise} \end{cases} \quad (23)$$

4. Experimentation

In this section, in order to verify the validity and generalizability of the model, we conducted experiments on two public datasets generated in real traffic scenarios. Before the experiments, we first describe the experimental setup, including the datasets used, the parameter settings, the baseline comparisons, and the evaluation metrics. We then present a series of studies, including ablation experiments, to illustrate the impact of the model components on the overall results. Finally, we present a detailed analysis and interpretable representation of the experimental results.

4.1. Experimental Datasets

In this study, traffic data from two regions, PeMS04 and PeMS08, were used (<http://pems.dot.ca.gov>, accessed on 1 July 2023). The PeMS data were located on the freeways of major metropolitan areas in California, and the traffic information of the city was obtained from the deployed traffic detectors. The detailed statistics are shown in Table 1.

The acquired data needed to be preprocessed, and the processing method was consistent with Bi-STAT [54]. The Z-score standard method was applied to normalize the data stream, and the goal of normalization was to uniformly map the data to be studied between the

one-dimensional intervals $[0, 1]$, which unified the data of different magnitudes to the same magnitude and ensured comparability among the data. The formula used is as follows.

$$x_{new} = \frac{x - \mu}{\pi} \quad (24)$$

where x is the original series data, μ is the sample mean, and π represents the standard deviation of the data, which becomes 0 for the mean and 1 for the variance after normalizing the data.

Table 1. Dataset statistics.

Datasets	Nodes	Edges	Time Interval	Duration	Time Steps Length
PeMS04	307	340	5 min	1 January 2018~28 February 2018	16,992
PeMS08	170	296	5 min	1 July 2016~31 August 2016	17,856

4.2. Parameter Settings

The experiments set the target time step, Q , and the historical time step, P , to 12 (indicating a time span of 60 min). After several training steps, the model framework hyperparameters were finalized. The dataset was divided according to three parts: the training set, the validation set, and the test set, where the proportions were 60%, 20%, and 20%, respectively. The range of hyperparameters was set manually based on experience: the learning rate included 0.01, 0.005, 0.001, and 0.0005; the dropout included 0.0, 0.1, 0.2, 0.3, 0.4, and 0.5; and the decay rate included 0.99, 0.95, 0.90, and 0.85. The effect of each different parameter combination was tried by loop traversal, and the hyperparameters with the best performance on the validation dataset were finally selected as the execution results. For these consecutive hyperparameter values, the sampling was performed at equal intervals. For each set of hyperparameters, the optimal parameter was determined using the minimum MAE of the validation set, and specific processing was performed.

For our WST-ANet model, the following settings were found to work best: set the dropout to 0.5, the decay rate to 0.99, and the learning rate to 0.001. When the prediction performance of the prediction model on the validation set was optimal, all the samples in the test set were iterated, and after several parameter adjustments and experiments, the training process ended to obtain the prediction results. The experiments were performed using the Windows operating system. The system was configured using Intel(R) Xeon(R) Gold 6226R CPU @ 2.90GHz, GPU: NVIDIA GeForce RTX 3090, compiled under PyCharm IDE, and based on the PyTorch framework and Python3.11 to realize the training and prediction of the model; the same framework was used for the comparison of the baseline experiments. The baseline experiments were also realized using the same framework.

4.3. Baselines

- (1) HA [26]: Prediction of traffic flow in future time slots by averaging a sequence of a fixed number of terms computed from the historical traffic flow.
- (2) LSTM [29]: Long Short-Term Memory Network, a special RNN model that processes longer sequences of signal data through input gates, output gates, and forgetting gates.
- (3) GRU [30]: Gated Recurrent Unit Network, a special RNN model that optimizes the parameter structure within the network to improve the convergence performance of processing sequence data.
- (4) GCN [31]: Graph Convolutional Network, which abstracts the traffic road network as a graph structure, aggregates the feature information between neighboring nodes through the graph convolution mechanism, and realizes feature update of traffic data between domains.
- (5) GAT [32]: Graph Attention Network, based on (3), the attention mechanism is introduced between nodes, so that each node can be adaptively weighted according to the features of its neighboring nodes; this adaptation can effectively aggregate and process the complex feature relationships between traffic data nodes.

- (6) DCRNN [36]: Bidirectional spatio-temporal adaptive transformer, which adopts encoder–decoder architecture and an adaptive mechanism to construct a spatio-temporal feature information extraction structure.
- (7) AGCRN [38]: Adaptive graph convolution recurrent network that enhances traditional graph convolution through adaptive graph generation and node-adaptive parameter learning, integrating into recurrent neural networks to capture more complex spatio-temporal correlations.
- (8) ST-CGCN [41]: A spatio-temporal complex graph convolutional network based on constructing complex correlation matrices through multi-feature and attention mechanisms to characterize dynamic spatio-temporal features.
- (9) GMAN [45]: Employs an encoder–decoder structure containing multiple spatio-temporal attentional blocks that use attentional mechanisms to model dynamic spatial and temporal correlations.
- (10) ST-WA [37]: A spatio-temporal perceptual attention network that randomly encodes time series to generate site-specific and time-varying model parameters to better capture spatio-temporal dynamics.
- (11) STPGCN [40]: A Spatio-Temporal Location-aware Graph Convolutional Network, which adaptively infers the correlation weights of three important spatio-temporal relationships through a spatio-temporal location-aware relationship inference module, aggregating and updating the node features to capture node-specific model features guided by location embedding.
- (12) ASTGCN [33]: Attention-based spatio-temporal graph convolutional network, combining the attention mechanism and spatio-temporal graph convolution to construct a convolutional model capable of capturing the spatio-temporal features to capture spatio-temporal dynamic correlations.
- (13) STSGCN [34]: A graph convolution spatio-temporal network based on a road network structure, using a graph convolution method to capture complex local spatio-temporal correlations, modeling spatio-temporal heterogeneity for mutually independent components.
- (14) AFDGCN [35]: A novel dynamic graph convolutional network with attention fusion functionality, which jointly models synchronous spatio-temporal correlations through a dynamic graph learner and a GRU.

