A comprehensive review of traffic flow prediction: from traditional models to deep learning architectures

Heng Zhang[1], Zhizhe Lin[1], Jinglin Zhou[2], Jiaxin Sun[3], Teng Zhou[1,4*] (iD) and Chunjie Cao[1]

[1] *School of Cyberspace Security, Hainan University, Haikou 570228, China*
[2] *School of Philosophy, Fudan University, Shanghai 200433, China*
[3] *Department of Economics, Ludwig Maximilian University of Munich, München 80539, Germany*
[4] *Yangtze Delta Region Institute, University of Electronic Science and Technology of China, Quzhou 324003, China*
* Corresponding author, E-mail: teng.zhou@hainanu.edu.cn

Abstract

Traffic congestion gradually becomes an increasingly serious problem in cities. Intelligent transportation systems (ITS) have been introduced to alleviate traffic congestion. Traffic flow prediction, as an important part of ITS, has become a research hotspot in recent years. This study presents machine learning techniques in traffic flow prediction. First, it systematically introduces the task of traffic flow prediction, including the research background and problem definition. Then, it reviews twelve benchmark datasets commonly used for traffic flow prediction and provides a detailed description of the data processing steps. After that, this study highlights the theoretical foundations and significant impact of Graph Convolutional Networks for traffic flow modeling. Finally, this study provides an in-depth analysis of the opportunities and challenges in this field and offers pertinent suggestions for future research. Overall, this study can help researchers quickly get started with traffic flow prediction.

Keywords: Intelligent transportation systems; Traffic flow modeling; Comprehensive review; Traffic flow prediction; Traffic flow theory

Introduction

With the rapid advancement of technology, urbanization has accelerated, resulting in a significant increase in the urban population and placing substantial pressure on urban traffic management[1,2]. The development of a city cannot be separated from efficient and intelligent traffic management. For this purpose, Intelligent transportation systems (ITS) were developed and have been widely adopted around the world[3,4]. The need for short-term traffic flow prediction has also become greater and more urgent with the rapid technological advancement and popular application of intelligent transportation systems[5].

Traffic flow prediction is a complex task. It aims to predict the future traffic state based on the input historical traffic state. Accurate traffic flow forecasting is essential for many intelligent transportation systems[6]. It helps traffic management implement interventions to alleviate traffic congestion[7], such as signalization control and manual road direction. Timely and reliable traffic flow information plays a remarkable role in alleviating traffic congestion, improving traffic operation efficiency, and supporting travel decision-making[8].

In addition to providing assistance in traffic management, traffic flow prediction also contributes to infrastructure development and resource allocation in cities[9]. The government can adjust urban development plans based on traffic flow during specific periods. The development of new roads and residential areas can optimize the utilization of road resources and help address the imbalance between the supply and demand of road capacity and travel demand.

Generally, the prevalent statistical approaches for traffic flow forecasting can be divided into parametric and non-parametric methods[10]. Traditional traffic flow prediction methods use parametric models to fit historical data and output predictions. However,

forecasting real-time and accurate traffic flow is highly challenging because of the complex spatiotemporal dependencies inherent in traffic flow[11,12]. Parametric models underperform in prediction due to the complex spatial and temporal correlations of traffic flow. Early parametric models, such as Historical Average (HA)[13], Autoregressive Integrated Moving Average (ARIMA)[14], and Vector Autoregression (VAR)[15] provided the foundation for traffic prediction.

In recent years, with the rise of machine learning, many traffic flow prediction models have been developed based on machine learning. Machine learning models are capable of capturing complex spatiotemporal dependencies from massive traffic flow data to provide more accurate predictions. Additionally, data-driven methods do not require prior knowledge of the underlying system dynamics or equations, making them flexible and adaptable to different datasets and scenarios[16].

Classical machine learning models, such as linear regression, decision trees, and support vector machines, have been widely used for traffic flow prediction. However, traffic flow sequences are easily affected by external factors, such as unexpected accidents or manual interventions in traffic control[17], and classical machine learning models cannot capture complex spatiotemporal dependencies[18].

