Transportation Research Part C 154 (2023) 104263

Contents lists available at ScienceDirect

Transportation Research Part C

journal homepage: www.elsevier.com/locate/trc

A novel spatio-temporal generative inference network for predicting the long-term highway traffic speed

Guojian Zou [a,b], Ziliang Lai [a,b], Changxi Ma [c], Ye Li [a,b,*], Ting Wang [a,b]

[a] The Key Laboratory of Road and Traffic Engineering, Ministry of Education, Tongji University, Shanghai, 201804, PR China
[b] College of Transportation Engineering, Tongji University, Shanghai 201804, PR China
[c] School of Traffic and Transportation, Lanzhou Jiaotong University, Lanzhou 730070, PR China

ARTICLE INFO

Keywords:
Long-term highway traffic speed prediction
Spatio-temporal correlation
Graph neural networks
Long short-term memory network
Generative inference
Multi-head self-attention

ABSTRACT

Accurately predicting the highway traffic speed can reduce traffic accidents and transit time, which is of great significance to highway management. Three essential elements should be considered in the long-term highway traffic speed prediction: (1) adaptability to speed fluctuation, (2) exploring the spatio-temporal correlation effectively, and (3) prediction of non-error propagation. This paper proposes a novel spatio-temporal generative inference network (STGIN) driven by data and long-term prediction. STGIN consists of three parts; semantic enhancement, spatio-temporal correlation extraction block (ST-Block), and generative inference. Semantic enhancement is first used to model the contextual semantics of traffic speed, improving the adaptability to speed fluctuations. The ST-Block is then used to extract the spatio-temporal correlations of the highway network. Finally, generative inference is used to pay attention to the correlation between historical- and target-sequences to generate the target hidden outputs rather than a dynamic step-by-step decoding way; it avoids long-term prediction error propagation in the spatial and temporal dimensions. The evaluation experiments use the monitoring data of highway in Yinchuan City, Ningxia Province, China. For long-term highway speed prediction, the experimental results demonstrate that the performance of the proposed method is better than that of the baseline methods.

1. Introduction

The highway is a critical way for people to travel and transport goods. It plays an essential role in the rapid economic development (Magazzino and Mele, 2021). Long-term highway traffic speed prediction is crucial for the intelligent transportation system (ITS), providing helpful information for travelers and traffic management departments (Liu et al., 2018; James et al., 2021; Yu et al., 2020). Typically, for long-term traffic speed prediction, historical observations are used to predict multiple future time step speeds. Long-term prediction requires more considerations than short-term prediction based on a single step, such as avoiding prediction error propagation.

The statistical methods, e.g., autoregressive integrated moving average (ARIMA) (Ahmed and Cook, 1979), and their improved methods are widely used in traffic-related time series prediction tasks (Duan et al., 2016; Wang et al., 2016). The advantage of these methods is that only a small batch of samples can roughly predict the target data. However, traffic speed prediction is a spatial–temporal correlation extraction problem affected by both temporal and spatial dimensions; they face the challenge of not being able

* Corresponding author at: The Key Laboratory of Road and Traffic Engineering, Ministry of Education, Tongji University, Shanghai, 201804, PR China.
E-mail addresses: 2010768@tongji.edu.cn (G. Zou), 2033402@tongji.edu.cn (Z. Lai), machangxi@mail.lzjtu.cn (C. Ma), JamesLI@tongji.edu.cn (Y. Li), 2110763@tongji.edu.cn (T. Wang).

https://doi.org/10.1016/j.trc.2023.104263
Received 11 February 2023; Received in revised form 4 July 2023; Accepted 17 July 2023
Available online 5 August 2023

to extract the nonlinear and complex spatio-temporal correlations of traffic data. Approaches based on traditional machine learning, such as support vector regression (SVR) (Hong, 2011), have been widely studied for traffic prediction tasks. These methods take into account the nonlinear correlations of the traffic data and have a more remarkable accuracy improvement than the statistical methods. However, traditional machine learning methods rely on complex mathematical algorithms and prior knowledge to obtain shallow data correlations, and it is not easy to extract deep, complex spatio-temporal correlations. Therefore, the accuracy and application value are difficult to be further improved.

In recent years, deep learning methods for automatic feature extraction have brought revolutions in computer vision and natural language processing (Fang et al., 2021; Otter et al., 2020), which attracted widespread application in cross-fields (Jin et al., 2020, 2022; Pan et al., 2020; Ma et al., 2021). Subsequently, some deep learning methods have been deployed in traffic speed prediction and achieved satisfactory performance (Yang et al., 2021; Jia et al., 2021; Qu et al., 2021), such as PSPNN (Yang et al., 2021). The highway traffic data conforms to discrete distribution, that is, non-Euclidean structure data. However, in these methods, convolutional neural networks (CNNs) are mainly used to extract the spatial correlation of the Euclidean structure data, such as pictures, and it is unsuitable for non-Euclidean structure data (Zafeiriou et al., 2022). Some current methods force the sampling of traffic data into a standard form and feed it to the CNNs, which may lose important spatial correlation information of the traffic data (Yang et al., 2021). The ideal way to express traffic data consists in maintaining the original spatial structure, and expressing it in a graph network. The latest researches have extended CNNs to graph neural networks (GNNs) (Luo et al., 2022), which can process data with the arbitrary graph structure. Especially, the methods based on GNNs and attention mechanisms have been employed successfully in traffic prediction tasks, including traffic speed prediction (Jin et al., 2023).

However, the existing GNN-based prediction models face three issues: compound-spatial and -temporal correlations, prediction error propagation, and speed fluctuations. Specifically, for the same time step, the traffic speed of the target road segment is affected by the vehicle speed in the upstream and downstream directions, defined as physical spatial dependency; for different time steps, the influence between road segments changes accordingly due to the dynamic evolution of traffic conditions, defined as dynamic spatial correlation. We may refer to these properties as compound spatial correlations. The speed of the target road segment varies dynamically in the temporal dimension and exhibits continuity on the time axis, which is defined as the inherent temporal dependency; for different time steps, the correlations between traffic speeds on the target road segment are distinct and vary with time, which is defined as the dynamic temporal correlation. We may refer to these properties as compound temporal correlations. In addition, for the prediction, the influence of historical time steps on target time steps cannot be ignored (Zheng et al., 2020; Park et al., 2020); nevertheless, it is difficult for these considered methods to solve the long-term prediction error propagation in the spatial and temporal dimensions, resulting in the accumulation of errors and limited model prediction accuracy. Moreover, traffic speeds are easily affected by other factors, such as during morning and evening rush hours, which fluctuate, making forecasting more difficult.

Predicting highway traffic speed is challenging because, compared to urban roads, the types of vehicles are more diverse, and the traffic condition is susceptible to external influences (e.g., agglomerate fog). In this paper, a novel spatio-temporal generative inference network (STGIN) for highway traffic speed prediction, is proposed. Semantic enhancement based on 1-D convolutional neural networks (1-D CNNs) is first used to model the contextual semantics of traffic speed, improving the adaptability to speed fluctuations. A spatio-temporal correlation extraction block (ST-Block) based on GNNs, temporal attention, and long short-term memory network (LSTM) is then designed to extract compound-spatial and -temporal dependencies of the traffic data. Finally, a specific transformer generates the long-term target hidden outputs in a single step by bridging the relationship between the historical and target sequences, called bridge transformer (BridgeTrans), and this architecture is defined as generative inference; it avoids long-term prediction error propagation in the spatial and temporal dimensions.

The contributions of this paper are summarized as follows:

1. In the STGIN, the semantic enhancement module is developed and implemented to model the contextual semantics of traffic speed, thereby enhancing adaptability to speed fluctuations; it prevents contextual semantic breaks caused by large speed fluctuations, hence limiting the model to learning the speed change regularity and affecting the prediction accuracy.
2. Highway traffic data is characterized by four dimensions, dynamic spatial correlation, physical spatial dependency, dynamic temporal correlation, and inherent temporal dependency. ST-Block is designed to extract these correlations, simulating the entire dynamic evolution process of highway network traffic speed with precision.
3. To avoid long-term prediction error propagation in the spatial and temporal dimensions, we design a generative inference that pays attention to the correlation between the historical sequence and the target sequence to generate the target hidden outputs, as opposed to a decoding method such as ST-GRAT (Park et al., 2020), which makes the dynamic step-by-step inference. In addition, residual connection (He et al., 2016) and batch normalization (BN) (Ioffe and Szegedy, 2015) are added to each layer of the network to avoid network feature loss and internal covariate shift.
4. Several experiments are conducted on the highway traffic dataset. The experimental results show that the proposed STGIN model outperforms all the baseline methods.

The remainder of this paper is organized as follows. In Section 2, the previous related studies are summarized. Section 3 describes the relative definition and problem statement. Section 4 details the proposed STGIN model. Section 5 presents the experiments and the results. Finally, the conclusion and future work are drawn in Section 6.

2. Related work

The existing traffic speed prediction techniques can be divided into three categories: statistical methods, traditional machine learning approaches, and deep learning algorithms.

Statistical methods have been successfully applied to traffic speed prediction tasks, including the history average model (HA) (Liu et al., 2019) and ARIMA model (Ahmed and Cook, 1979; Duan et al., 2016; Wang et al., 2016). However, traffic speed is nonlinear, and the statistical methods are based on prior knowledge, theoretical assumptions, and simple mathematical statistics. Therefore, they face difficulties in accurately predicting the traffic speed. Traditional machine learning approaches alleviate the problems encountered by statistical methods and are used in traffic speed prediction tasks, including gradient boosting decision tree (GBDT) (Wu et al., 2020b), SVR (Hong, 2011), support vector machine (SVM) (Vanajakshi and Rilett, 2004), hidden markov model (HMM) (Jiang and Fei, 2016), markov chain (MC) (Shin and Sunwoo, 2018), and traffic factor state network (TFSN) based on high-order multivariate markov models (Zhang et al., 2020), due to their ability to capture nonlinear correlations. Existing studies have proved that traffic speed prediction is affected by both temporal and spatial dimensions (Zhou et al., 2019). However, these traditional machine learning approaches are incapable of modeling the complex spatio-temporal correlations of traffic data.

In recent years, deep learning has achieved a high performance when dealing with regression problems. In the early studies, the researchers found that neural network algorithms are more suitable for receiving and processing complex and nonlinear traffic data than traditional machine learning techniques (Csikós et al., 2015; Jia et al., 2016; Tang et al., 2017); examples include artificial neural networks (ANN) (Csikós et al., 2015), deep belief network (DBN) (Jia et al., 2016), and fuzzy neural network (FNN) (Tang et al., 2017). Since traffic speed prediction is a typical time series prediction problem, it is essential to extract the temporal correlation. Recurrent neural networks (RNNs) are deep learning methods that can efficiently extract the temporal correlations of the data and are used for time series prediction tasks (Zhang et al., 2021b; Ong et al., 2016). At present, several traffic prediction methods use RNNs and its variants as temporal extractors, in order to improve the prediction accuracy (Qu et al., 2021; Ma et al., 2015; Gu et al., 2019; Meng et al., 2020; Yi and Bui, 2020; Wang et al., 2019). For instance, the integration of LSTM with bidirectional LSTM (Bi-LSTM) is employed for traffic flow prediction (Ma et al., 2021). Qu et al. (2021) proposes the features injected recurrent neural networks (FI-RNNs), which combines sequential temporal data with contextual factors and uses a staked RNN and sparse autoencoder to mine the potential relationship between traffic state and its contexts. However, if only RNNs are utilized to process the temporal correlation of the traffic data, the impact of the spatial correlation on the prediction may be overlooked.

To solve the problems encountered by RNNs, the spatio-temporal prediction models based on CNNs have been designed and widely used for traffic speed prediction (Yang et al., 2021; Lv et al., 2018; Song et al., 2017; Jia et al., 2021; Zhou et al., 2019; Lu et al., 2020; Zang et al., 2018). For example, Zhou et al. (2019) propose a novel speed prediction method, referred to as the spatio-temporal and deep tensor neural networks (ST-DTNN), which is mainly used for large-scale urban networks with mixed road types. To predict the lane-level short-term traffic speed, Lu et al. (2020) propose a novel mixed deep learning (MDL), which consists of a convolutional long short-term memory (Conv-LSTM) layer, a convolutional layer, and a fully connected layer. Yang et al. (2021) propose a path-based speed prediction neural network (PSPNN) composed of CNN and a bidirectional LSTM (Bi-LSTM) network, which extracts the temporal and spatial correlations of historical data to perform path-based speed prediction. However, these methods extract the spatial correlation using CNNs, whereas typical CNNs only deal with this issue in Euclidean space and are, therefore, inadequate for non-Euclidean space.

Recently, GNNs have been widely used for traffic speed prediction, since they can process non-Euclidean structure data (James et al., 2021; Zhang et al., 2021a; Lu et al., 2022; Zhao et al., 2022). For instance, Li et al. (2018) propose a diffusion convolutional recurrent neural network (DCRNN), a deep learning framework incorporating spatial and temporal dependency into traffic prediction. Yu et al. (2018) propose a novel spatio-temporal graph convolutional network (STGCN), which utilizes graph convolution to extract spatial correlation and employs 1-D convolution to model temporal dependency. Zhao et al. (2019) propose a novel temporal graph convolutional network (T-GCN) for traffic prediction, which is combined with the graph convolutional network (GCN) and gate recurrent unit (GRU). These approaches extend the previous works to GCN; however, over-depend on handling pre-defined directed graphs. Therefore, Bai et al. (2020) proposed an adaptive graph convolutional recurrent network (AGCRN) forecast traffic effectively without relying on pre-defined graphs. Wu et al. (2019) introduced a novel graph neural network architecture called Graph-WaveNet, which develops a novel adaptive dependency matrix. This matrix is learned through node embedding to capture hidden spatial dependencies effectively. To address the issue of multivariate time series lacking an explicit graph structure, Wu et al. (2020a) introduced a framework (referred to as MTGNN in our paper) that consists of three components: the graph learning layer, which captures the graph adjacency matrix adaptively based on data, and the graph convolution module and temporal convolution module.