4.4. Evaluation Metrics

This experiment used mean absolute error (MAE), root mean square error (RMSE), and mean absolute percentage error (MAPE) to evaluate the prediction performance of the model, and demonstrate the quality of the prediction effect by comparing the MAE, RMSE, and MAPE results. MAE, RMSE, and MAPE assess the accuracy of traffic flow prediction from different perspectives, integrating the magnitude, distribution, and relative value of prediction errors; therefore, they are widely used.

RMSE measures the dispersion of an observation and its true value by averaging the sum of the squares and then taking the square root of the two. The formula is as follows:

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad (25)$$

MAE is expressed through the mean of the absolute errors, which better reflects the prediction value errors. The formula is as follows:

$$MAE = \frac{1}{n} \sum_{i=1}^n |\hat{y}_i - y_i| \quad (26)$$

MAPE is used as a relative measure to compare the accuracy of model predictions using absolute values to avoid positive and negative errors canceling each other out. The formula is as follows:

$$MAPE = \frac{1}{n} \sum_{i=1}^n \left| \frac{\hat{y}_i - y_i}{y_i} \right| \quad (27)$$

where y_i denotes the observed true value at time step i , \hat{y}_i denotes the model predicted value at time step i , and n denotes the number of time steps.

5. Analysis of Results

5.1. Performance Analysis

After studying the different model algorithms, it was found that, among the models listed in Table 2, the HA model performed lower than the other benchmark models, reflecting the difficulty of urban traffic flow prediction in the long term, which also reflects the fact that deep learning methods perform better than statistical methods. For example, in the PeMS04 dataset, the MAE of LSTM, GRU, GCN, GAT, DCRNN, AGCRN, ST-CGCN, GMAN, ST-WA, STPGCN, AFDGCN, STSGCN, ASTGCN, and WST-Anet was lower than HA by 23.981%, 36.76%, 8.388%, 10.018%, 34.457%, 46.989%, 45.341%, 48.961%, 49.382%, 49.093%, 30.450%, 44.097%, 41.703%, and 51.124%, respectively, which indicates that the performance evaluation indexes of HA, MAE, RMSE, and MAPE performed the worst. In addition, the MAE of GRU was 16.811%, which was 7.430% lower than that of LSTM in the PeMS04 and PeMS08 datasets, indicating that GRU is better than LSTM in capturing temporal correlation. This is because the HA processing performance is limited by the structure when dealing with nonlinear time series, and these deep learning methods are more suitable for extracting the correlation between nonlinear traffic data. The MAE and RMSE of the GAT model on the PeMS04 dataset were 1.780% and 3.557% lower than that of the GCN model, and the MAE and RMSE were 0.856% and 2.219% lower than the GCN model, respectively. As a result, GAT added an attention technique to aggregate feature information in graph topologies and outperformed GCN in prediction performance. However, the GCN and GAT models can only capture the spatial correlation of traffic flow data, and learning characteristics in the time dimension is problematic. To learn the temporal correlation of traffic flow data, the GCN and GAT models must be paired with a temporal prediction model or a temporal feature learning module.

Table 2. Comparison of the average performance of different models on PeMS04 and PeMS08.

Module	PEMS04			PEMS08		
	MAE	RMSE	MAPE (%)	MAE	RMSE	MAPE (%)
HA	38.03	59.24	27.88%	34.86	52.04	24.07%
LSTM	28.91	37.93	33.31%	23.15	34.46	21.86%
GRU	24.05	35.51	24.88%	21.43	25.58	21.59%
GCN	34.84	51.43	25.45%	35.14	49.12	22.26%
GAT	34.22	50.99	25.07%	33.89	48.03	23.32%
DCRNN	24.93	36.38	15.48%	17.86	27.84	11.46%
AGCRN	20.16	32.12	11.42%	16.77	27.28	11.99%
ST-CGCN	20.79	33.62	14.21%	17.84	26.43	11.37%
GMAN	19.41	31.06	13.55%	14.51	24.68	10.45%
ST-WA	19.25	28.54	13.05%	14.44	23.61	11.32%
STPGCN	19.36	30.97	11.75%	14.53	24.62	9.94%
AFDGCN	26.45	37.50	14.46%	19.09	31.01	12.62%
STSGCN	21.26	33.68	13.96%	17.44	26.82	11.01%
ASTGCN	22.17	35.69	16.45%	18.88	29.17	11.34%
WST-Anet (ours)	18.59	30.03	11.79%	13.92	24.04	10.39%

Spatio-temporal correlation is an important factor in traffic prediction. The prediction accuracy of GCN and GAT is lower than that of DCRNN, AGCRN, ST-CGCN, GMAN, ST-WA, STPGCN, AFDGCN, STSGCN, ASTGCN, and WST-ANet because these models only consider spatial correlation and ignore the influence of the spatio-temporal features of traffic data. DCRNN combines GCN-based diffusion convolution with GRU to build a diffusion convolution gating unit that learns the temporal correlation of traffic flow data and outperforms AFDGCN in prediction performance. Compared with DCRNN, ST-CGCN and WST-ANet add robust factors to constrain the robustness of the model, and obtain a better prediction performance. The impact of considered spatio-temporal features on traffic flow was investigated in terms of deep fine-grained spatio-temporal dependence and dynamic aggregation by AGCRN, ST-WA, STSGCN, ASTGCN, STPGCN, and WST-ANet. As shown in Figure 4, for example, in the PeMS08 dataset, WST-ANet improved the MAE of AGCRN, ST-WA, STSGCN, STPGCN, and ASTGCN by 20.47%, 4.07%, 3.74%, 4.38%, 18.94%, and 26.68% than that of AGCRN, ST-WA, STSGCN, STPGCN, and ASTGCN, which suggests that our proposed model was relatively good in terms of performance.