Deep learning approaches are not limited by the stationarity assumption and achieve better performance in time series prediction[19,20]. Some modern deep neural networks, such as Recurrent Neural Networks (RNN)[21−23], Long Short-Term Memory Networks (LSTM)[24−26], and Convolutional Neural Networks (CNN)[27−29], have achieved outstanding results in traffic flow prediction[30]. These deep neural networks can extract complex spatiotemporal features from traffic flow data. They capture both temporal and spatial dependencies of traffic flow and output more accurate predictions. However, approaches based on CNNs and RNNs can only handle

standard gridded data and neglect non-Euclidean correlations generated by complex road networks[31].

Graph Neural Networks (GNNs) are a recent research hotspot. Graphs can represent many real-life objects, such as social relationships and knowledge graphs, among others. The traffic network is also a 'natural graph'[32,33]. In deep learning, graphs can be used to represent complex road network relationships, thus capturing the spatial dependencies of traffic flow. GCNs, when integrated with RNNs or Gated CNNs, achieve better performance in spatial and temporal modeling[34].

Despite the significant progress of deep learning in traffic flow prediction, there are still some challenges. For example, challenges remain in terms of model efficiency, transferability, abnormal event handling, and interpretability. This also brings many opportunities for future research.

Traffic congestion occurs all over the world, and traffic flow prediction is needed in many regions. However, these regions have different levels of economic development. A deep learning network requires a large number of parameters and long training times[35]. Some regions may not be able to provide enough computational power to train the model. Therefore, it makes sense to develop lightweight but effective traffic flow prediction models. Such models can be extended to economically underdeveloped regions to provide effective traffic flow prediction services, thus improving local traffic conditions.

In addition, due to regional economic conditions, certain regions may not be able to provide sufficient infrastructure to capture traffic flow data. Therefore, traffic flow data is lacking in these regions. The deficiency and low-quality training data may cause these models to fall into local minima[36]. Meanwhile, the traffic conditions in each region are influenced by local customs, culture, and road construction, and a traffic flow prediction model trained in a different region cannot be well adapted to the local traffic conditions. Therefore, it is necessary to develop an effective pre-training model and a domain-adaptive fine-tuning strategy.

While existing deep learning models show good prediction performance, their parameters still lack interpretability. As an auxiliary system for traffic management authorities to make decisions, traffic flow prediction models need to be sufficiently interpretable to make their predictions more trustworthy[37].

It is evident from the literature that there have been various surveys on traffic flow prediction. While most of these articles have investigated the role of machine learning (ML) in predicting traffic flow, the focus of this study differs from that of existing surveys.

● More concern is placed on open-source traffic flow prediction models that can be reproduced. Meanwhile, 15 publicly available datasets are reviewed, and details on how they can be used are provided, so that beginners can get started quickly.

● Although GNNs are used by most traffic flow prediction models, this review goes beyond GNNs to include new models from recent years. Examples include the Transformer model, pre-training models, and differential equation models.

● The future directions proposed are not limited to designing more complex models, but are more oriented towards real-world problems—for example, large-scale deployment, generalization capability of models, and multi-factor combination capability.

Background

Here, the traffic flow prediction task is first described. The traffic flow prediction task aims to forecast future traffic conditions based on historical traffic flow data, taking into account a variety of external factors, such as weather and time of day[38]. In general, three common metrics for traffic flow prediction are traffic flow, traffic speed, and traffic density[33]. The choice of metric depends on the dataset employed.

Problem definition

Traffic flow rate is the total number of vehicles passing a given point during a specific time period[39]. It measures how many vehicles pass through a given point on the roadway and is critical to understanding the overall usage of the roadway network and identifying potential congestion points. Traffic speed is the average speed at which vehicles are detected traveling at a target location during the same time period[40]. Traffic speed indicates road conditions: high speed implies smooth traffic, while low speed implies congestion. Traffic density is the number of vehicles per unit length at a given time[41]. Higher traffic densities generally result in slower speeds and a higher likelihood of traffic jams, while lower densities indicate smoother traffic flow.

Traffic flow, traffic speed, and traffic density are all important metrics for characterizing road network performance[42]. In practice, it is common to use one of the three metrics for forecasting purposes due to data availability. For example, some datasets may provide detailed speed and flow information but lack density data, while others may provide all three metrics. The choice of metric usually depends on the specific requirements of the prediction task and data availability.