Furthermore, approaches like STMFFN (Wang et al., 2023) and Lastjormer (Fang et al., 2022) have further enhanced the modeling of spatial and temporal dependencies by incorporating more intricate attention mechanisms with GCNs. For example, Guo et al. (2019) propose a novel attention-based spatial–temporal graph convolutional network (ASTGCN) model for traffic forecasting. More specifically, the spatio-temporal attention mechanism captures the dynamic spatio-temporal correlations, and graph convolutions are employed to capture the spatial patterns and common standard convolutions to describe the temporal features. Zheng et al. (2020) proposed a graph multi-attention network (GMAN) to predict long-term traffic speed at different locations on the traffic network. Park et al. (2020) propose a novel spatio-temporal graph attention (ST-GRAT) method based on the self-attention mechanism. It efficiently captures the spatio-temporal dynamic correlations in the road network, and improves the prediction accuracy. Zhao et al. (2022) proposes a novel attention-based dynamic spatio-temporal graph convolutional network (ADSTGCN), which is used to model dynamic spatio-temporal correlations and improve long-term prediction performance.

Fig. 1. (a) Example of highway road network. Each road segment is represented by one color. (b) The highway network can be represented as a graph, while the road segments can be represented as nodes.

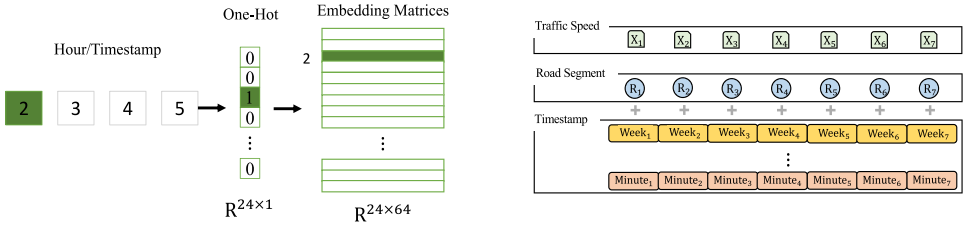


Fig. 2. Left: Example of hour embedding. **Right:** The node representation consists of three types of primary data information: traffic speed embedding, road segment embedding, and timestamp embeddings (including week, day, hour, and minute).

However, these methods neglect traffic speed fluctuations on observations and do not consider the compound-spatial and -temporal correlations of the traffic data simultaneously for speed prediction. In addition, long-term prediction error propagation in the spatial and temporal dimensions, such as GMAN (Zheng et al., 2020) and ST-GRAT (Park et al., 2020), cannot be avoided. Moreover, some models, such as ADSTGCN (Zhao et al., 2022), do not account for the influence of historical time steps on target time steps during the prediction phase. In this paper, inspired by the recent studies on graph neural networks in traffic speed prediction, a novel long-term highway traffic speed prediction model, referred to as STGIN, is proposed.

3. Preliminary

In this section, we first present several preliminaries and define our problem formally.

Definition 1 (*Road Network*)**.** The highway is a directed topological network connected in physical space, as shown in Fig. 1(a). In order to perform the essential preliminary work for modeling traffic data, we must map the highway network in the geographical area to the logical space that the computer can comprehend. Each line segment represents a highway road segment (cf. Fig. 1(a)), while each road segment can be mapped to the graph network node, and the connections between the road segments are abstracted as edges (cf. Fig. 1(b)). The input highway graph is defined as $G = (V, E, A)$, where V represents the nodes, E denotes the edges, $A \in R^{N\times N}$ is the adjacency matrix, and N represents the number of nodes. More precisely, when A_{v_i,v_j} is 'one', indicating a directed connection edge between node v_i and node v_j (and vice versa), while 'zero' indicates no directed connection.

In contrast to the conventional road network, the sensor is a node in the graph, and road segments are regarded as nodes in this paper, as depicted in Fig. 1(a). As defined by physical spatial dependency, the traffic condition of the red road segment is influenced by upstream (i.e., blue road segment) and downstream (i.e., pink, green, and yellow road segments) flows, and this connectivity does not change over time. As shown in Fig. 1(b), the road network is a closed-loop graph; the traffic flow of the red node originates from the blue (first-order upstream neighbor) and then flows into the pink, green, and yellow nodes (first-order downstream neighbors).

Definition 2 (*Embedding Statement*)**.** Three types of input data information are included for traffic speed prediction: traffic speed, road segment embedding, and timestamp embeddings. The traffic speed input to STGIN at time step $t \in T$ is $X_t \in R^{N\times d_X}$, road segment embedding at time step t is $XR_t \in R^{N\times d}$, timestamp embeddings at time step t including minute embedding $XM_t \in R^{N\times d}$, hour embedding $XH_t \in R^{N\times d}$, day embedding $XD_t \in R^{N\times d}$, and week embedding $XW_t \in R^{N\times d}$. The timestamp and road segment embedded methods are similar to the embedded method in the BERT (Kenton and Toutanova, 2019), which is mapped to the dense matrix through one-hot (Zou et al., 2023), as shown in Fig. 2 left. The total timesteps are represented by T; $d_X = 1$ denotes the input speed dimension at time step t; to consider computation cost, the dimension of d is set to 64 based on prior knowledge.

Unlike GMAN, the road segment embeddings in this paper trained in the training phase are the same as the BERT (Zheng et al., 2020; Kenton and Toutanova, 2019). We also fed road segment and timestamp embeddings $XR + XM + XH + XD + XW$ to a two-layer feed-forward neural network and obtained spatio-temporal embeddings (STE) as additional information input to our proposed model, as shown in Fig. 2 right.

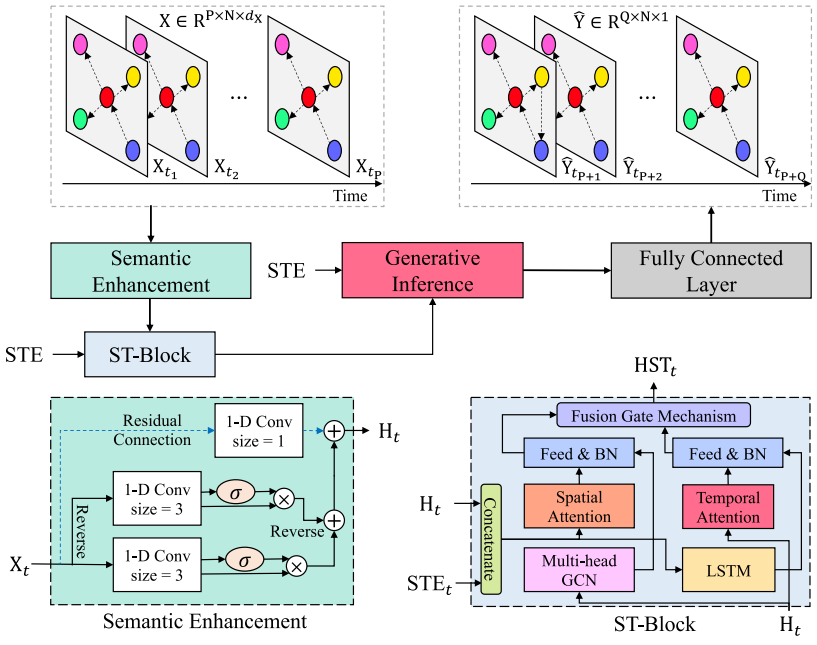


Fig. 3. The framework of the proposed STGIN.

Definition 3 (*Problem Statement*)**.** STGIN aims at predicting the long-term traffic speed in each highway road segment. Assume that input time steps length is P and the prediction time steps length is Q. Given the historical sequence of observations $X = \left\{X_{t_1}, \ldots, X_{t_P}\right\} \in R^{P\times N\times d_X}$ of N nodes in P time steps and $STE \in R^{(P+Q)\times N\times d}$ of N nodes in P + Q time steps, we aim to predict the target sequence values of Q time steps for N nodes, expressed as $\hat{Y} = \left\{\hat{Y}_{t_{P+1}}, \ldots, \hat{Y}_{t_{P+Q}}\right\} \in R^{Q\times N\times 1}$.

4. Proposed approach

4.1. Framework overview

Fig. 3 depicts the framework of the proposed STGIN for the extraction of spatio-temporal correlations and the prediction of long-term highway traffic speed. First, contextual semantics of traffic speed is modeled using a semantic enhancement module based on 1-D CNNs. The ST-Block is then designed to derive complex spatial and temporal correlations from the traffic data. Multi-head graph attention network (multi-head GAT) and multi-head graph convolutional network (multi-head GCN) are developed to extract compound spatial correlations; long short-term memory network (LSTM) and temporal attention are employed to extract compound temporal correlations. Finally, instead of a dynamic step-by-step decoding method, the target hidden outputs are generated using generative inference based on BridgeTrans by paying attention to the correlation between historical and target sequences. Moreover, residual connection and BN are added to the entire STGIN. Each part of the proposed method is detailed in the sequel.

4.2. Semantic enhancement

Neural language processing (NLP) always employs contextual semantics to characterize the logical context relationship within a sentence (Saif et al., 2016). As in the sentence 'I like playing basketball', the context conforms to the grammatical constraint (I-Subject, like-Verb, and playing basketball-Object). The relationship between words in the sentence from left to right and vice versa is substantial. For the traffic speeds, as with a sentence, the observed values are continuous in both directions, which is also referred to as contextual semantics in this paper.

The complexity of the traffic speed variations on the highway network makes prediction challenging. For instance, between 6:45 and 7:15 a.m., during the morning rush hour, the traffic speed drops significantly, resulting in context semantic disruption. Contextual semantic continuity is crucial for predictive models, reflecting the continuity of traffic speed changes. However, gaps in contextual semantics can make it difficult for the model to react positively to irregular variations in traffic speed. Fortunately, 1-D convolutional neural networks (1-D CNNs) are employed with an NLP-inspired n-gram concept to capture local context semantics (Zhou et al., 2022). Therefore, we use 1-D CNNs with kernel size k to model the contextual semantics of traffic speed, and this processing is defined as semantic enhancement, as shown in Fig. 3. The advantage of a semantic enhancement module is that sensitive to k time steps traffic speed, smoothing contextual semantic interruptions. Fig. 4 is a case of processing context

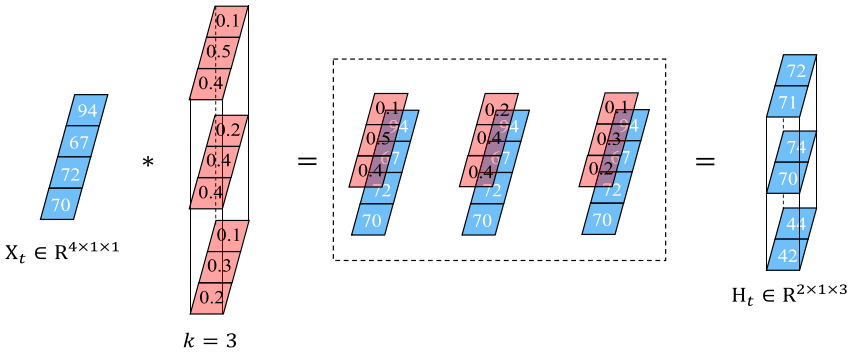


Fig. 4. 1-D CNN.

semantics interruption, which occurs when the traffic speed drops from 94 km/h to 67 km/h. We use a kernel to perform non-padding convolution on the input traffic speed $X_t \in R^{4\times1\times1}$ and achieve contextual semantic values $H_t \in R^{2\times1\times3}$ continuity.

In general, the semantic enhancement module works as follow. As Fig. 3 shows, the semantic enhancement module contains two 1-D CNNs with kernel size $k = 3$ followed by read gate unit, respectively, and a 1-D CNN with kernel size $k = 1$ used for a residual connection. The 0 padding method is used to keep the temporal channel dimension consistent. For example, when kernel size $k = 3$, three neighbor time steps of $v \in V$ are mapped into one element, as shown in Fig. 4. However, padding caused the last two elements of the temporal dimension to be abnormal in the 1-D CNN with kernel size $k = 3$; a reverse trick is used to solve this problem. As a result, semantic enhancement module works can be defined as,

$$\begin{cases} H_1 = \sigma\left(f\left(X * \omega_1\right)\right) \odot f\left(X * \omega_1\right) \\ H_2 = \sigma\left(f\left(reverse\left(X\right) * \omega_2\right)\right) \odot f\left(reverse\left(X\right) * \omega_2\right) \\ H = H_1 + reverse\left(H_2\right) + f\left(X * \omega_3\right) \end{cases} \tag{1}$$

where $\odot$ denotes the Hadamard product, $*$ denotes convolution calculation, ω represents convolution kernel, σ represents a sigmoid function, and $\sigma \odot f$ represents read gate unit. The σ function is similar to the attention mechanism used to control the output flow of 1-D CNN, and the function value ranges are between [0.0,1.0]. The initial input of semantic enhancement module is $X \in R^{P\times N\times d_X}$, and the output is $H \in R^{P\times N\times d}$.

4.3. ST-Block architecture

In the spatial and temporal dimensions, there are four characteristics of the traffic network: physical spatial dependency, dynamic spatial correlation, inherent temporal dependency, and dynamic temporal correlation. The working process of the ST-Block is shown in Fig. 3, and we utilize the multi-head GAT and multi-head GCN in the ST-Block to model compound spatial correlations and use temporal attention and LSTM to model the compound temporal correlations. Assume that the input of ST-Block is $STE[:P] \in R^{P\times N\times d}$ and $H \in R^{P\times N\times d}$, and output is $HST \in R^{P\times N\times d}$, in which the hidden state of road segment $v \in V$ at time step $t \in T_P$ is $hst_{v,t} \in R^d$. The outputs of multi-head GAT, multi-head GCN, temporal attention, and LSTM in the l^{th} layer are $HDS^l \in R^{P\times N\times d}$, $HPS^l \in R^{P\times N\times d}$, $HDT^l \in R^{P\times N\times d}$, and $HIT^l \in R^{P\times N\times d}$, respectively, while the dynamic spatial correlation, physical spatial dependency, dynamic temporal correlation, and inherent temporal dependency of road segment $v \in V$ at time step $t \in T_P$ are $hds^l_{v,t} \in R^d$, $hps^l_{v,t} \in R^d$, $hdt^l_{v,t} \in R^d$, and $hit^l_{v,t} \in R^d$, where P denotes the input series length of ST-Block.