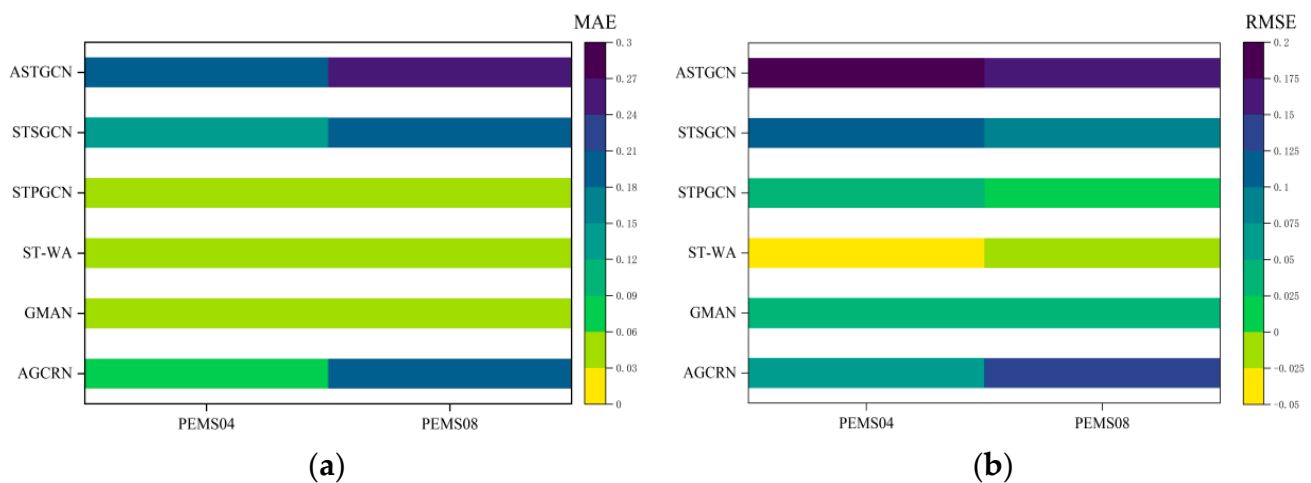


Figure 4. (a) denotes the change in MAE of different models compared to WST-ANet model; (b) denotes the change in RMSE of different models compared to WST-ANet model.

5.2. Analysis of Predicted Results

Different models differ in the short-term prediction and long-term prediction processes. The spatio-temporal-dependent models based on GATs and attentional mechanisms include ASTGCN, ST-WA, and AFDGCN. Compared with ST-WA, the prediction accuracies of ASTGCN and AFDGCN are lower at 12 prediction steps due to the non-Euclidean structure of the highway traffic data. In addition, the coupled mode of spatial correlation and temporal dependence may lead to the better prediction performance of STPGCN than ASTGCN. Compared with the baseline model, the WST-ANet model in this paper took into account the drive of external weather factors on the basis of the existing spatio-temporal dependence correlation. It learnt the more pertinent and deeper dependence of the traffic nodes in terms of spatial distances, adaptive dynamics changes, and weather interaction sensing relationships, while utilizing spatio-temporal feature embedding and spatial attention mechanisms to capture the dynamics between nodes.

Therefore, the WST-ANet model predicts better than the baseline model in the short time-series prediction process. For example, in the traffic flow prediction task for the PeMS04 dataset in Figure 5 at the 6th prediction step (30 min ahead), the MAE predicted by WST-ANet was compared with the AGCRN, ST-WA, STSGCN, STPGCN, and ASTGCN models, increased by 18.949%, 8.598%, 27.389%, 5.326%, and 13.505%, respectively. RMSE increased by 16.194%, 7.699%, 20.446%, 2.986%, and 8.836%, respectively, and MAPE was different in the first eight prediction steps. With the degree of improvement, starting from the 9th prediction step, the performance of ST-WA was better than that of WST-ANet

in this article. Similarly, in the PeMS08 dataset, the same problem existed. Starting from the 6th prediction step, the MAPE of ST-WA was a bit better than that of WST-ANet. Overall, the WST-ANet model proposed in this paper integrates weather factors with spatio-temporal features interactively sensed. The short-term traffic prediction considering the exogenous factors of weather is more conducive to practical applications, e.g., the more relevant short-term traffic prediction can inhibit the accumulation of the error in the long-term traffic process, enabling the transportation agencies to optimize the traffic scheme based on the prediction more accurately.