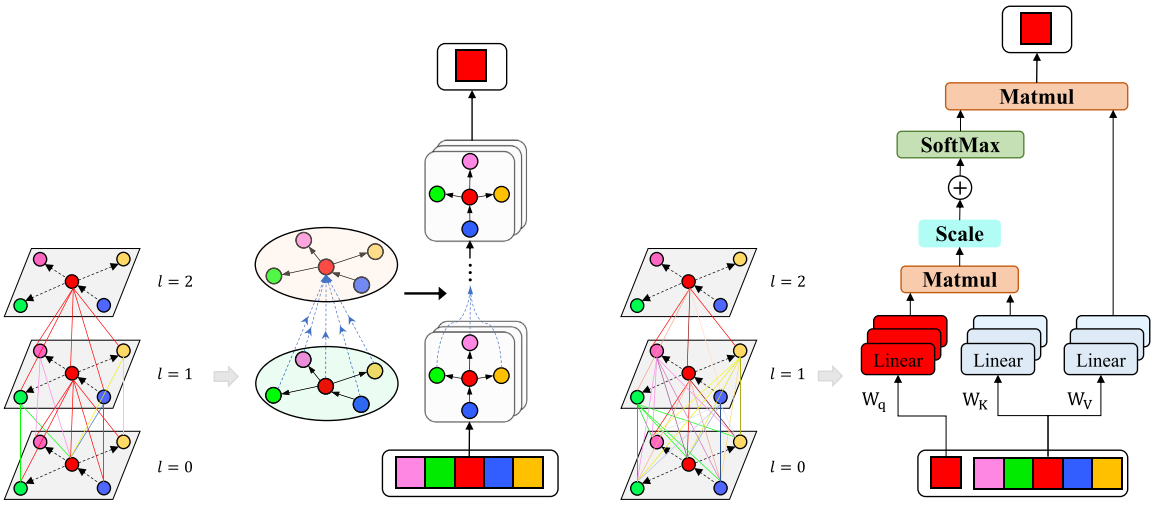
Since this study uses a nonlinear transformation function at high frequencies, it is first defined as:

$$f(x) = \mathrm{ReLU}\left(xW + b\right) \tag{2}$$

where x represents the input data, W and b denote the learnable parameters, and ReLU is the nonlinear activation function.

4.3.1. Physical spatial dependency

The traffic network is a directed graph, so the study of traffic speed must take into account the physical spatial dependency, i.e. the upstream and downstream traffic speeds that are directly associated with the target road segment. As seen in Fig. 5(a), the traffic speed on a red target road segment in the traffic network is directly related to one blue upstream road segment and three downstream yellow, green, and pink road segments. Additionally, for the example of the pink road segment, which perceives the first-order neighbor's (i.e., red) traffic information and updates the information itself, corresponding to the first layer graph, and then the second layer, based on the previous step, the blue segment, obtains the second-order neighbors' (i.e., yellow, green, and blue) traffic information. Therefore, a stacked multi-head GCN is designed to model the traffic direction and diffusion process, focusing on the traffic speed from several subspaces of road segments. The physical spatial dependency aggregation is used as input to a standard nonlinear transformation layer, in order to generate the l^{th} layer embedding of the graph nodes, as shown in Eqs. (3)



(a) Physical spatial dependency (b) Dynamic spatial correlation

Fig. 5. (a) Example of using a 2-layer multi-head GCN to extract the physical spatial dependency of traffic data from the bottom up, with local to global correlations. (b) Example of using a 2-layer multi-head GAT to extract the dynamic spatial correlation of traffic data, where each layer obtains the global correlations.

and (4). For node $v \in \mathrm{V}$, at time step $t \in \mathrm{T_P}$, the correlations between node v and v's first-order neighbors V_v in the m^{th} head GCN is,

$$hps_{v,t}^{l,m} = f\left(\widetilde{\mathrm{D}}^{-0.5}\widetilde{\mathrm{A}}_v\widetilde{\mathrm{D}}^{-0.5}\mathrm{HPS}_t^{l-1,m}\mathrm{W}^{l-1,m}\right) \tag{3}$$

$$hps_{v,t}^{l} = \mathrm{BN}\left(\|_{m=1}^{\mathrm{M}}\ hps_{v,t}^{l,m}\mathrm{W}_{ps}\right) + hps_{v,t}^{l-1} \tag{4}$$

where $\widetilde{\mathrm{D}}^{-0.5}\widetilde{\mathrm{A}}_v\widetilde{\mathrm{D}}^{-0.5}$ denotes the normalized adjacency matrix with added self-connections, $\widetilde{\mathrm{A}}_v = \mathrm{A}_v + \mathrm{I}_v$ is the adjacency matrix of the graph with added self-connections, $\mathrm{I} \in \mathrm{R}^{\mathrm{N}\times\mathrm{N}}$ represents the identity matrix, and $\widetilde{\mathrm{D}} \in \mathrm{R}^{\mathrm{N}\times\mathrm{N}}$ is the degree matrix of $\widetilde{\mathrm{A}} \in \mathrm{R}^{\mathrm{N}\times\mathrm{N}}$; $\|$ represents the concatenation operation; BN represents the batch normalization. Multi-head GCN can be built by stacking multi-convolutional layers in parallel; as Eqs. (3) and (4), the topological relationship between nodes $v \in \mathrm{V}$ and v's first-order neighbors V_v can be obtained, and the topological structure of the traffic network and the attributes of the road segment are encoded in order to obtain the physical spatial dependency $hps_{v,t} \in \mathrm{R}^d$ of node $v \in \mathrm{V}$ at time step $t \in \mathrm{T_P}$. The initial input of the multi-head GCN is $\mathrm{H} \in \mathrm{R}^{\mathrm{P}\times\mathrm{N}\times d}$, and the output is $\mathrm{HPS} \in \mathrm{R}^{\mathrm{P}\times\mathrm{N}\times d}$.

4.3.2. Dynamic spatial correlation

Besides the physical spatial dependency, the traffic speed of the target road segment is affected by other road segments, and the influence weight changes dynamically over time, called dynamic spatial correlation. For example, as shown in Fig. 5(b), the target red road segment is influenced by downstream road segments (i.e., yellow, green, and pink) in the traffic network during the congestion period, and the influence weight may be weakened as the congestion is relieved. To model the dynamic spatial correlation, we design a spatial attention approach base on stacked multi-head GAT to adaptively model the correlation between the target road segment and other road sections of the highway network, concentrating on road sections with huge correlations. In addition, unlike multi-head GCN, the first layer graph corresponds to a one-layer multi-head GAT model, whereas the second layer graph corresponds to a deep model that can extract more rich features than a shallow model, which is why Transformer (Vaswani et al., 2017) employs a deep multi-head self-attention mechanism for machine translation. For node $v_i \in \mathrm{V}$, at time step $t \in \mathrm{T_P}$, the correlation coefficient between nodes v_i and $v_j \in \mathrm{V}$ in the l^{th} spatial attention layer is,

$$\alpha_{v_i,v_j}^{l,m} = \frac{exp\left(s_{v_i,v_j}^{l,m}\right)}{\sum_{v_r\in\mathrm{V}} exp\left(s_{v_i,v_r}^{l,m}\right)} \tag{5}$$

where $s_{v_i,v_j}^{l,m} \in \mathrm{R}$ denotes the relevance between v_i and v_j in the l^{th} layer, V represents the input nodes of ST-Block.

The relevance $s_{v_i,v_j}^{l,m}$ can be obtained by the inner product of the query vector of node v_i and the key vector of node v_j,

$$s_{v_i,v_j}^{l,m} = \frac{\left\langle f_q^m\left(hds_{v_i,t}^{l-1}\right), f_k^m\left(hds_{v_j,t}^{l-1}\right)\right\rangle}{\sqrt{d}} \tag{6}$$

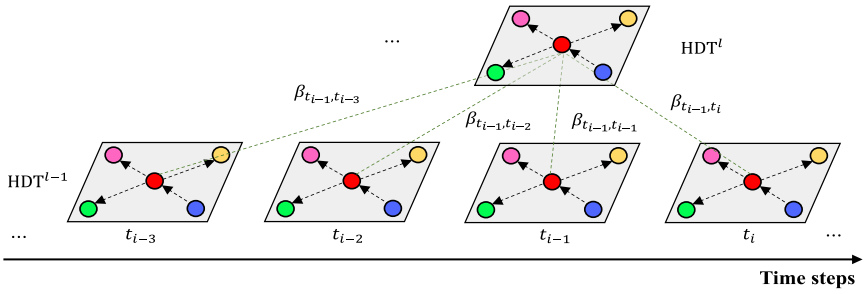


Fig. 6. Temporal attention model of the temporal correlation between different time steps.

where f_q^m and f_k^m are respectively the nonlinear transformation functions in the m^{th} head attention of the query vector and the key vector, and $\langle *, * \rangle$ represents the inner product operator.

After obtaining the correlation coefficient $\alpha_{v_i,v_j}^{l,m} \in \mathrm{R}$ between nodes v_i and v_j in the m^{th} head attention, the l^{th} layer dynamic spatial correlation $hds_{v_i,t}^l \in \mathrm{R}^d$ of node v_i at time step t can be formulated as,

$$hds_{v_i,t}^{l,m} = \sum_{v_r \in \mathrm{V}} \alpha_{v_i,v_r}^{l,m} f_v^m \left(hds_{v_r,t}^{l-1} \right) \tag{7}$$

$$hds_{v_i,t}^{l} = \mathrm{BN} \left(\|_{m=1}^{\mathrm{M}} hds_{v_i,t}^{l,m} \mathrm{W}_{ds} \right) + hds_{v_i,t}^{l-1} \tag{8}$$

where f_v^m is the nonlinear transformation function in the m^{th} head attention of the value vector. The final dynamic spatial correlation $hds_{v_i,t} \in \mathrm{R}^d$ of node v_i can be calculated using Eqs. (5)–(8) at time step t. The initial input of spatial attention is $[\mathrm{H}, \mathrm{STE}[:\mathrm{P}]] \in \mathrm{R}^{\mathrm{P}\times\mathrm{N}\times 2d}$, and the output is $\mathrm{HDS} \in \mathrm{R}^{\mathrm{P}\times\mathrm{N}\times d}$.

The spatial correlations $hds_{v_i,t} \in \mathrm{R}^d$ and $hps_{v_i,t} \in \mathrm{R}^d$ of road segment v_i can be calculated using Eqs. (3)–(8) at time step $t \in \mathrm{T_P}$, and the compound spatial correlation $hs_{v_i,t} \in \mathrm{R}^d$ can be formulated as,

$$hs_{v_i,t} = \mathrm{BN} \left(f \left(\left[hds_{v_i,t}, hps_{v_i,t} \right] \right) \right) \tag{9}$$

where $[*, *]$ represents the binary concatenation operation. The initial inputs of feed-forward layer are $\mathrm{HPS} \in \mathrm{R}^{\mathrm{P}\times\mathrm{N}\times d}$ and $\mathrm{HDS} \in \mathrm{R}^{\mathrm{P}\times\mathrm{N}\times d}$, and the output is $\mathrm{HS} \in \mathrm{R}^{\mathrm{P}\times\mathrm{N}\times d}$.

4.3.3. Dynamic temporal correlation

The traffic speed on each road segment in a traffic network is influenced by the previous traffic speeds and affects the future. The influence weight varies dynamically over time, known as dynamic temporal correlation (Park et al., 2020). For instance, the traffic speed during the morning rush hour is negatively influenced by earlier traffic speeds, which may gradually exacerbate congestion before releasing it. In this study, we design a temporal attention approach to adaptively model the correlations between different time steps, as shown in Fig. 6.

For road segment node $v \in \mathrm{V}$, at time step t_i, the correlation coefficient between time steps t_i and t_j in the l^{th} temporal attention layer is,

$$\beta_{t_i,t_j}^{l,m} = \frac{exp\left(\mu_{t_i,t_j}^{l,m} \right)}{\sum_{t_r \in \mathrm{T_P}} exp\left(\mu_{t_i,t_r}^{l,m} \right)} \tag{10}$$

where $\mu_{t_i,t_j}^{l,m}$ denotes the relevance between t_i and t_j in the l^{th} layer, $\mathrm{T_P}$ denotes a set of time steps before t_P.

The relevance can be obtained by the inner product of the query vector of node v at time step t_i and the key vector of node v at time step t_j,

$$\mu_{t_i,t_j}^{l,m} = \frac{\left\langle f_q^m \left(hdt_{v,t_i}^{l-1} \right), f_k^m \left(hdt_{v,t_j}^{l-1} \right) \right\rangle}{\sqrt{d}} \tag{11}$$

Once the correlation coefficient $\beta_{t_i,t_j}^{l,m}$ in the m^{th} head attention is obtained, the l^{th} layer temporal correlation hdt_{v,t_i}^l of node v at time step t_i can be formulated as,

$$hdt_{v,t_i}^{l,m} = \sum_{t_r \in \mathrm{T_P}} \beta_{t_i,t_r}^{l,m} f_v^m \left(hdt_{v,t_r}^{l-1} \right) \tag{12}$$

$$hdt_{v,t_i}^{l} = \mathrm{BN} \left(\|_{m=1}^{\mathrm{M}} hdt_{v,t_i}^{l,m} \mathrm{W}_{dt} \right) + hdt_{v,t_i}^{l-1} \tag{13}$$

The final temporal correlation $hdt_{v,t_i} \in \mathrm{R}^d$ of node $v \in \mathrm{V}$ can be calculated using Eqs. (10)–(13) at time step t_i. The initial input of temporal attention is $\mathrm{H} \in \mathrm{R}^{\mathrm{P}\times\mathrm{N}\times d}$, and the output is $\mathrm{HDT} \in \mathrm{R}^{\mathrm{P}\times\mathrm{N}\times d}$.

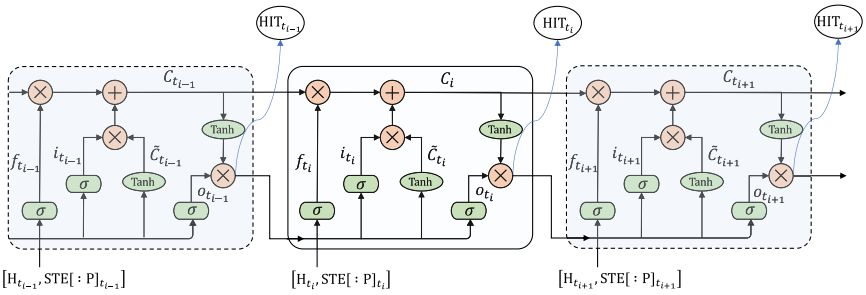


Fig. 7. The architecture of the LSTM model.

4.3.4. Inherent temporal dependency

The traffic speed of the target road segment changes dynamically and presents continuity with time, called inherent temporal dependency. For example, in general, the traffic speed of the target road segment changes continuously rather than randomly; that is, the traffic speed at the current moment is consistent with the previous. The inherent temporal dependency is one of the most important characteristics of the traffic data. The temporal attention mechanism (Park et al., 2020) is applied to model the correlations between different time steps. However, the traffic trend is challenging to describe via the attention mechanism, necessitating the modeling account for continuously dynamic speeds. Fortunately, a long short-term memory network (LSTM) has a unique character that records the traffic trend and learns the traffic pattern through a memory unit. Several studies tackled the temporal dimension (Qu et al., 2021; Ma et al., 2015; Gu et al., 2019; Meng et al., 2020; Yi and Bui, 2020; Wang et al., 2019). LSTM is the mainstream method of inherent temporal dependency extraction, and the working process is presented in Fig. 7.

Long short-term memory network is used to extract the inherent temporal correlation of node $v \in \mathrm{V}$. We assume that i, f and o respectively represent the input gate, forget gate, and output gate, $\left[\mathrm{H}_{v,t_i}, \mathrm{STE}[\ :\ \mathrm{P}]_{v,t_i}\right] \in \mathrm{R}^{2d}$ is the input of node v at time step t_i, $hit_{v,t_i} \in \mathrm{R}^d$ is the output at time step t_i, σ represents the sigmoid function, $tanh$ denotes the Tanh function, W and b represent the weight and bias parameters of the LSTM, respectively. The steps of LSTM for the extraction of inherent temporal correlation are summarized as follows,

Step 1: LSTM selectively forgets the feature information of the cell state $C_{v,t_{i-1}}$ at time step t_i:

$$\begin{cases} f_{v,t_i} = \sigma\left(\left[\mathrm{H}_{v,t_i}, \mathrm{STE}[\ :\ \mathrm{P}]_{v,t_i}\right]\mathrm{W}_f + hit_{v,t_{i-1}}\mathrm{W}_f + b_f\right) \\ C'_{v,t_i} = f_{v,t_i} \odot C_{v,t_{i-1}} \end{cases} \tag{14}$$

Step 2: LSTM selects the important information from the input features, which is then used to update the state cell C'_{v,t_i}:

$$\begin{cases} \tilde{C}_{v,t_i} = tanh\left(hit_{v,t_{i-1}}\mathrm{W}_C + \left[\mathrm{H}_{v,t_i}, \mathrm{STE}[\ :\ \mathrm{P}]_{v,t_i}\right]\mathrm{W}_C + b_C\right) \\ i_{v,t_i} = \sigma\left(hit_{v,t_{i-1}}\mathrm{W}_i + \left[\mathrm{H}_{v,t_i}, \mathrm{STE}[\ :\ \mathrm{P}]_{v,t_i}\right]\mathrm{W}_i + b_i\right) \\ C_{v,t_i} = C'_{v,t_i} + i_{v,t_i} \odot \tilde{C}_{v,t_i} \end{cases} \tag{15}$$

Step 3: finally, the LSTM output is determined:

$$\begin{cases} o_{v,t_i} = \sigma\left(hit_{v,t_{i-1}}\mathrm{W}_o + \left[\mathrm{H}_{v,t_i}, \mathrm{STE}[\ :\ \mathrm{P}]_{v,t_i}\right]\mathrm{W}_o + b_o\right) \\ hit_{v,t_i} = o_{v,t_i} \odot tanh\left(C_{v,t_i}\right) \end{cases} \tag{16}$$

The final inherent temporal correlation $hit_{v,t_i} \in \mathrm{R}^d$ of node $v \in \mathrm{V}$ can be calculated using Eqs. (14)–(16) at time step t_i. The initial input of LSTM is $[\mathrm{H}, \mathrm{STE}[\ :\ \mathrm{P}]] \in \mathrm{R}^{\mathrm{P}\times\mathrm{N}\times 2d}$, and the output is $\mathrm{HIT} \in \mathrm{R}^{\mathrm{P}\times\mathrm{N}\times d}$.

The temporal correlations $hit_{v,t_i} \in \mathrm{R}^d$ and $hdt_{v,t_i} \in \mathrm{R}^d$ of road $v \in \mathrm{V}$ can be calculated using Eqs. (10)–(16) at time step t_i, and the compound temporal correlation ht_{v,t_i} can be formulated as,

$$ht_{v,t_i} = \mathrm{BN}\left(f\left(hit_{v,t_i} + hdt_{v,t_i}\right)\right) \tag{17}$$

The initial inputs of feed-forward layer are $\mathrm{HIT} \in \mathrm{R}^{\mathrm{P}\times\mathrm{N}\times d}$ and $\mathrm{HDT} \in \mathrm{R}^{\mathrm{P}\times\mathrm{N}\times d}$, and the output is $\mathrm{HT} \in \mathrm{R}^{\mathrm{P}\times\mathrm{N}\times d}$.

4.3.5. Fusion gate mechanism

To obtain the final spatio-temporal correlations, we designed a fusion gate mechanism to adaptively fuse the compound spatial correlations $\mathrm{HS} \in \mathrm{R}^{\mathrm{P}\times\mathrm{N}\times d}$ and compound temporal correlations $\mathrm{HT} \in \mathrm{R}^{\mathrm{P}\times\mathrm{N}\times d}$, do not add additional parameters, and the output is

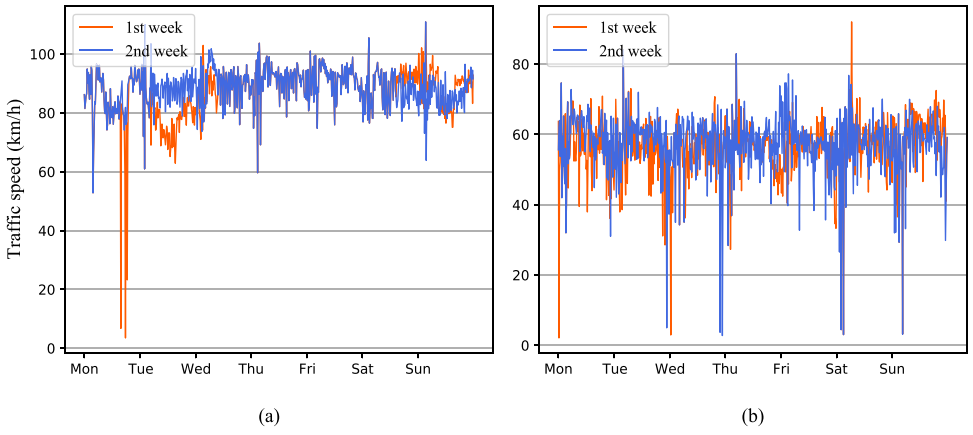


Fig. 8. Example traffic speeds of two road segments on different weeks.

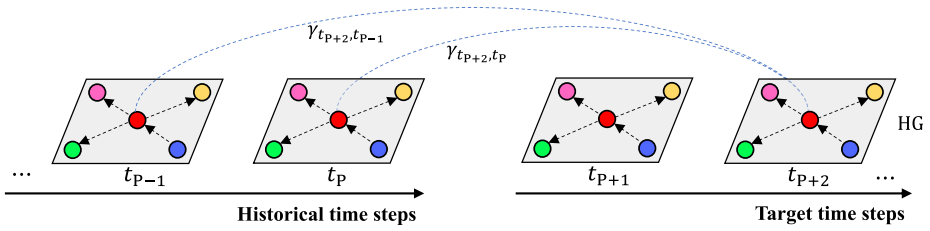


Fig. 9. Generative inference calculates the correlation between historical time step and target time step directly.

$\mathrm{HST} \in \mathrm{R}^{\mathrm{P}\times\mathrm{N}\times d}$. The working process of fusion gate mechanism is,

$$\mathrm{HST} = \mathcal{Z} \odot \mathrm{HS} + (1 - \mathcal{Z}) \odot \mathrm{HT} \tag{18}$$

with

$$\mathcal{Z} = \sigma\left(\mathrm{HS} \odot \mathrm{HT}\right) \tag{19}$$

where σ represents *sigmoid* activation function, and $\mathcal{Z} \in \mathrm{R}^{\mathrm{P}\times\mathrm{N}\times d}$ is weight vector that controls the flow of spatial and temporal representations at each time step.

4.4. Generative inference

Existing approaches have trouble resolving the error propagation problem; the accuracy of the forecast of the target node is bound by the prediction results of its previous time steps and other nodes. For instance, using ST-GRAT to forecast the second-time step speed (Park et al., 2020), the inference process incorporates the first-time step results and employs spatial attention to highlight the interdependencies between anticipated variables. In this dynamic decoding procedure, the prediction errors in the first-time step will permeate to the second-time step, causing an accumulation of mistakes. Fortunately, a crucial property that motivates us, as depicted in Fig. 8, is that the traffic speed trends are comparable during the identical period on the target road segment, defined as traffic patterns. In addition, the two distinct color curves overlap approximately in the partial road segment shown in Fig. 8(a). Subsequently, a generative style transformer is devised based on a multi-head self-attention mechanism that replaces the decoder to infer future long-term traffic speeds by one step, and this architecture is termed generative inference.

In this paper, we design a bridge transformer (BridgeTrans) to pay attention to the relationships between historical spatio-temporal features $\mathrm{HST} \in \mathrm{R}^{\mathrm{P}\times\mathrm{N}\times d}$ and target sequence $\mathrm{STE}[\mathrm{P}:\mathrm{P}+\mathrm{Q}] \in \mathrm{R}^{\mathrm{Q}\times\mathrm{N}\times d}$ without speed information to generate the target hidden outputs $\mathrm{HG} \in \mathrm{R}^{\mathrm{Q}\times\mathrm{N}\times d}$ rather than a dynamic step-by-step decoding way. For example, to generate the target hidden outputs $\mathrm{HG}_{t_{\mathrm{P}+i}} \in \mathrm{R}^{\mathrm{N}\times d}$ at time step $t_{\mathrm{P}+i}$, we assess the correlation between target time step $t_{\mathrm{P}+i}$ and historical time step $t_j \in \mathrm{T_P}$, avoiding the prediction error propagation in the spatial and temporal dimensions, as shown in Fig. 9.

For node $v \in \mathrm{V}$, the correlation coefficient $\gamma^{l,m}_{t_{\mathrm{P}+i},t_j}$ between target time step $t_{\mathrm{P}+i}$ and historical time step t_j in the l^{th} BridgeTrans layer is calculated,

$$\gamma^{l,m}_{t_{\mathrm{P}+i},t_j} = \frac{exp\left(\lambda^{l,m}_{t_{\mathrm{P}+i},t_j}\right)}{\sum_{t_r\in\mathrm{T_P}} exp\left(\lambda^{l,m}_{t_{\mathrm{P}+i},t_r}\right)} \tag{20}$$

where $\lambda^{l,m}_{t_{P+i},t_j}$ denotes the relevance between time steps t_{P+i} and t_j in the l^{th} layer, which can be obtained by the inner product of the query vector of node v at time step t_{P+i} and the key vector of node v at time step t_j,

$$\lambda^{l,m}_{t_{P+i},t_j} = \frac{\left\langle f^m_q\left(hg^{l-1}_{v,t_{P+i}}\right), f^m_k\left(hst_{v,t_j}\right)\right\rangle}{\sqrt{d}} \tag{21}$$

After obtaining the correlation coefficient $\gamma^{l,m}_{t_{P+i},t_j}$ in the m^{th} head attention, the l^{th} layer hidden output $hg^{l}_{v,t_{P+i}}$ of node v at time step t_{P+i} can be updated using Eqs. (20) and (21).

$$hg^{l,m}_{v,t_{P+i}} = \sum_{t_r \in T_P} \gamma^{l,m}_{t_{P+i},t_r} f^m_v\left(hst_{v,t_r}\right) \tag{22}$$

$$hg^{l}_{v,t_{P+i}} = \mathrm{BN}\left(\|^{M}_{m=1}\, hg^{l,m}_{v,t_{P+i}} W_g\right) + hg^{l-1}_{v,t_{P+i}} \tag{23}$$

The final hidden output $hg_{v,t_{P+i}} \in R^d$ of node $v \in V$ at time step t_{P+i} can be calculated using Eqs. (20)–(23). The initial input of generative inference is $HST \in R^{P\times N\times d}$ and $STE[P:P+Q] \in R^{Q\times N\times d}$, and the hidden outputs of generative inference is $HG \in R^{Q\times N\times d}$.

4.5. Loss function

For the highway traffic speed prediction, the hidden outputs of the generative inference are directly fed into the fully connected layer in order to generate the predicted values:

$$\hat{Y} = HG \bullet W_{\hat{Y}} \tag{24}$$

where $W_{\hat{Y}} \in R^{d\times 1}$ represents the weight parameter of the fully connected layer, and $\bullet$ is a matrix multiplication operation.

The loss function of STGIN is defined as the mean absolute error (MAE) between observed values $Y \in R^{Q\times N\times 1}$ and predicted values $\hat{Y} \in R^{Q\times N\times 1}$,

$$L(\theta) = \frac{1}{Q\times N}\sum_{j=1}^{N}\sum_{i=1}^{Q}\left|Y_{i,j} - \hat{Y}_{i,j}\right| + \frac{\lambda}{2}\|\theta\|^2 \tag{25}$$

where λ is the regularization parameter, and θ denotes all the learnable parameters in the STGIN.

5. Experiments

5.1. Data description

The highway traffic data used in this study is provided by the ETC intelligent monitoring sensors at the gantries and the toll stations of the highway in Yinchuan City, Ningxia Province, China. The 66 ETC intelligent monitoring sensors record the vehicle driving data in real time, including 13 highway toll stations (each toll station contains an entrance and exit) and 40 highway gantries. Therefore, these monitoring sensors divide the highway network into 108 road segments, as shown in Fig. 10. The traffic speed of each road segment is measured at a certain frequency, such that one sample is measured every 15 min, and therefore the time series form of the traffic speed is obtained. The highway traffic data includes three factors: traffic speed, timestamp, and road segment index. The time span is from June 1, 2021 to August 31, 2021. The road segment index does not change over time, and there are 108 road segments in total, that is, 108 road segment indexes. In the experiment, 70% of the data are used as the training set, 10% of the data are used as the validation set, and the remaining 20% are considered as the test set.[1]

5.2. Baselines and metrics

The proposed model for highway traffic speed prediction is compared with the following prediction methods:

ARIMA, is a traditional time series prediction method that combines the moving average and autoregressive components in order to model the historical time series data (Duan et al., 2016).