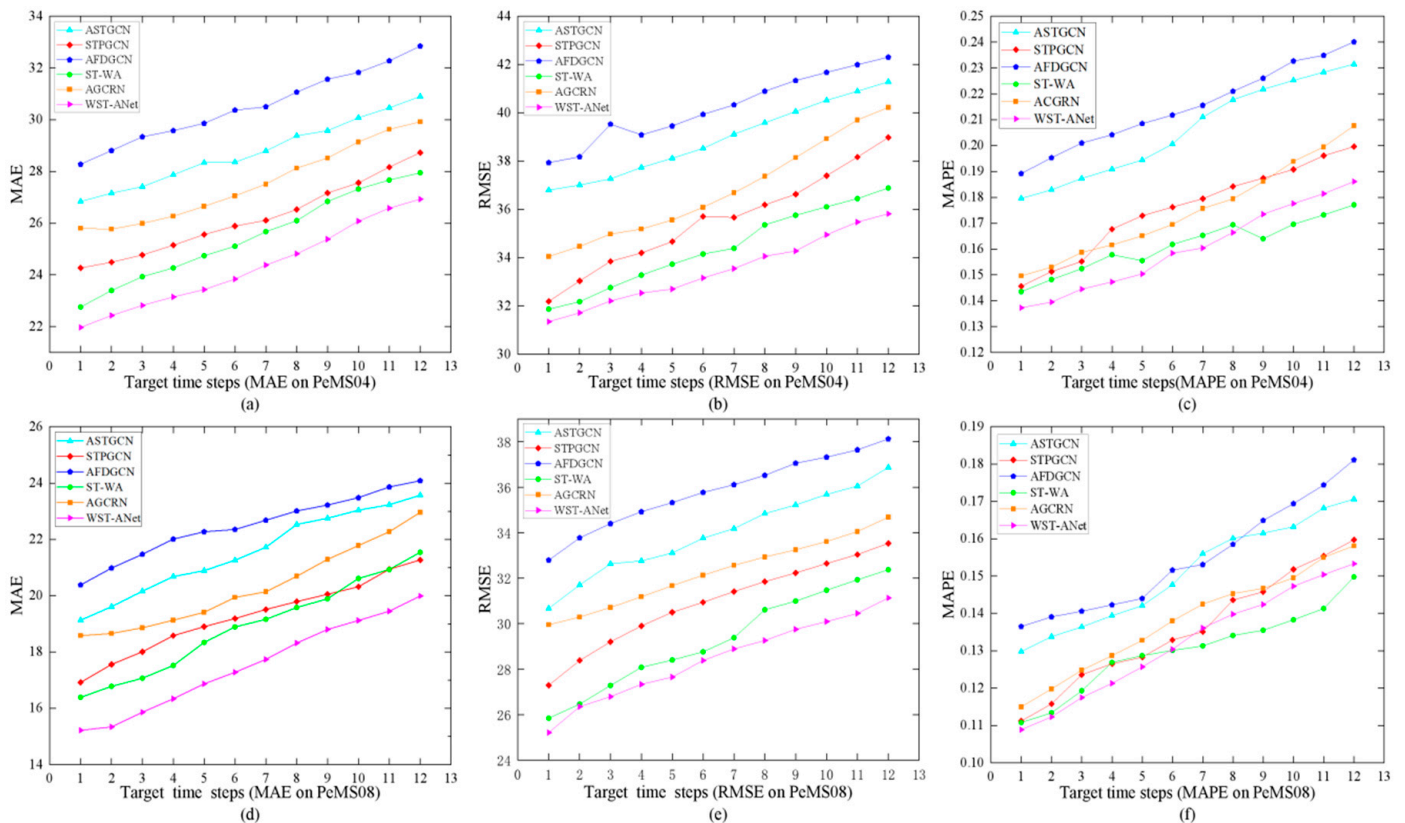


Figure 5. (a–c) denote the changes in MAE, RMSE, and MAPE metrics for different step-size prediction performance under PeMS04 dataset; (d–f) denote the changes in MAE, RMSE, and MAPE metrics for different step-size prediction performance under PeMS08 dataset.

5.3. Ablation Study

This section investigates the contribution of each component to the model in PeMSD8, prescribing the following WST-ANet variables. In order to verify the validity of the proposed model components, this section performs an ablation study of the different modules of the model. Each component of the model has an irreplaceable role in extracting the spatio-temporal dependencies, so this section investigates the contribution of each component to the model in PeMSD8:

- W/O DGM: indicates the removal of the dynamic graph structure module. The spatial convolutional layer of the graph is constructed using directly introduced spatial feature structures.
- W/O TWE: denotes that the WST-ANet model removes the sequence feature embedding of temporal weather interactions.
- W/O SWE: denotes that the WST-ANet model removes the positional feature embedding of spatial weather interactions.
- W/O CrossAtt: indicates that the WST-ANet model removes the cross-attention mechanism module that connects the encoder and decoder

The MAE and RMSE prediction errors of WST-ANet and its variants for PeMS04 and PeMS08 are shown in Figure 6. It is clear that the model with missing feature components did not predict as well as the complete model in terms of prediction accuracy. W/O DGM, W/O TWE, and W/O SWE had similar variations in the two data sets, i.e., the magnitude of the variation differences was small relative to the complete model. For example, in the PeMS04 dataset, the MAE accuracies of W/O DGM, W/O TWE, and W/O SWE were 3.21, 2.58, and 2.85 lower than those of the intact WST-ANet, and the RMSE was 3.20, 2.14, and 2.68 lower than that of the intact WST-ANet. The MAE and RMSE showed an overall steady-state trend of change.

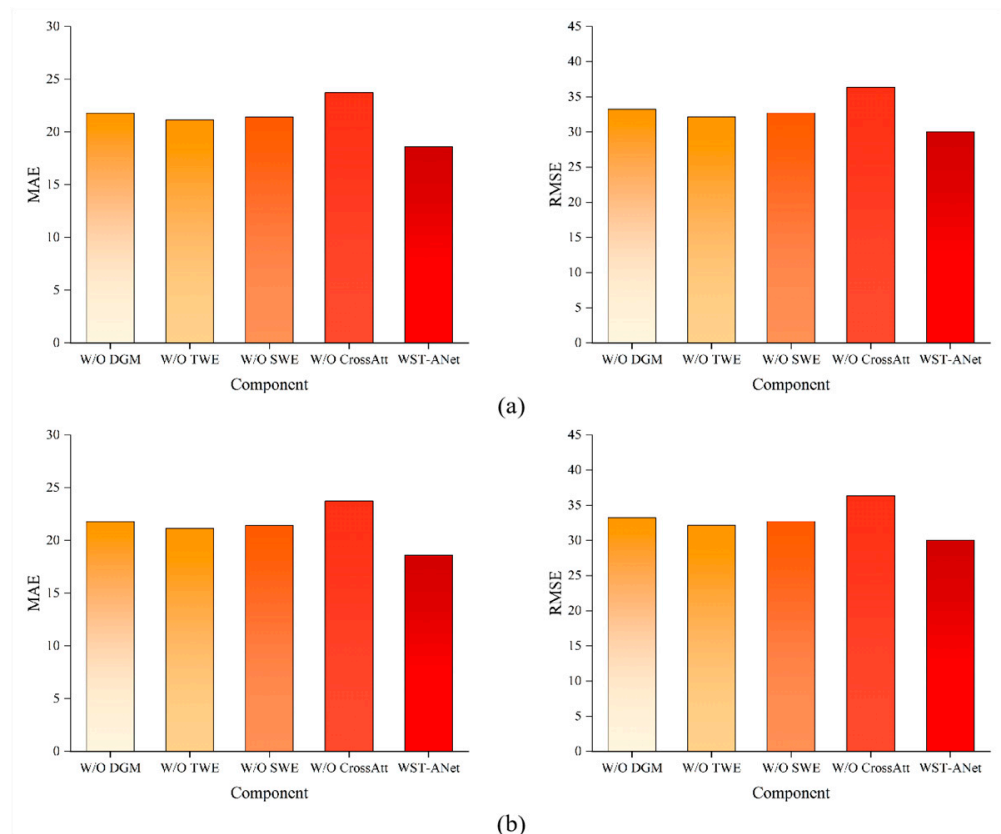


Figure 6. Ablation results. (a) Represents the MAE and RMSE of the ablation model under the PeMS04 dataset; (b) represents the MAE and RMSE of the ablation model under the PeMS08 dataset.

However, the accuracy of the MAE and RMSE of the model with the removal of the cross-attention mechanism component varied more, decreasing by 5.15 and 6.34, respectively. This was mainly due to the fact that the cross-attention mechanism enabled the decoder to focus on the output of the encoder to obtain the encoder information related to the current decoding position, which had a certain convergence effect on the accurate generation of the output sequence. Similarly, the PeMS08 dataset exists in the ablation analysis results with the above; therefore, it will not be repeated here. In summary, the complete model structure of WST-ANet outperformed the other models with missing components in terms of prediction accuracy, suggesting that WST-ANet can capture complex correlations driven by weather complexities, and that the inclusion of such weather features improves prediction accuracy to some extent.

5.4. Visualization Results

There was variability in the prediction accuracy caused by different weather drivers. For example, the prediction of data in the PeMS04 dataset in Figures 7a and 8a revealed that sensor 117, with a prediction step of 6 (i.e., 30 min), predicted a better fit in the morning of

February 27 in the three time periods of 0:00–2:00, 8:00–12:00, and 16:00–20:00 compared with February 20, due to the fact that the 27th fell under a heavy rainy day, while the 20th fell under a sunny day. This also demonstrated that our model was able to make better predictions in non-sunny environments. Similarly, the results of the study of data in the PeMS08 dataset showed that the non-sunny day prediction results had a better fit. The results showed that WST-ANet captured the polymorphic traffic patterns at different nodes under weather-driven conditions.

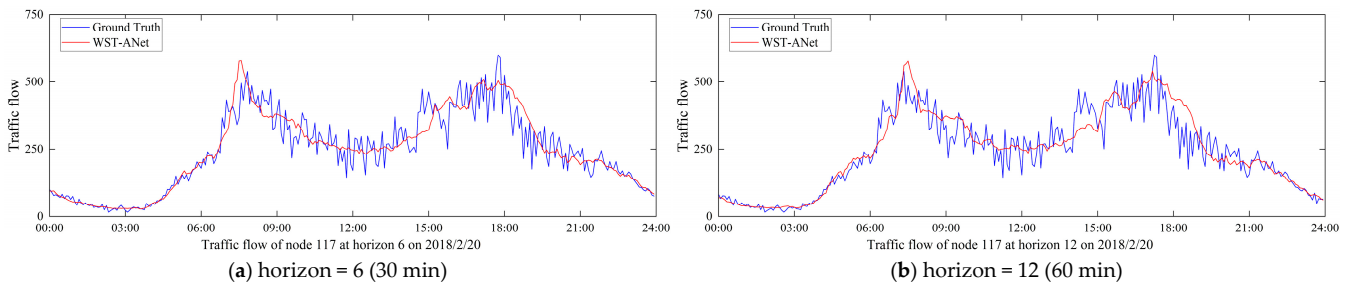


Figure 7. (a,b) represents the PeMS04 dataset on 20 February 2018. Visualization of traffic flow in one day at node 117.

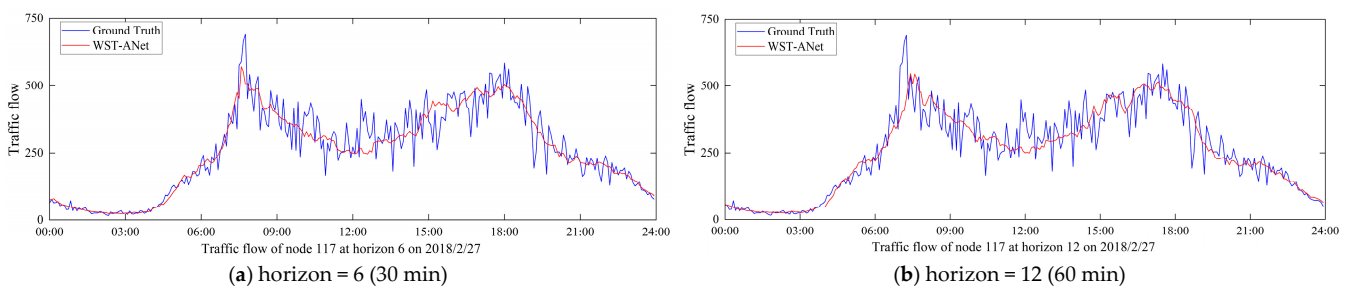


Figure 8. (a,b) represents the PeMS04 dataset on 27 February 2018. Visualization of traffic flow in one day at node 117.

To further visualize the understanding and evaluation of the model proposed in this paper, we performed ground visualization displays of the predicted and true values (from the test set) for different datasets. As an example, the predictions of the PeMS08 data are shown in Figures 9 and 10, where Sensor 15 on PeMS08 predicts and displays the predicted and true values (from the test set) for August 20, 2016 and August 27, 2016 for steps 6 and 12, respectively. The two time periods that were one week apart showed different traffic patterns. From the results of the two datasets, it can be observed that the smaller the prediction step size, the better the prediction accuracy the prediction shows; it can also predict the fluctuation of the data more accurately. As the prediction step size increased, the error accumulated throughout the period, resulting in a certain degree of decrease in the prediction accuracy, but overall, it still better predicted the trend of the data. In addition, we can observe that the surface traffic truth curve was very irregular and fluctuated a lot, while our model effectively adapted to these sudden trend changes and made predictions as close as possible to the real situation. From a planetary point of view, our proposed WST-ANet model recognized weather elements interactively and completely learned the traffic flow characteristics of the real road network.

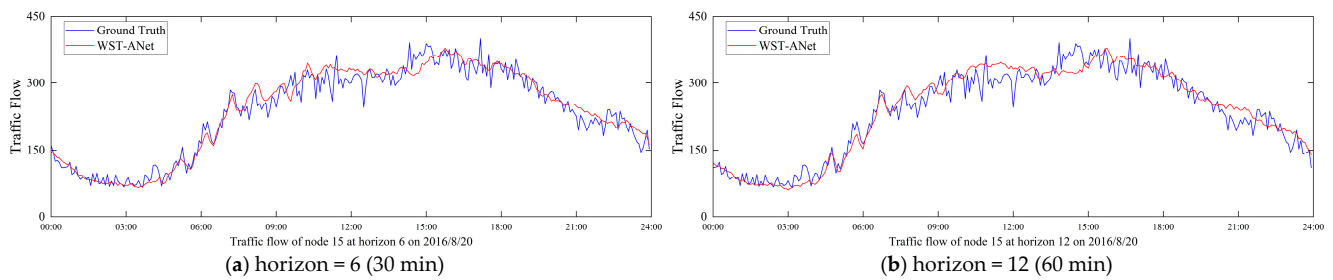


Figure 9. (a,b) represents the PeMS08 dataset on 20 August 2016. Visualization of traffic flow in one day at node 15.