SVR, is a traditional pattern classification and regression technique called support vector regression model for short-term traffic forecasting (Hong, 2011).

LSTM_BILSTM, is an enhanced time series model based on long short-term memory neural network (LSTM) and bidirectional LSTM (Bi-LSTM), which uses the multiple LSTM and Bi-LSTM layers stacked together to merge the time information (Ma et al., 2021).

FI-RNNs, which combines the time series data and uses a stacked RNN and a sparse autoencoder, in order to learn the sequential features of the traffic data (Qu et al., 2021).

PSPNN, which is composed of a hierarchical CNN and a bidirectional LSTM (Bi-LSTM) network, that extract the temporal and spatial correlations of the historical data, in order to perform the path-based speed prediction (Yang et al., 2021).

[1] https://github.com/zouguojian/STGIN/tree/main/data.

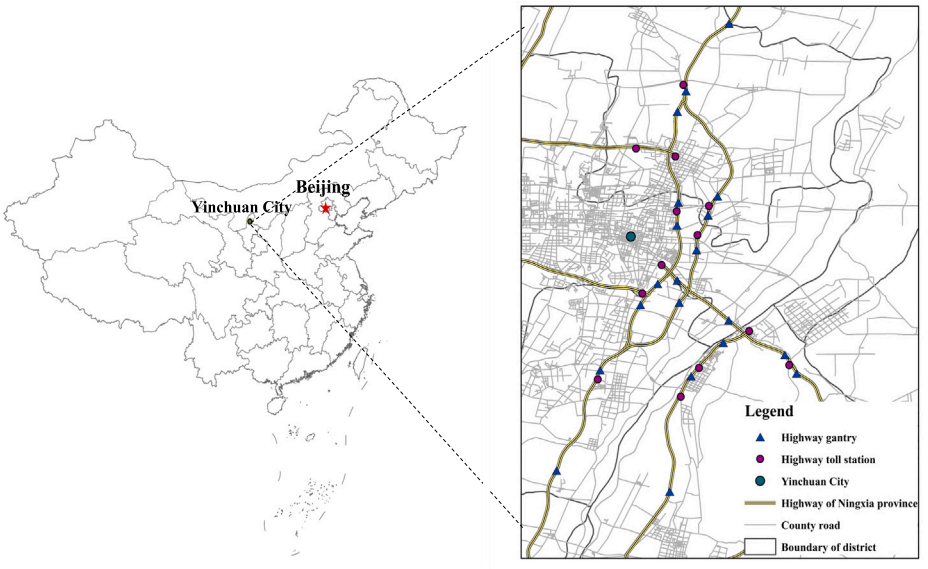


Fig. 10. Study area.

MDL, is used to predict the lane-level short-term traffic speed. It consists of a convolutional long and short-term memory (Conv-LSTM) layer, a convolutional layer, and a fully connected layer (Lu et al., 2020).
T-GCN, which combines the graph convolutional network (GCN) and gate recurrent unit (GRU) to model the spatio-temporal correlations (Zhao et al., 2019).
DCRNN, is a diffusion convolutional recurrent neural network incorporating spatial and temporal dependency into traffic prediction (Li et al., 2018).
STGCN, is a novel deep learning framework, spatio-temporal graph convolutional networks (STGCN), which combines graph convolution with 1-D convolution (Yu et al., 2018).
AGCRN, is an adaptive graph convolutional recurrent network based on the node adaptive parameter learning module, data-adaptive graph generation module, and recurrent networks to automatically captures fine-grained spatial and temporal dependencies (Bai et al., 2020).
ASTGCN, is an attention-based spatial–temporal graph convolutional network to model dynamic spatial–temporal correlations for traffic forecasting (Guo et al., 2019), and the version released here only consists of the recent component.[2]
Graph-WaveNet, is a graph neural network architecture that captures the hidden spatial correlation in the data and handles very long sequences (Wu et al., 2019).
GMAN, is a graph multi-attention network that utilizes spatio-temporal- and transform-attention to model the spatial and temporal features of historical and future time steps (Zheng et al., 2020).
ST-GRAT, it is a novel spatio-temporal graph attention model based on self-attention mechanism that effectively captures dynamic spatio-temporal correlations of the road network (Park et al., 2020).
MTGNN, is a general graph neural network for multivariate time series data, which extracts the uni-directed relations through a graph learning module and captures the spatial and temporal dependencies via a mix-hop propagation- and a dilated inception-layer (Wu et al., 2020a).

In order to evaluate the prediction performance of the STGIN model, three metrics are used to determine the difference between the observed values $Y \in R^{Q\times N\times 1}$ and the predicted values $\hat{Y} \in R^{Q\times N\times 1}$: the mean absolute error (MAE), root mean square error (RMSE), and mean absolute percentage error (MAPE). These metrics are presented in Eqs. (26)–(28), respectively.

$$\mathrm{MAE} = \frac{1}{D\times Q}\sum_{j=1}^{Q}\sum_{i=1}^{D}\left|Y_{i,j} - \hat{Y}_{i,j}\right| \tag{26}$$

$$\mathrm{RMSE} = \sqrt{\frac{1}{D\times Q}\sum_{j=1}^{Q}\sum_{i=1}^{D}\left(Y_{i,j} - \hat{Y}_{i,j}\right)^2} \tag{27}$$

$$\mathrm{MAPE} = \frac{100\%}{D\times Q}\sum_{j=1}^{Q}\sum_{i=1}^{D}\frac{\left|Y_{i,j} - \hat{Y}_{i,j}\right|}{Y_{i,j}} \tag{28}$$

[2] https://github.com/guoshnBJTU/ASTGCN-r-pytorch.

Table 1
Model hyperparameter.

Layer name	Hyperparameter	Values
Multi-head GCN	Hidden nodes	64
	Number of blocks	2
	Number of heads (M)	1
Multi-head GAT	Hidden nodes	64
	Number of block	2
	Number of heads (M)	8
LSTM	Hidden nodes × number of layers	64 × 1
Temporal attention	Hidden nodes	64
	Number of blocks	2
	Number of heads (M)	8
Residual connection	Hidden nodes × number of layers	64 × 1
Feed forward	Hidden nodes × number of layers	64 × 1
Generative inference	Hidden nodes	64
	Number of blocks	1
	Number of heads (M)	8
Fully connected layer	Hidden nodes × number of layers	128 × 1 64 × 1 1 × 1
–	Batch size	32
–	Dropout	0.3
–	Decay rate	0.9
–	Learning rate	0.001
–	Epochs	100
–	λ	0.001
–	Training method	Adam

where D is the number of samples in test set. Note that low MAE, RMSE, and MAPE values indicate a more accurate prediction performance.

5.3. Experimental settings

We applied the grid search in the proposed STGIN method to find the optimal model on the validation dataset. Especially among all candidate hyperparameter selections, every possibility is tried through loop traversal, and the hyperparameter group with the best performance on the validation dataset is selected as the final result. Note for these continuous hyperparameter values, sample at equal intervals. For each hyperparameter group, the optimal parameters of the proposed STGIN model and baseline techniques are determined during the training process with minimal MAE on the validation set, and specific processing follows,

In the experiment, the maximum number of epochs is 100; and the batch size is 32, which divides the training set into 183 iterations in a single epoch. Updating the model's parameters via backpropagation with a batch of data is called one iteration. Specifically, we evaluate the prediction model on the validation set after one epoch. If the MAE on the validation set is improved, the model parameters are updated and recorded to replace the last one saved. In addition, when the forecasting performance of the prediction model on the validation set is optimal, the training process ends after many parameter adjustments and experiments. We use an early-stop mechanism in all experiments, and the number of early-stop epochs is set to 10, defined as patience. The early-stop mechanism means the training stops early if the MAE on the validation set is not decreased under the patience before the maximum number of epochs. Finally, the prediction result is obtained by iterating all the samples in the test set. We set the target time steps Q and historical time steps P to 12, respectively, representing the time span is 360 min.

After multiple training steps, the final model framework hyperparameters are determined. Table 1 presents the number of layers, hidden nodes, and related hyperparameters of the STGIN model. We implement the STGIN and baselines in TensorFlow and PyTorch. The server's one NVIDIA Tesla V100S-PCIE-32 GB GPU and 24 CPU cores are used for model training and testing. Note that the **implementation codes** of the proposed model-**STGIN** and **baseline methods** are open source, and are available at the personal **GitHub homepage**.[3]

5.4. Experimental results

5.4.1. Predicting performance comparison

The current traffic speed prediction studies mainly focus on short-term prediction, which may not be enough to meet the needs of the actual application scenarios. The long-term prediction of highway traffic speed is a challenging task, and it is related to the precise

[3] https://github.com/zouguojian/STGIN.

Table 2
Performance comparison of different approaches for long-term highway traffic speed prediction.

Model	Horizon 3			Horizon 6			Horizon 12			Average		
	MAE	RMSE	MAPE	MAE	RMSE	MAPE	MAE	RMSE	MAPE	MAE	RMSE	MAPE
ARIMA	6.031	9.415	13.036%	6.150	9.555	13.211%	6.236	9.670	13.336%	6.131	9.544	13.183%
SVR	5.735	9.535	12.655%	5.929	9.737	12.952%	6.074	9.865	13.210%	5.893	9.695	12.906%
LSTM_BILSTM	5.750	9.336	20.106%	6.064	9.619	20.485%	6.499	10.008	21.021%	6.067	9.625	20.472%
FI-RNNs	5.644	9.256	12.673%	5.953	9.583	13.197%	6.368	10.010	14.975%	5.949	9.579	13.479%
PSPNN	5.601	9.181	12.629%	5.814	9.419	13.075%	6.112	9.723	13.586%	5.824	9.428	13.211%
MDL	5.558	9.181	12.798%	5.656	9.282	12.677%	5.809	9.433	13.200%	5.644	9.266	12.927%
T-GCN	5.648	9.235	12.698%	5.700	9.308	12.839%	5.868	9.496	13.664%	5.726	9.334	12.980%
STGCN	5.306	8.887	12.401%	5.455	9.022	12.716%	5.604	9.166	12.907%	5.442	9.010	12.663%
DCRNN	5.277	8.860	11.917%	5.346	8.946	11.977%	5.462	9.065	12.224%	5.346	8.939	12.008%
AGCRN	5.247	8.923	12.363%	5.308	8.964	12.493%	5.375	9.027	13.116%	5.298	8.963	12.395%
ASTGCN	5.453	9.014	13.222%	5.583	9.172	13.185%	5.752	9.290	13.671%	5.571	9.137	13.270%
Graph-WaveNet	5.210	**8.804**	12.448%	5.267	**8.866**	12.232%	5.357	**8.954**	12.102%	5.264	**8.861**	12.166%
GMAN	5.318	8.930	12.762%	5.333	8.951	12.679%	5.403	9.032	12.757%	5.350	8.966	12.741%
ST-GRAT	5.556	9.216	13.296%	5.544	9.107	13.099%	5.943	9.464	13.182%	5.623	9.185	13.061%
MTGNN	5.441	9.071	13.476%	5.512	9.158	13.644%	5.613	9.268	14.324%	5.514	9.158	13.668%
STGIN	**5.119**	8.899	**11.618%**	**5.142**	8.935	**11.622%**	**5.225**	9.001	**11.666%**	**5.154**	8.940	**11.630%**

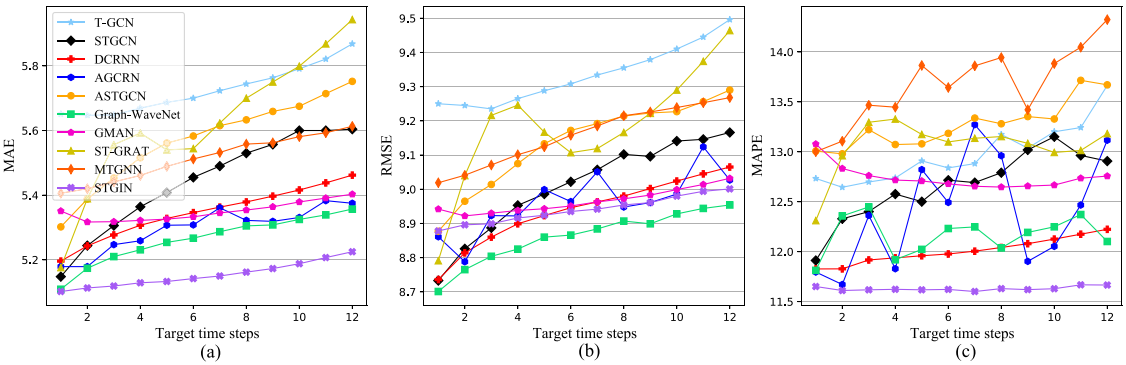


Fig. 11. Long-term highway traffic speed prediction ability of prediction models with GNN module. (a) MAE changes of different models in different target time steps; (b) RMSE changes; (c) MAPE changes.

control of the highway traffic in the future. In order to verify the advantages of STGIN in long-term prediction, Table 2 presents the performance of the baseline models and the proposed model for the highway traffic speed prediction in the next different target time steps, respectively. For instance, the input time steps 6:00–9:00 are applied to anticipate the traffic speed during 9:00–12:00 (i.e., twelve time steps).

The performance of ARIMA is lower than that of all the other baseline models, which also demonstrates the difficulty of long-term highway traffic speed prediction. The traditional machine learning and deep learning methods perform better than the statistical methods. For example, compared with ARIMA for the next twelve-time steps prediction, SVR, LSTM_BILSTM, FI-RNNs, PSPNN, MDL, T-GCN, STGCN, DCRNN, AGCRN, ASTGCN, Graph-WaveNet, GMAN, ST-GRAT, MTGNN, and STGIN increased MAE by 4.039%, 1.055%, 3.059%, 5.271%, 8.629%, 7.073%, 12.661%, 14.684%, 15.723%, 10.052%, 16.470%, 14.598%, 9.034%, 11.190%, 18.956%, respectively. This is due to the fact that they are more suitable for extracting nonlinear correlations of traffic data. Therefore, the existing researches are gradually moving from statistical methods to machine learning methods.

Spatial correlation is another essential element in traffic forecasting. The prediction precisions of LSTM_BILSTM and FI-RNNs are lower than PSPNN, MDL, T-GCN, STGCN, DCRNN, AGCRN, Graph-WaveNet, and MTGNN because these models merely consider temporal correlation and ignore the impact of spatial correlation of traffic data. For instance, compared with LSTM_BILSTM and FI-RNNs for horizon twelve prediction, MDL reduced MAE by 10.617% and 8.778%, respectively; lowered the RMSE by 5.745% and 5.664%; improved MAPE by 37.206% and 11.853%. In addition, compared with MDL for horizon twelve prediction, STGCN, DCRNN, AGCRN, Graph-WaveNet, and MTGNN reduced by 3.529%, 5.973%, 7.471%, 7.781%, and 3.374% in terms of MAE, respectively; by 3.933%, 5.280%, 6.130%, 6.733%, and 2.303% for the next twelve time steps. These comparisons verified the ability of GCNs to model the spatial property in non-Euclidean space than CNNs. Note that the spatio-temporal prediction model T-GCN based on GCN and GRU applied in non-Euclidean space, its prediction accuracy is relatively lower than MDL because T-GCN is challenging to converge and needs training of almost 5000 epochs due to its combination limitation.

In the baseline models, the spatio-temporal dependency models based on GATs and attention mechanisms include ASTGCN, GMAN, and ST-GRAT. Compared with ASTGCN, GMAN, and ST-GRAT, the lower prediction precision of PSPNN and MDL because highway traffic data is non-Euclidean structure. Compared with MDL for the horizon six, ASTGCN, GMAN, and ST-GRAT improved

Table 3
Performance of the different time steps prediction for distinguished variants.

Model	Horizon 6			Horizon 12			Average		
	MAE	RMSE	MAPE	MAE	RMSE	MAPE	MAE	RMSE	MAPE
w/o semantic enhancement	5.234	9.107	12.564%	5.340	9.197	12.765%	5.255	9.127	12.601%
w/o dynamic spatial correlation	5.185	9.035	12.775%	5.276	9.124	12.943%	5.200	9.048	12.796%
w/o physical spatial dependency	5.169	9.058	11.499%	5.246	9.127	11.662%	5.179	9.069	11.548%
w/o dynamic temporal correlation	5.233	9.064	11.522%	5.296	9.109	11.671%	5.243	9.066	11.560%
w/o inherent temporal dependency	5.162	8.977	**11.378%**	5.247	9.045	11.545%	5.174	8.982	**11.422%**
w/o fusion gate mechanism	5.231	9.079	11.478%	5.314	9.145	11.654%	5.245	9.088	11.515%
w/o ST-Block	5.226	9.053	11.503%	5.273	9.085	11.623%	5.227	9.052	11.514%
w/o generative inference	5.278	9.363	11.539%	5.343	9.407	**11.512%**	5.257	9.308	11.496%
STGIN (ours)	**5.142**	**8.935**	11.622%	**5.225**	**9.001**	11.666%	**5.154**	**8.940**	11.630%

by 1.291%, 5.711%, and 1.980% in terms of MAE, respectively; by 1.293%, 5.209%, and 0.372% for the next twelve time steps. In addition, GMAN has a better prediction performance than ASTGCN, which the coupling mode of spatial correlation and temporal dependency may cause. For instance, GMAN lowered MAE, RMSE, and MAPE by 2.070%, 1.320%, and 3.130% compared with ASTGCN for the next twelve time steps. Moreover, compared with ST-GRAT for the next twelve time steps, GMAN reduced MAE, RMSE, and MAPE by 2.070%, 1.320%, and 3.130%, respectively. The experimental results demonstrate that decoding is the main factor that decides whether error propagation is in spatial and temporal dimensions. In this paper, the research methods of GNNs, especially GMAN, inspire our work.

Absorbing the experience of GNNs in modeling spatial correlation, we construct a spatio-temporal generative inference network for long-term highway traffic speed prediction. The proposed model first uses a semantic enhancement module to model the contextual semantics of traffic speed, uses the ST-Block to extract the spatio-temporal correlations of the highway network, and designs a generative inference to eliminate the problem of long-term prediction error propagation in the spatial and temporal dimensions. As depicted in Table 2 and Fig. 11, the proposed method outperforms all spatio-temporal baseline techniques regarding the three metrics. For example, compared with T-GCN, STGCN, DCRNN, AGCRN, ASTGCN, Graph-WaveNet, GMAN, ST-GRAT, and MTGNN for horizon twelve prediction, STGIN improved MAE by 10.958%, 6.763%, 4.339%, 2.791%, 9.162%, 2.464%, 3.294%, 12.081%, and 6.913%, respectively; 9.990%, 5.292%, 3.591%, 2.718%, 7.485%, 2.090%, 3.664%, 8.341%, and 6.529% for the next twelve time steps. We argue that long-term traffic speed forecasting is more beneficial to practical applications, e.g., it allows transportation agencies to have more time to take actions to optimize the traffic according to the prediction. Therefore, when STGIN obtains a high level of prediction performance, the benefits become more apparent in the long-term horizon (e.g., 180 min ahead).

Note that the proposed STGIN need to fit speed fluctuation, which leads to prediction shift in rare time steps (shown in the following Figs. 13 and 14), causing the performance to be inferior to partial baselines, such as Graph-WaveNet, in terms of RMSE. This phenomenon will be further demonstrated in Sections 5.4.4 and 5.4.5.

5.4.2. Influence of each component

To verify the effectiveness of each essential component of the proposed STGIN model, eight variants are compared in this part. W/O semantic enhancement, like GMAN, uses a fully connected layer to replace. W/O dynamic- and physical-spatial correlations, STGIN respectively removes the spatial attention and multi-head GCN model. W/O dynamic- and inherent-temporal correlations, STGIN respectively discards the temporal attention and LSTM network. W/O fusion gate, using the addition operation to replace the adaptively combined mechanism. W/O ST-Block, STGIN uses a STAtt Block of GMAN instead. W/O generative inference, STGIN uses dynamic step-by-step decoding to replace. As Table 3 shows, each component's contribution to STGIN is heightened in the following comparisons,

W/O Semantic Enhancement The prediction precision without semantic enhancement module is lower than STGIN because it ignores the contextual semantic breaks caused by speed fluctuations. For example, for horizon twelve forecasting, compared with the STGIN, the MAE, RMSE, and MAPE increased by 2.154%, 2.131%, and 8.609%, respectively; by 1.922%, 2.049%, and 7.706% for the next twelve time steps. Irresistible factors, such as agglomerate fog, typically cause traffic speed fluctuations on the highway network. The experimental results indicate that the semantic enhancement component is an effective strategy for resolving the speed fluctuation problem in speed prediction.

W/O Dynamic Spatial Correlation & **W/O Physical Spatial Dependency** Compared with STGIN, *w/o physical spatial dependency module* respectively, increased MAE and RMSE by 0.400% and 1.381% for horizon twelve; by 0.483% and 1.422% for the next twelve time steps. In addition, compared with STGIN for horizon twelve, *w/o dynamic spatial correlation module* respectively increased MAE, RMSE, and MAPE by 0.967%, 1.348%, and 9.867%; by 0.885%, 1.194%, and 9.112% for the next twelve time steps. Moreover, the performance of STGIN without a physical spatial module is superior to that without a dynamic spatial correlation module for both metrics (MAE and MAPE). The experimental results demonstrate that the dynamic and physical spatial correlation modules are positive for long-term highway traffic speed prediction, and the contribution of the dynamic spatial correlation module is higher than the physical module.

W/O Dynamic Temporal Correlation & **W/O Inherent Temporal Dependency** Compared with STGIN for horizon twelve, *w/o inherent temporal dependency module* increased MAE and RMSE by 0.419% and 0.486%; by 0.387% and 0.468% for the next twelve time steps. In addition, compared with STGIN for horizon twelve, *w/o dynamic temporal correlation module* respectively increased

Table 4
Computation cost during the training and inference phases.

Model	Parameters	Training/(100 iterations) (batch size = 32)		Inference (batch size = 1)	
		Time cost	GPU memory usage	Time cost	GPU memory usage
ARIMA[a]	–	268.155 (min)	–	8.285 (min)	–
SVR[a]	–	68.321 (min)	–	24.214 (min)	–
FI-RNNs	60,301	0.131 (min)	1019 MiB	1.774 (min)	507 MiB
LSTM_BILSTM	294,017	0.360 (min)	1523 MiB	2.633 (min)	515 MiB
PSPNN	158,508	0.293 (min)	4325 MiB	2.399 (min)	1965 MiB
MDL	285,377	0.276 (min)	5093 MiB	2.224 (min)	5037 MiB
T-GCN	37,844	0.037 (min)	1011 MiB	0.094 (min)	501 MiB
STGCN	193,921	0.612 (min)	1805 MiB	5.244 (min)	1625 MiB
DCRNN	372,353	0.972 (min)	2703 MiB	4.723 (min)	1615 MiB
AGCRN	750,660	0.166 (min)	2591 MiB	1.550 (min)	1871 MiB
ASTGCN	106,358	0.132 (min)	2085 MiB	0.247 (min)	1729 MiB
Graph-WaveNet	307,420	0.127 (min)	2141 MiB	0.240 (min)	1677 MiB
GMAN	907,201	0.821 (min)	8707 MiB	0.928 (min)	531 MiB
ST-GRAT	2,238,849	0.431 (min)	9515 MiB	5.187 (min)	1669 MiB
MTGNN	264,476	0.091 (min)	1853 MiB	0.216 (min)	1653 MiB
STGIN	208,257	0.182 (min)	3181 MiB	0.382 (min)	941 MiB

[a]Means the model train one time on the whole training set.

MAE, RMSE, and MAPE by 1.341%, 1.186%, and 0.043%; by 1.698% and 1.390% in terms of MAE and RMSE for the next twelve time steps. Moreover, the performance of STGIN without an inherent temporal dependency component is superior to that without a dynamic temporal correlation module regarding the three metrics. Furthermore, the performance of STGIN without a dynamic temporal correlation component is lower than dynamic or physical spatial correlation modules across all metrics. The experiments evaluate that the dynamic and inherent temporal correlation modules are essential for long-term highway traffic speed prediction, and the dynamic temporal correlation module contributes more than the inherent module and even outperforms the dynamic and physical spatial correlation modules.

W/O Fusion Gate Mechanism The performance without an adaptive fusion module is inferior to STGIN because the contributions of spatial and temporal features are deemed equally for long-term highway traffic speed prediction. For instance, compared with the STGIN for horizon twelve prediction, the MAE and RMSE increased by 1.675% and 1.575%, respectively; by 1.735% and 1.629% for the next twelve time steps. The experimental results indicate that the influence of spatial and temporal characteristics is distinguished for speed prediction, and the fusion gate mechanism is a preferable solution for adaptively combining distinct features.

W/O ST-Block An enhanced architecture of the STAtt Block proposed in GMAN (Zheng et al., 2020) is proposed; as a result, *w/o ST-Block* is inferior to STGIN regarding the MAE and RMSE metrics. For instance, *w/o ST-Block* increased MAE and RMSE by 0.910% and 0.925% compared with STGIN for horizon twelve prediction, respectively; by 1.397% and 1.237% for the next twelve time steps. These results suggest that STAtt Block's reliance on dynamic-spatial and -temporal correlations is insufficient and that physical spatial correlation and inherent temporal dependence have a crucial impact on spatio-temporal correlation extraction processing.

W/O Generative Inference Dynamic step-by-step decoding based on Transformer (Vaswani et al., 2017) is applied to replace the architecture of generative inference. Without generative inference, leading to poorer prediction precisions than STGIN. For example, for horizon six forecasting, compared with the STGIN model, the MAE and RMSE increased by 2.577% and 4.571%, respectively; by 2.208% and 4.315% for horizon twelve prediction. In addition, compared with all other variants, *w/o generative inference architecture* produces the worst forecast. The experimental results demonstrate that the generative inference strategy is an efficient method for preventing error propagation in long-term highway speed prediction, with a higher priority than other components.

However, except *w/o semantic enhancement module* and *w/o dynamic spatial correlation component*, variants outperform STGIN in terms of MAPE. This result may be due to the fact that these components are positive for predicting long-term highway speeds, but lack sensitivity at low speeds.