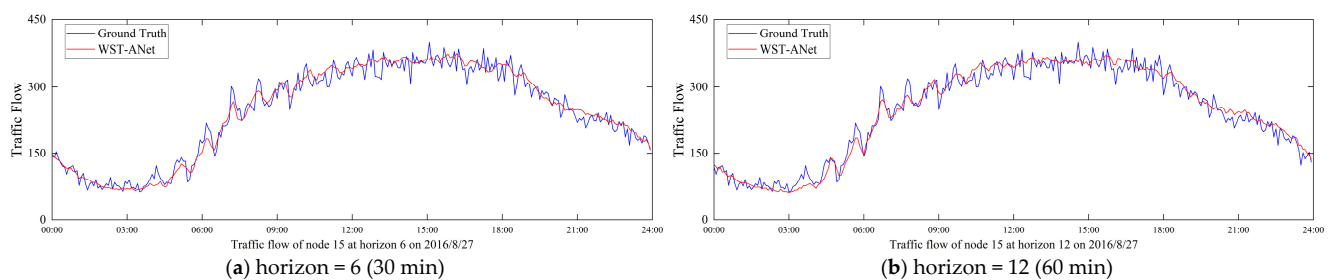


Figure 10. (a,b) represents the PeMS08 dataset on 27 August 2016. Visualization of traffic flow in one day at node 15.

6. Conclusions and Future Work

In order to improve the accuracy of traffic prediction in complex scenarios, we proposed a weather interaction-aware spatio-temporal attention network (WST-ANet) model based on the interactive perception of weather. Compared with existing traffic prediction models, the WST-ANet model interactively fuses the contextual semantics of traffic flows using a feature-embedding module from the spatio-temporal and weather-driven processes of the traffic road network, which is used to improve the deep adaptability of the road network to weather-driven factors. Combined with the dynamic graph module to provide the adjacency matrix for dynamic feature capture, the polynomial features in the spatial graph convolutional layer were then dynamically adjusted by the attention scores obtained from the encoder–decoder constructed by the WSTA Block to extract the relevance of spatio-temporal feature aggregation in the urban road network. In addition, by weighting the aggregated encoder feature information through the cross-attention mechanism and focusing on the spatio-temporal dependence constraint convergence, the extraction efficiency after model feature learning was higher. Our experimental results in real datasets show the superiority of WST-ANet over correlated baselines in traffic prediction. Therefore, the relevant departments can consider the weather factor as a more refined way of traffic forecasting, optimize the traffic signal control system more accurately according to the forecast data on traffic flow, and adjust the control timing of the signals dynamically, so as to improve the science and effectiveness of traffic management.

Our model has the following limitations: (1) only a single weather factor, temperature, was considered during the experiments and other important factors that may affect traffic flow, such as rainfall, wind speed, and humidity, were ignored; (2) the validation experiments only covered traffic flow as a pattern, and its broad applicability needs to be further verified; and (3) the effects of weather and spatially unbalanced traffic variations (e.g., the morning and evening rush hours) during special time periods were not considered in detail.

In the future, we will investigate the relationships between traffic and weather, as well as consider particular time periods for modeling and forecasting performance evaluation. For example, the influence of severe weather conditions on traffic flow and the impact of traffic congestion on weather pollution. Furthermore, we intend to further investigate and

develop the deep learning models so that they may be used in a broader range of traffic prediction domains.

Author Contributions: Conceptualization, J.W. (Jian Wang) and H.Z.; methodology, C.C.; software, H.Z. and C.C.; data curation, J.W. (Jianlong Wang) and K.G.; writing—original draft preparation, H.Z. and D.L.; writing—review and editing, J.W. (Jianlong Wang) and K.G. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Natural Science Foundation of Beijing (Grant No. 8222011).

Data Availability Statement: The data presented in this study are available in the article.

Conflicts of Interest: Author Jianlong Wang was employed by the company Changjiang Space Information Technology Engineering Co., Ltd. (Wuhan). The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