5.4.3. Computation cost

Table 2 presents the prediction performance comparisons, and the computation costs of the baselines and the proposed model on twelve target time steps are shown in Table 4,

Table 4 presents the computation cost of STGIN and baseline methods for predicting highway traffic speed, including total parameters, time cost, and GPU memory usage. According to Tables 2 and 4, the ARIMA and SVR possess high time costs but weak prediction performance in both the training and inference stages. T-GCN consumes less time and GPU memory than other deep learning baselines during the training and inference phases, but its performance is inferior. For the three optimal baselines shown in Table 2, AGCRN, Graph-WaveNet, and GMAN, achieving high prediction performance with Graph-WaveNet requires less time and fewer parameters, as shown in Table 4. Additionally, due to the difference in data loading, the GPU memory consumption of GMAN in the inference stage is less than that of AGCRN and Graph-WaveNet, and the forecasting accuracy is greater than that of other baselines. In this paper, we seek to devise a faster, more efficient, low-complexity model that uses less GPU memory and maintains high prediction accuracy. Therefore, STGIN is proposed as having superior performance, and its model complexity is less

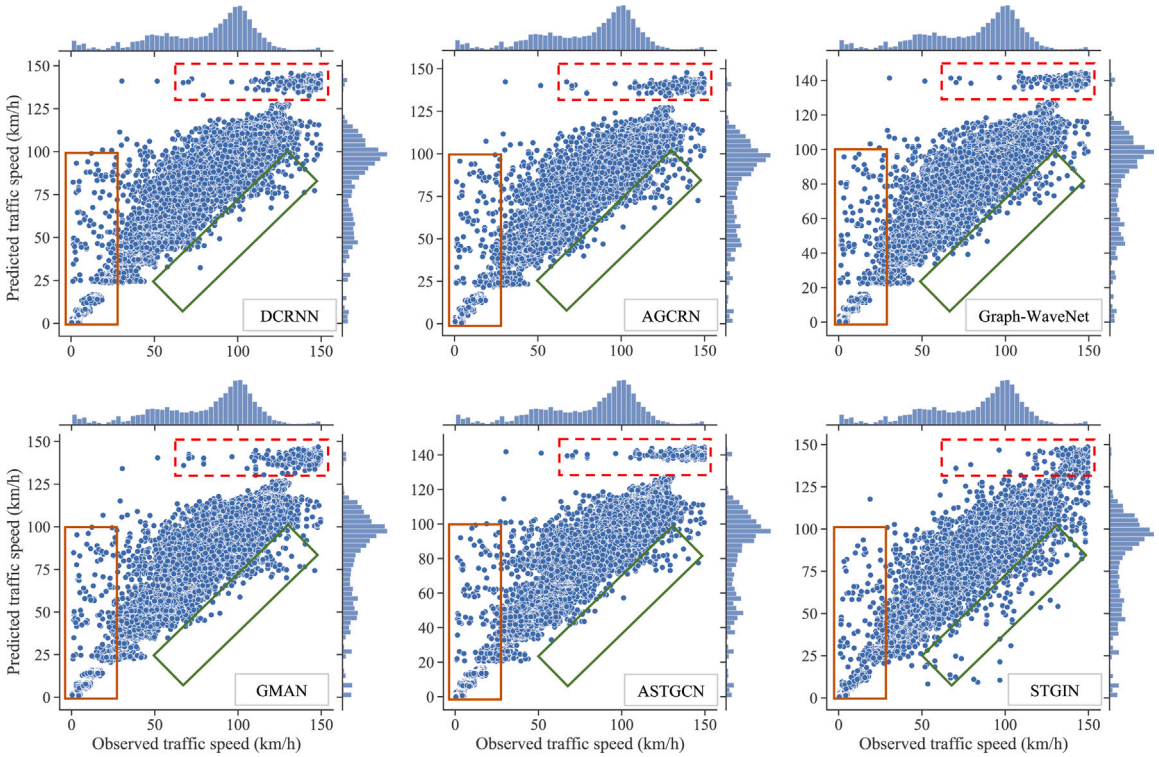


Fig. 12. Degree of fit between the observed and predicted traffic speed values. The blue dots indicate the degree of deviation between the observed and predicted values, and the blue histograms represent the distribution of observed and predicted.

than that of AGCRN, Graph-WaveNet, and GMAN. In the training phase, the time cost and GPU memory usage of STGIN are higher than those of AGCRN and Graph-WaveNet, whereas in the inference phase, the GPU memory usage is minimal, and the time cost outperforms AGCRN and GMAN.

We prefer a faster, more efficient, low-complexity model that uses less GPU memory while maintaining high prediction precision. STGIN provides long-term forecasts in a single pass, reducing the time required for inference compared to baselines such as DCRNN, GMAN, and ST-GRAT. The computation cost further validates the superiority of STGIN in long-term highway traffic speed prediction.

5.4.4. Fitting performance

According to the above comparisons, long-term highway traffic speed prediction is a challenging task. To better demonstrate the performance of STGIN, we compare it with the other five baseline models and visualize the fitting results. Six models, three optimal baselines based on GCNs, and other optimum methods based on GATs, including DCRNN, AGCRN, Graph-WaveNet, GMAN, ASTGCN, and STGIN. Fig. 12 shows the visualization results of the predictive fit ability over twelve target time steps, and we note the following four findings:

1. As is well-known, highway traffic velocities are affected by traffic incidents, which result in speed fluctuations, particularly a decrease in speed. When the traffic speeds are below 30 km/h, the blue dots are contained in the brown box for STGIN; however, some of these dots are out of the brown box in different degrees for other models. The performance of STGIN on low-traffic speed prediction may benefit from modeling the different scales of traffic patterns.
2. In addition, compared with baselines, the proposed STGIN model performs satisfactorily in the traffic speed range between 30 km/h and 130 km/h, and the distribution of observations is the same as predicted. In particular, all but a few blue dots are near the diagonal.
3. Moreover, when the traffic speeds exceed 100 km/h, STGIN presents a significant performance compared with baseline models. In the red box, for instance, the discrete degree of blue dots contained in STGIN is lower than in the other five models, and the distributions of observed and predicted keep consistent, as indicated by the blue histograms. The comparison results indicate that STGIN has excellent fitting performance at various scope traffic speeds and may have promising application potential.
4. Furthermore, a few blue dots deviate from the diagonal for the proposed model. For example, in the green box, when the traffic speeds exceed 50 km/h, the discrete degree of blue dots contained in STGIN is higher than in the other five models. This phenomenon is from the STGIN fit the dynamic traffic fluctuation, which caused model prediction shift in rare cases;

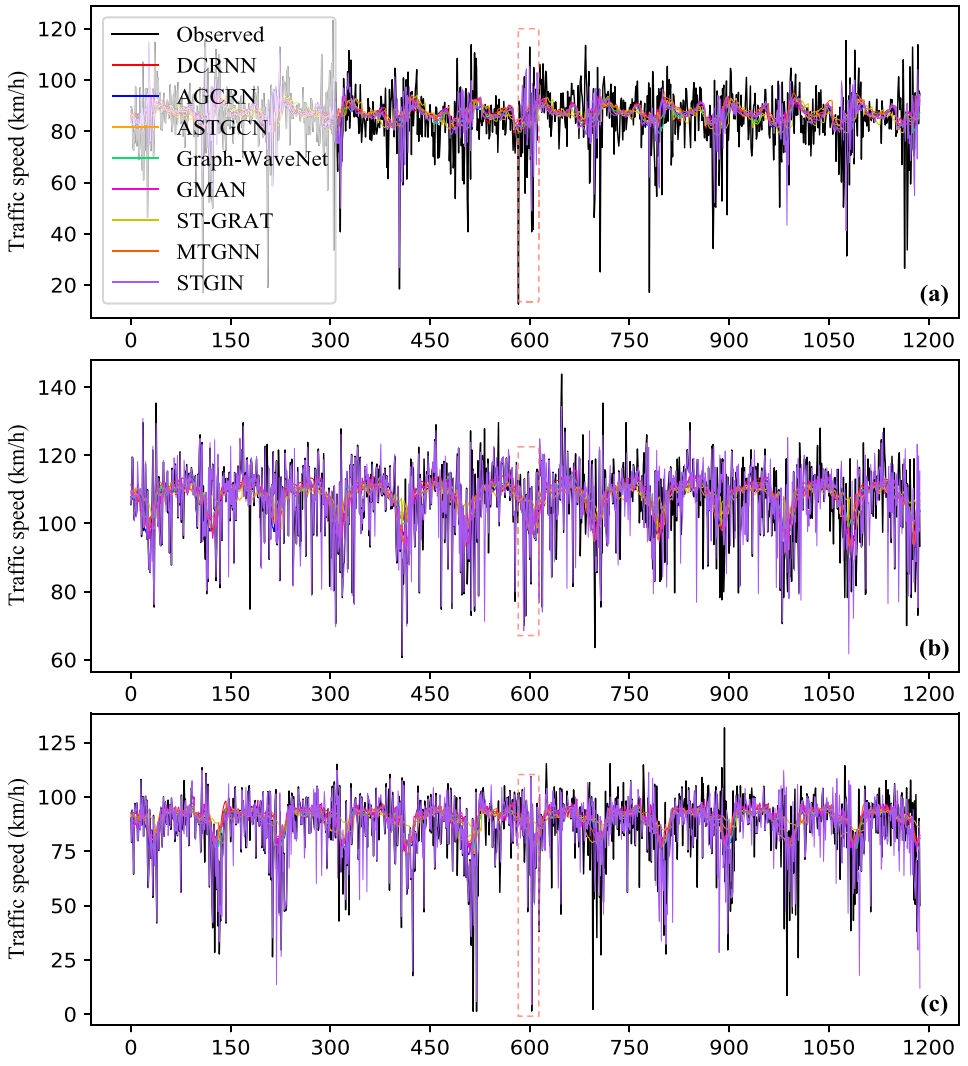


Fig. 13. The visualization results for predicting the next twelve time steps.

however, the baselines just keep prediction in range with traffic speeds (proved in the following Fig. 13), making it challenging to adapt to the complex real-world traffic condition. The practical applicability value of STGIN is confirmed once more.

5.4.5. Case study

We selected three road segments and visualized the prediction results for the next twelve time steps forecasting on the test set, as shown in Fig. 13. In the experiment, one hundred continuous samples are randomly sliced from the test set, and the samples' time interval is 2021.8.13 17:30 to 2021.8.26 05:30. Fig. 13 shows that STGIN accurately fit the changing trend of traffic speed and adapt to complex speed fluctuations. Because of the limited legend space, only seven baseline models based on GNNs are chosen and compared, DCRNN, AGCRN, ASTGCN, Graph-WaveNet, GAMN, ST-GRAT, and MTGNN. For example, traffic speeds vary greatly between different types of road segments; nonetheless, the performance of STGIN remains constant when compared to baseline models, including samples ranging from 30 to 50 (between 360 and 600 in Fig. 13). STGIN achieves consistently better results than other baselines, makes forecasts that are close to real data, and conquers speed oscillations.

To further illustrate the prediction performance of the proposed model and baselines on these three selected road segments, we visualize the historical, observed, and forecasted values (the samples correspond to the pink box in Fig. 13), as shown in Fig. 14. What can we learn from the graphic? First, the lack of apparent regularity in the changes in historical values, as depicted in Fig. 14(b), makes speed forecasting problematic. Second, it is difficult to identify a direct correlation between past and future observed values, such as in Fig. 14(c). Third, traffic speed fluctuation exists in the road segments, such as time steps sixteen and nineteen, as shown in Fig. 14(b). Finally, under the aforementioned three circumstances, our proposed model accurately predicts traffic speed, and

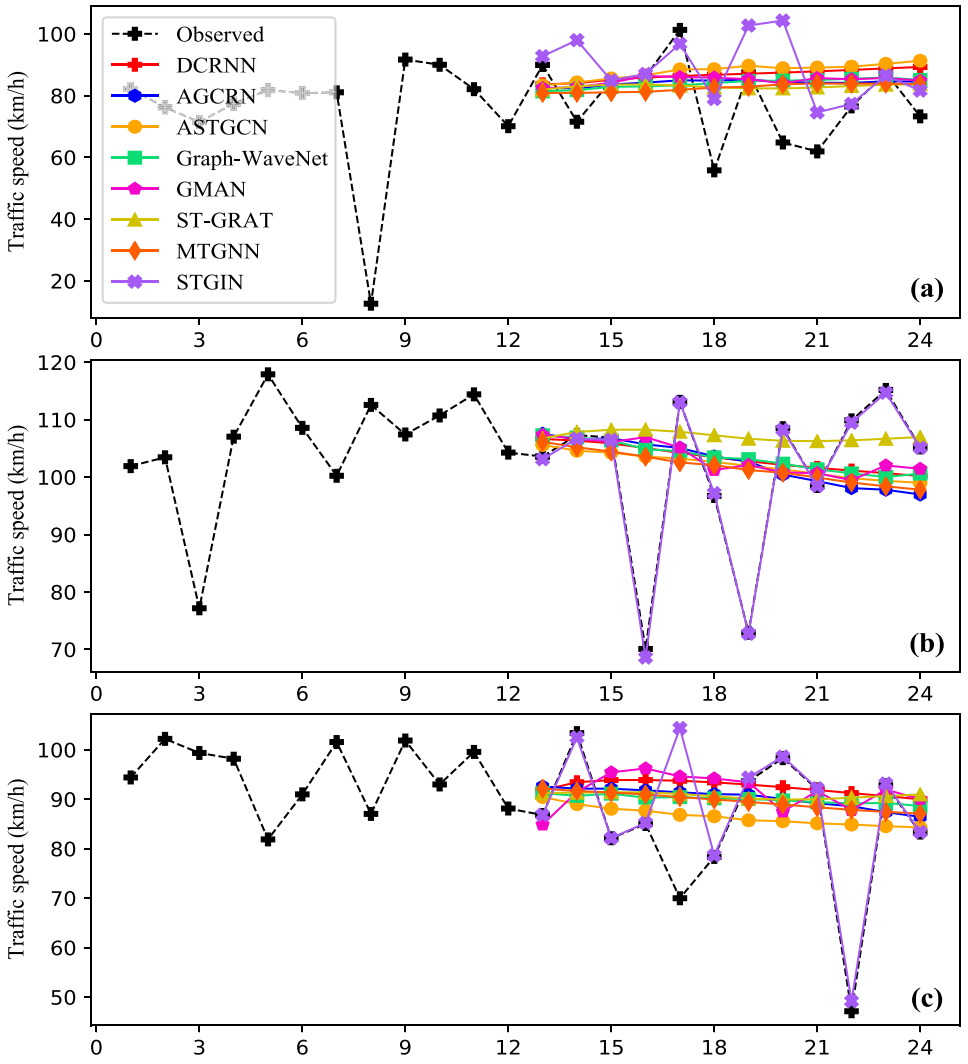


Fig. 14. Prediction vs. ground truth for segments (a), (b), and (c), respectively, 1–12 is historical time steps, and 13–24 is predicted horizons.

forecasting patterns are consistent with observations when compared to the other seven baseline techniques, as shown in Fig. 14(a), (b), and (c).

There are still obstacles to traffic speed prediction, but forecasting systems are also evolving. For instance, we can use a semantic enhancement module to smooth the speed fluctuation; observing the upstream and downstream traffic conditions, as supplementary information, can assist in determining the speed of the target road segment, particularly when it is challenging to find a direct association between previous and future observed values. Fortunately, we developed a novel approach, and the prediction performance of the suggested method has been verified.

6. Conclusion

This paper proposed a novel spatio-temporal generative inference network (STGIN), which can accurately predict the long-term highway traffic speed. Specially, we first proposed a semantic enhancement module to model the contextual semantics of traffic speed, improving the sensitivity and adaptability to speed fluctuations. We then designed a ST-Block to extract the spatio-temporal correlations of the highway network. We further proposed a generative inference to avoid long-term prediction error propagation in the spatial and temporal dimensions. Experiments on real-world dataset show that STGIN achieves state-of-the-art results compared with the baseline models, with a more evident advantage in long-term highway traffic speed prediction. In addition, comparing the prediction performance of all STGIN variants, each component exhibits a positive effect on prediction. Moreover, the computation cost further validates the superiority of STGIN in long-term highway traffic speed prediction. In future work, we expect to add traffic flow prediction tasks to the proposed STGIN model, gather related tasks to reach more robust predictions, and finally explore

advanced models for future predictions. In addition, we consider incorporating existing tricks such as Graph-WaveNet into our STGIN model, achieving a more accurate prediction performance.

CRediT authorship contribution statement

Guojian Zou: Data curation, Writing – original draft , Visualization, Investigation, Writing – reviewing & editing. **Ziliang Lai:** Conceptualization, Methodology. **Changxi Ma:** Conceptualization. **Ye Li:** Supervision. **Ting Wang:** Software, Validation, Writing – reviewing & editing.

Data availability

Data will be made available on request

Acknowledgments

This research were supported by the project of the National Key R&D Program of China (No. 2018YFB1601301), the National Natural Science Foundation of China (No. 71961137006).

References

Ahmed, M.S., Cook, A.R., 1979. Analysis of freeway traffic time-series data by using Box-Jenkins techniques. p. 722.
Bai, L., Yao, L., Li, C., Wang, X., Wang, C., 2020. Adaptive graph convolutional recurrent network for traffic forecasting. Adv. Neural Inf. Process. Syst. 33, 17804–17815.
Csikós, A., Viharos, Z.J., Kis, K.B., Tettamanti, T., Varga, I., 2015. Traffic speed prediction method for urban networks—an ANN approach. In: 2015 International Conference on Models and Technologies for Intelligent Transportation Systems (MT-ITS). IEEE, pp. 102–108.
Duan, P., Mao, G., Zhang, C., Wang, S., 2016. STARIMA-based traffic prediction with time-varying lags. In: 2016 IEEE 19th International Conference on Intelligent Transportation Systems. ITSC, IEEE, pp. 1610–1615.
Fang, W., Love, P.E., Ding, L., Xu, S., Kong, T., Li, H., 2021. Computer vision and deep learning to manage safety in construction: Matching images of unsafe behavior and semantic rules. IEEE Trans. Eng. Manage..
Fang, Y., Zhao, F., Qin, Y., Luo, H., Wang, C., 2022. Learning all dynamics: Traffic forecasting via locality-aware spatio-temporal joint transformer. IEEE Trans. Intell. Transp. Syst. 23 (12), 23433–23446.
Gu, Y., Lu, W., Qin, L., Li, M., Shao, Z., 2019. Short-term prediction of lane-level traffic speeds: A fusion deep learning model. Transp. Res. C 106, 1–16.
Guo, S., Lin, Y., Feng, N., Song, C., Wan, H., 2019. Attention based spatial-temporal graph convolutional networks for traffic flow forecasting. In: Proceedings of the AAAI Conference on Artificial Intelligence. vol. 33, pp. 922–929, (01).
He, K., Zhang, X., Ren, S., Sun, J., 2016. Deep residual learning for image recognition. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 770–778.
Hong, W.-C., 2011. Traffic flow forecasting by seasonal SVR with chaotic simulated annealing algorithm. Neurocomputing 74 (12–13), 2096–2107.
Ioffe, S., Szegedy, C., 2015. Batch normalization: Accelerating deep network training by reducing internal covariate shift. In: International Conference on Machine Learning. PMLR, pp. 448–456.
James, J., Markos, C., Zhang, S., 2021. Long-term urban traffic speed prediction with deep learning on graphs. IEEE Trans. Intell. Transp. Syst..
Jia, D., Chen, H., Zheng, Z., Watling, D., Connors, R., Gao, J., Li, Y., 2021. An enhanced predictive cruise control system design with data-driven traffic prediction. IEEE Trans. Intell. Transp. Syst..
Jia, Y., Wu, J., Du, Y., 2016. Traffic speed prediction using deep learning method. In: 2016 IEEE 19th International Conference on Intelligent Transportation Systems. ITSC, IEEE, pp. 1217–1222.
Jiang, B., Fei, Y., 2016. Vehicle speed prediction by two-level data driven models in vehicular networks. IEEE Trans. Intell. Transp. Syst. 18 (7), 1793–1801.
Jin, G., Cui, Y., Zeng, L., Tang, H., Feng, Y., Huang, J., 2020. Urban ride-hailing demand prediction with multiple spatio-temporal information fusion network. Transp. Res. C 117, 102665.
Jin, G., Li, F., Zhang, J., Wang, M., Huang, J., 2022. Automated dilated spatio-temporal synchronous graph modeling for traffic prediction. IEEE Trans. Intell. Transp. Syst..
Jin, G., Liang, Y., Fang, Y., Huang, J., Zhang, J., Zheng, Y., 2023. Spatio-temporal graph neural networks for predictive learning in urban computing: A survey. arXiv preprint arXiv:2303.14483.
Kenton, J.D.M.-W.C., Toutanova, L.K., 2019. BERT: Pre-training of deep bidirectional transformers for language understanding. In: Proceedings of NAACL-HLT. pp. 4171–4186.
Li, Y., Yu, R., Shahabi, C., Liu, Y., 2018. Diffusion convolutional recurrent neural network: Data-driven traffic forecasting. In: International Conference on Learning Representations.
Liu, Z., Li, Z., Wu, K., Li, M., 2018. Urban traffic prediction from mobility data using deep learning. IEEE Netw. 32 (4), 40–46.
Liu, L., Qiu, Z., Li, G., Wang, Q., Ouyang, W., Lin, L., 2019. Contextualized spatial–temporal network for taxi origin-destination demand prediction. IEEE Trans. Intell. Transp. Syst. 20 (10), 3875–3887.
Lu, Z., Lv, W., Xie, Z., Du, B., Xiong, G., Sun, L., Wang, H., 2022. Graph sequence neural network with an attention mechanism for traffic speed prediction. ACM Trans. Intell. Syst. Technol. 13 (2), 1–24.
Lu, W., Rui, Y., Ran, B., 2020. Lane-level traffic speed forecasting: A novel mixed deep learning model. IEEE Trans. Intell. Transp. Syst..
Luo, G., Zhang, H., Yuan, Q., Li, J., Wang, F.-Y., 2022. ESTNet: embedded spatial-temporal network for modeling traffic flow dynamics. IEEE Trans. Intell. Transp. Syst. 23 (10), 19201–19212.
Lv, Z., Xu, J., Zheng, K., Yin, H., Zhao, P., Zhou, X., 2018. Lc-rnn: A deep learning model for traffic speed prediction.. In: IJCAI. pp. 3470–3476.
Ma, C., Dai, G., Zhou, J., 2021. Short-term traffic flow prediction for urban road sections based on time series analysis and LSTM_BILSTM method. IEEE Trans. Intell. Transp. Syst..
Ma, X., Tao, Z., Wang, Y., Yu, H., Wang, Y., 2015. Long short-term memory neural network for traffic speed prediction using remote microwave sensor data. Transp. Res. C 54, 187–197.
Magazzino, C., Mele, M., 2021. On the relationship between transportation infrastructure and economic development in China. Res. Transp. Econ. 88, 100947.
Meng, X., Fu, H., Peng, L., Liu, G., Yu, Y., Wang, Z., Chen, E., 2020. D-LSTM: Short-term road traffic speed prediction model based on GPS positioning data. IEEE Trans. Intell. Transp. Syst..

Ong, B.T., Sugiura, K., Zettsu, K., 2016. Dynamically pre-trained deep recurrent neural networks using environmental monitoring data for predicting PM 2.5. Neural Comput. Appl. 27 (6), 1553–1566.
Otter, D.W., Medina, J.R., Kalita, J.K., 2020. A survey of the usages of deep learning for natural language processing. IEEE Trans. Neural Netw. Learn. Syst. 32 (2), 604–624.
Pan, Z., Zhang, W., Liang, Y., Zhang, W., Yu, Y., Zhang, J., Zheng, Y., 2020. Spatio-temporal meta learning for urban traffic prediction. IEEE Trans. Knowl. Data Eng..
Park, C., Lee, C., Bahng, H., Tae, Y., Jin, S., Kim, K., Ko, S., Choo, J., 2020. ST-GRAT: A novel spatio-temporal graph attention networks for accurately forecasting dynamically changing road speed. In: Proceedings of the 29th ACM International Conference on Information & Knowledge Management. pp. 1215–1224.
Qu, L., Lyu, J., Li, W., Ma, D., Fan, H., 2021. Features injected recurrent neural networks for short-term traffic speed prediction. Neurocomputing 451, 290–304.
Saif, H., He, Y., Fernandez, M., Alani, H., 2016. Contextual semantics for sentiment analysis of Twitter. Inf. Process. Manage. 52 (1), 5–19.
Shin, J., Sunwoo, M., 2018. Vehicle speed prediction using a Markov chain with speed constraints. IEEE Trans. Intell. Transp. Syst. 20 (9), 3201–3211.
Song, C., Lee, H., Kang, C., Lee, W., Kim, Y.B., Cha, S.W., 2017. Traffic speed prediction under weekday using convolutional neural networks concepts. In: 2017 IEEE Intelligent Vehicles Symposium. IV, IEEE, pp. 1293–1298.
Tang, J., Liu, F., Zou, Y., Zhang, W., Wang, Y., 2017. An improved fuzzy neural network for traffic speed prediction considering periodic characteristic. IEEE Trans. Intell. Transp. Syst. 18 (9), 2340–2350.
Vanajakshi, L., Rilett, L.R., 2004. A comparison of the performance of artificial neural networks and support vector machines for the prediction of traffic speed. In: IEEE Intelligent Vehicles Symposium, 2004. IEEE, pp. 194–199.
Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, Ł., Polosukhin, I., 2017. Attention is all you need. In: Advances in Neural Information Processing Systems. pp. 5998–6008.
Wang, J., Chen, R., He, Z., 2019. Traffic speed prediction for urban transportation network: A path based deep learning approach. Transp. Res. C 100, 372–385.
Wang, H., Liu, L., Dong, S., Qian, Z., Wei, H., 2016. A novel work zone short-term vehicle-type specific traffic speed prediction model through the hybrid EMD–ARIMA framework. Transp. B: Transp. Dyn. 4 (3), 159–186.
Wang, Y., Ren, Q., Li, J., 2023. Spatial–temporal multi-feature fusion network for long short-term traffic prediction. Expert Syst. Appl. 224, 119959.
Wu, Z., Pan, S., Long, G., Jiang, J., Chang, X., Zhang, C., 2020a. Connecting the dots: Multivariate time series forecasting with graph neural networks. In: Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. pp. 753–763.
Wu, Z., Pan, S., Long, G., Jiang, J., Zhang, C., 2019. Graph wavenet for deep spatial-temporal graph modeling. In: Proceedings of the 28th International Joint Conference on Artificial Intelligence. pp. 1907–1913.
Wu, W., Xia, Y., Jin, W., 2020b. Predicting bus passenger flow and prioritizing influential factors using multi-source data: Scaled stacking gradient boosting decision trees. IEEE Trans. Intell. Transp. Syst. 22 (4), 2510–2523.
Yang, H., Liu, C., Zhu, M., Ban, X., Wang, Y., 2021. How fast you will drive? Predicting speed of customized paths by deep neural network. IEEE Trans. Intell. Transp. Syst..
Yi, H., Bui, K.-H.N., 2020. An automated hyperparameter search-based deep learning model for highway traffic prediction. IEEE Trans. Intell. Transp. Syst..
Yu, B., Lee, Y., Sohn, K., 2020. Forecasting road traffic speeds by considering area-wide spatio-temporal dependencies based on a graph convolutional neural network (GCN). Transp. Res. C 114, 189–204.
Yu, B., Yin, H., Zhu, Z., 2018. Spatio-temporal graph convolutional networks: a deep learning framework for traffic forecasting. In: Proceedings of the 27th International Joint Conference on Artificial Intelligence. pp. 3634–3640.
Zafeiriou, S., Bronstein, M., Cohen, T., Vinyals, O., Song, L., Leskovec, J., Lio, P., Bruna, J., Gori, M., 2022. Guest editorial: Non-euclidean machine learning. IEEE Trans. Pattern Anal. Mach. Intell. 44 (2), 723–726.
Zang, D., Ling, J., Wei, Z., Tang, K., Cheng, J., 2018. Long-term traffic speed prediction based on multiscale spatio-temporal feature learning network. IEEE Trans. Intell. Transp. Syst. 20 (10), 3700–3709.
Zhang, W., Feng, Y., Lu, K., Song, Y., Wang, Y., 2020. Speed prediction based on a traffic factor state network model. IEEE Trans. Intell. Transp. Syst. 22 (5), 3112–3122.
Zhang, Z., Li, Y., Song, H., Dong, H., 2021a. Multiple dynamic graph based traffic speed prediction method. Neurocomputing 461, 109–117.
Zhang, B., Zou, G., Qin, D., Lu, Y., Jin, Y., Wang, H., 2021b. A novel encoder-decoder model based on read-first LSTM for air pollutant prediction. Sci. Total Environ. 765, 144507.
Zhao, J., Liu, Z., Sun, Q., Li, Q., Jia, X., Zhang, R., 2022. Attention-based dynamic spatial-temporal graph convolutional networks for traffic speed forecasting. Expert Syst. Appl. 204, 117511.
Zhao, L., Song, Y., Zhang, C., Liu, Y., Wang, P., Lin, T., Deng, M., Li, H., 2019. T-gcn: A temporal graph convolutional network for traffic prediction. IEEE Trans. Intell. Transp. Syst. 21 (9), 3848–3858.
Zheng, C., Fan, X., Wang, C., Qi, J., 2020. Gman: A graph multi-attention network for traffic prediction. In: Proceedings of the AAAI Conference on Artificial Intelligence. vol. 34, pp. 1234–1241, (01).
Zhou, Y., Li, J., Chi, J., Tang, W., Zheng, Y., 2022. Set-CNN: A text convolutional neural network based on semantic extension for short text classification. Knowl.-Based Syst. 257, 109948.
Zhou, L., Zhang, S., Yu, J., Chen, X., 2019. Spatial–temporal deep tensor neural networks for large-scale urban network speed prediction. IEEE Trans. Intell. Transp. Syst. 21 (9), 3718–3729.
Zou, G., Lai, Z., Ma, C., Tu, M., Fan, J., Li, Y., 2023. When will we arrive? A novel multi-task spatio-temporal attention network based on individual preference for estimating travel time. IEEE Trans. Intell. Transp. Syst. 1–15. http://dx.doi.org/10.1109/TITS.2023.3276916.