References

1. Wang, C.; Tian, R.; Hu, J.; Ma, Z. A trend graph attention network for traffic prediction. *Inf. Sci.* **2023**, *623*, 275–292. [\[CrossRef\]](#)
2. Tu, Y.; Lin, S.; Qiao, J.; Liu, B. Deep traffic congestion prediction model based on road segment grouping. *Appl. Intell.* **2021**, *51*, 8519–8541. [\[CrossRef\]](#)
3. Wang, K.; Liu, L.; Liu, Y.; Li, G.; Zhou, F.; Lin, L. Urban regional function guided traffic flow prediction. *Inf. Sci.* **2023**, *634*, 308–320. [\[CrossRef\]](#)
4. Lv, Z.; Zhang, S.; Xiu, W. Solving the Security Problem of Intelligent Transportation System with Deep Learning. *IEEE Trans. Intell. Transp. Syst.* **2021**, *22*, 4281–4290. [\[CrossRef\]](#)
5. Haydari, A.; Yilmaz, Y. Deep reinforcement learning for intelligent transportation systems: A survey. *IEEE Trans. Intell. Transp. Syst.* **2020**, *23*, 11–32. [\[CrossRef\]](#)
6. Rani, P.; Sharma, R. Intelligent transportation system for internet of vehicles based vehicular networks for smart cities. *Comput. Electr. Eng.* **2023**, *105*, 108543. [\[CrossRef\]](#)
7. Zhou, F.; Yang, Q.; Zhong, T.; Chen, D.; Zhang, N. Variational Graph Neural Networks for Road Traffic Prediction in Intelligent Transportation Systems. *IEEE Trans. Ind. Inform.* **2021**, *17*, 2802–2812. [\[CrossRef\]](#)
8. Yin, X.; Wu, G.; Wei, J.; Shen, Y.; Qi, H.; Yin, B. Deep Learning on Traffic Prediction: Methods, Analysis and Future Directions. *IEEE Trans. Intell. Transp. Syst.* **2021**, *23*, 4927–4943. [\[CrossRef\]](#)
9. Yuan, H.; Li, G. A Survey of Traffic Prediction: From Spatio-Temporal Data to Intelligent Transportation. *Data Sci. Eng.* **2021**, *6*, 63–85. [\[CrossRef\]](#)
10. Chandra, S.; Al-Deek, H. Predictions of Freeway Traffic Speeds and Volumes Using Vector Autoregressive Models. *J. Intell. Transp. Syst.* **2009**, *13*, 53–72. [\[CrossRef\]](#)
11. Yu, H.-F.; Rao, N.; Dhillon, I.S. Temporal regularized matrix factorization for high-dimensional time series prediction. In Proceedings of the Proceedings of the 30th International Conference on Neural Information Processing Systems, Barcelona, Spain, 5–10 December 2016; pp. 847–855.
12. Lippi, M.; Bertini, M.; Frasconi, P. Short-Term Traffic Flow Forecasting: An Experimental Comparison of Time-Series Analysis and Supervised Learning. *IEEE Trans. Intell. Transp. Syst.* **2013**, *14*, 871–882. [\[CrossRef\]](#)
13. Pan, Z.; Zhang, W.; Liang, Y.; Zhang, W.; Yu, Y.; Zhang, J.; Zheng, Y. Spatio-Temporal Meta Learning for Urban Traffic Prediction. *IEEE Trans. Knowl. Data Eng.* **2022**, *34*, 1462–1476. [\[CrossRef\]](#)
14. Ma, X.; Dai, Z.; He, Z.; Ma, J.; Wang, Y.; Wang, Y. Learning Traffic as Images: A Deep Convolutional Neural Network for Large-Scale Transportation Network Speed Prediction. *Sensors* **2017**, *17*, 818. [\[CrossRef\]](#)
15. Zhang, J.; Zheng, Y.; Qi, D. Deep Spatio-Temporal Residual Networks for Citywide Crowd Flows Prediction. In Proceedings of the AAAI Conference on Artificial Intelligence, San Francisco, CA, USA, 4–9 February 2017. [\[CrossRef\]](#)
16. Ali, A.; Zhu, Y.; Zakarya, M. Exploiting dynamic spatio-temporal correlations for citywide traffic flow prediction using attention based neural networks. *Inf. Sci.* **2021**, *577*, 852–870. [\[CrossRef\]](#)
17. Kipf, T.N.; Welling, M. Semi-supervised classification with graph convolutional networks. *arXiv* **2016**, arXiv:1609.02907.
18. Defferrard, M.; Bresson, X.; Vandergheynst, P. Convolutional neural networks on graphs with fast localized spectral filtering. *Adv. Neural Inf. Process. Syst.* **2016**, *29*. [\[CrossRef\]](#)
19. Elias, V.R.; Martins, W.A.; Werner, S. Extended adjacency and scale-dependent graph Fourier transform via diffusion distances. *IEEE Trans. Signal Inf. Process. Over Netw.* **2020**, *6*, 592–604. [\[CrossRef\]](#)
20. Theofilatos, A.; Yannis, G. A review of the effect of traffic and weather characteristics on road safety. *Accid. Anal. Prev.* **2014**, *72*, 244–256. [\[CrossRef\]](#) [\[PubMed\]](#)
21. Bi, H.; Ye, Z.; Zhu, H. Data-driven analysis of weather impacts on urban traffic conditions at the city level. *Urban. Clim.* **2022**, *41*, 101065. [\[CrossRef\]](#)

22. Shi, X.; Qi, H.; Shen, Y.; Wu, G.; Yin, B. A spatial–temporal attention approach for traffic prediction. *IEEE Trans. Intell. Transp. Syst.* **2020**, *22*, 4909–4918. [[CrossRef](#)]
23. Yao, W.; Qian, S. From Twitter to traffic predictor: Next-day morning traffic prediction using social media data. *Transp. Res. Part. C Emerg. Technol.* **2021**, *124*, 102938. [[CrossRef](#)]
24. Yang, T.; Yu, X.; Ma, N.; Zhao, Y.; Li, H. A novel domain adaptive deep recurrent network for multivariate time series prediction. *Eng. Appl. Artif. Intell.* **2021**, *106*, 104498. [[CrossRef](#)]
25. Du, S.; Li, T.; Yang, Y.; Horng, S.-J. Multivariate time series forecasting via attention-based encoder–decoder framework. *Neurocomputing* **2020**, *388*, 269–279. [[CrossRef](#)]
26. Storani, F.; Di Pace, R.; Bruno, F.; Fiori, C. Analysis and comparison of traffic flow models: A new hybrid traffic flow model vs. benchmark models. *Eur. Transp. Res. Rev.* **2021**, *13*, 58. [[CrossRef](#)]
27. Williams, B.M.; Hoel, L.A. Modeling and forecasting vehicular traffic flow as a seasonal ARIMA process: Theoretical basis and empirical results. *J. Transp. Eng.* **2003**, *129*, 664–672. [[CrossRef](#)]
28. Kumar, S.V. Traffic Flow Prediction using Kalman Filtering Technique. *Procedia Eng.* **2017**, *187*, 582–587. [[CrossRef](#)]
29. Ma, X.; Tao, Z.; Wang, Y.; Yu, H.; Wang, Y. Long short-term memory neural network for traffic speed prediction using remote microwave sensor data. *Transp. Res. Part. C Emerg. Technol.* **2015**, *54*, 187–197. [[CrossRef](#)]
30. Chung, J.; Gulcehre, C.; Cho, K.; Bengio, Y. Empirical evaluation of gated recurrent neural networks on sequence modeling. *arXiv* **2014**, arXiv:1412.3555.
31. Chen, X.; Chen, S.; Yao, J.; Zheng, H.; Zhang, Y.; Tsang, I.W. Learning on attribute-missing graphs. *IEEE Trans. Pattern Anal. Mach. Intell.* **2020**, *44*, 740–757. [[CrossRef](#)]
32. Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A.N.; Kaiser, Ł.; Polosukhin, I. Attention is all you need. *Adv. Neural Inf. Process. Syst.* **2017**, *30*. [[CrossRef](#)]
33. Zhang, C.; James, J.; Liu, Y. Spatial-temporal graph attention networks: A deep learning approach for traffic forecasting. *IEEE Access* **2019**, *7*, 166246–166256. [[CrossRef](#)]
34. Guo, S.; Lin, Y.; Feng, N.; Song, C.; Wa, H. Attention Based Spatial-Temporal Graph Convolutional Networks for Traffic Flow Forecasting. In Proceedings of the AAAI Conference on Artificial Intelligence, Honolulu, HI, USA, 21 January–1 February 2019.
35. Song, C.; Lin, Y.; Guo, S.; Wan, H. Spatial-temporal synchronous graph convolutional networks: A new framework for spatial-temporal network data forecasting. In Proceedings of the AAAI Conference on Artificial Intelligence, New York, NY, USA, 7–12 February 2020; pp. 914–921.
36. Luo, X.; Zhu, C.; Zhang, D.; Li, Q. Dynamic Graph Convolution Network with Spatio-Temporal Attention Fusion for Traffic Flow Prediction. *arXiv* **2023**, arXiv:2302.12598.
37. Li, Y.; Yu, R.; Shahabi, C.; Liu, Y. Diffusion convolutional recurrent neural network: Data-driven traffic forecasting. *arXiv* **2017**, arXiv:1707.01926.
38. Cirstea, R.-G.; Yang, B.; Guo, C.; Kieu, T.; Pan, S. Towards spatio-temporal aware traffic time series forecasting. In Proceedings of the 2022 IEEE 38th International Conference on Data Engineering (ICDE), Kuala Lumpur, Malaysia, 9–12 May 2022; pp. 2900–2913.
39. Bai, L.; Yao, L.; Li, C.; Wang, X.; Wang, C. Adaptive graph convolutional recurrent network for traffic forecasting. *Adv. Neural Inf. Process. Syst.* **2020**, *33*, 17804–17815.
40. Wu, Z.; Pan, S.; Long, G.; Jiang, J.; Zhang, C. Graph wavenet for deep spatial-temporal graph modeling. *arXiv* **2019**, arXiv:1906.00121.
41. Zhao, Y.; Lin, Y.; Wen, H.; Wei, T.; Jin, X.; Wan, H. Spatial-Temporal Position-Aware Graph Convolution Networks for Traffic Flow Forecasting. *IEEE Trans. Intell. Transp. Syst.* **2023**, *24*, 8650–8666. [[CrossRef](#)]
42. Bao, Y.; Huang, J.; Shen, Q.; Cao, Y.; Ding, W.; Shi, Z.; Shi, Q. Spatial–Temporal Complex Graph Convolution Network for Traffic Flow Prediction. *Eng. Appl. Artif. Intell.* **2023**, *121*, 106044. [[CrossRef](#)]
43. Zhang, W.; Yao, R.; Du, X.; Liu, Y.; Wang, R.; Wang, L. Traffic flow prediction under multiple adverse weather based on self-attention mechanism and deep learning models. *Phys. A Stat. Mech. Appl.* **2023**, *625*, 128988. [[CrossRef](#)]
44. Nigam, A.; Srivastava, S. Weather impact on macroscopic traffic stream variables prediction using recurrent learning approach. *J. Intell. Transp. Syst.* **2023**, *27*, 19–35. [[CrossRef](#)]
45. Geng, X.; Li, Y.; Wang, L.; Zhang, L.; Yang, Q.; Ye, J.; Liu, Y. Spatiotemporal multi-graph convolution network for ride-hailing demand forecasting. In Proceedings of the AAAI Conference on Artificial Intelligence, Honolulu, HI, USA, 27 January–1 February 2019; pp. 3656–3663.
46. Zheng, C.; Fan, X.; Wang, C.; Qi, J. Gman: A graph multi-attention network for traffic prediction. In Proceedings of the AAAI Conference on Artificial Intelligence, New York, NY, USA, 7–12 February 2020; pp. 1234–1241.
47. Zou, G.; Lai, Z.; Wang, T.; Liu, Z.; Bao, J.; Ma, C.; Li, Y.; Fan, J. Multi-task-based spatiotemporal generative inference network: A novel framework for predicting the highway traffic speed. *Expert Syst. Appl.* **2024**, *237*, 121548. [[CrossRef](#)]
48. Zhang, F.; Feng, N.; Liu, Y.; Yang, C.; Zhai, J.; Zhang, S.; He, B.; Lin, J.; Du, X. PewLSTM: Periodic LSTM with Weather-Aware Gating Mechanism for Parking Behavior Prediction. In Proceedings of the International Joint Conference on Artificial Intelligence, Yokohama, Japan, 11–17 July 2020; pp. 4424–4430.
49. Wang, Y.; Ren, Q.; Li, J. Spatial–temporal multi-feature fusion network for long short-term traffic prediction. *Expert Syst. Appl.* **2023**, *224*, 119959. [[CrossRef](#)]

50. Grover, A.; Leskovec, J. node2vec: Scalable feature learning for networks. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, New York, NY, USA, 13–17 August 2016; pp. 855–864.
51. Zheng, J.; Liu, H.; Feng, Y.; Xu, J.; Zhao, L. CASF-Net: Cross-attention and cross-scale fusion network for medical image segmentation. *Comput. Methods Programs Biomed.* **2023**, *229*, 107307. [[CrossRef](#)] [[PubMed](#)]
52. Tian, R.; Wang, C.; Hu, J.; Ma, Z. Multi-scale spatial-temporal aware transformer for traffic prediction. *Inf. Sci.* **2023**, *648*, 119557. [[CrossRef](#)]
53. Huang, S.; Wu, Q. Robust pairwise learning with Huber loss. *J. Complex.* **2021**, *66*, 101570. [[CrossRef](#)]
54. Chen, C.; Liu, Y.; Chen, L.; Zhang, C. Bidirectional Spatial-Temporal Adaptive Transformer for Urban Traffic Flow Forecasting. *IEEE Trans. Neural Netw. Learn. Syst.* **2022**, *34*, 6913–6925. [[CrossRef](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.