It is important to note that many data mining methods have been proposed due to increased data availability. Traffic flow prediction uses not only the metrics to be predicted, but also additional data. The most common type of additional data is periodic embedding[43]. Periodic embeddings are designed by extracting timestamps from traffic flow datasets. Common periodic embeddings include 'time of day', which represents the relative position of the current time within a day, and 'day of week', which represents the relative position of the current day within a week[44]. Periodic embedding allows the traffic flow prediction model to learn the periodicity of traffic flow. In addition, data such as weather and holidays[45] can also be incorporated into the traffic flow prediction model as embeddings.

Traffic flow prediction, as a key factor for measuring road traffic, can be grouped into coarse-grained and fine-grained types, which differ essentially in their applications to traffic planning and strategic decision-making. The main content of fine-grained traffic flow prediction is based on current and past traffic flow data, using time series analysis methods to predict future traffic flow over short-term horizons such as seconds, minutes, or half an hour[46].

Formally, the traffic prediction problem can be stated as follows: given a series of historical traffic data points, the objective is to predict traffic conditions at one or more future time steps[47]. This involves creating a predictive model that can accurately capture the temporal dependencies and patterns in the traffic data. The input to this model typically includes a sequence of traffic observations over a specified period, and the output is the forecasted traffic condition at a future time step. Thus, the traffic flow prediction task can be defined as follows:

Given the current time t, the prediction and history input windows is defined as follows: the prediction window of T' time steps is [t+1, t+2, ..., t+T'] and the history input window of T time steps time steps is [t−T+1, t−T+2, ..., t]. The ground truth of traffic flow is denoted as X, partitioned into historical data X = [$X_{t−T+1}$, $X_{t−T+2}$, ..., X_t] and labels Y = [Y_{t+1}, Y_{t+2}, ..., $Y_{t+T'}$]. The predicted output is denoted as $\hat{Y} = [\hat{Y}_{t+1}, \hat{Y}_{t+2}, \cdots, \hat{Y}_{t+T'}]$. Assuming that the machine learning model is a function f, the traffic flow prediction task can be defined as Eq. (1):

$$\hat{Y} = f([X_{t-T+1}, X_{t-T+2}, \ldots, X_t]) \qquad (1)$$

The goal of the traffic flow prediction task is to design a model that takes historical data X as input and produces a set of predictions $\hat{Y}$ close to the labels Y. Machine learning uses a loss function L to measure the difference between Y and $\hat{Y}$. Formally, the goal of machine learning for traffic flow prediction is to find a set of trainable parameters θ^* that minimizes the loss function L. The detailed definition is as Eq. (2):

$$\theta^* = \arg\min_{\theta^*} L\left(y_{t+T'}, \hat{y}_{t+T'}; \theta^*\right) \tag{2}$$

Performance metrics

After model training is complete, three metrics are typically used to evaluate predictive performance. They are Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and Mean Absolute Percentage Error (MAPE)[48]. The MAE is usually used as the model's loss function. The formulas of the three metrics are as Eqs (3)−(5):

$$MAE = \frac{1}{n} \sum_{i=1}^{n} |y_i - \hat{y}_i| \tag{3}$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (y_i - \hat{y}_i)^2} \tag{4}$$

$$MAPE = \frac{1}{n} \sum_{i=1}^{n} \left|\frac{\hat{y}_i - y_i}{y_i}\right| \times 100\% \tag{5}$$

Model list

Table 1 shows a list of machine learning models for traffic flow prediction collected in this review. Specific details of these models, as published in the papers, are discussed in section Machine Learning for Traffic Flow Prediction, including research gaps and innovations.

Datasets and preprocessing

Datasets

In this section, the 15 datasets used in the collected papers are summarized. For each dataset, the following information is provided: metric types, time range, number of sensors, number of road edges, sample rate, and zero value rate. In addition to conventional loop detector data, recent studies[54,66] have increasingly utilized grid-based datasets, commonly used in trajectory prediction tasks. These datasets are commonly employed as baselines in experiments. An overview of the datasets is given in Table 2.

Table 1. Model list.

Model	Tag	Source	Year
DCRNN[49]	RNN/GNN	ICLR 2018	2018
STGCN[50]	CNN/GCN	IJCAI 2018	2018
GraphWaveNet[51]	TCN/GCN	IJCAI 2019	2019
STSGCN[52]	GCN	AAAI 2020	2020
STAWnet[53]	TCN	IET intelligent transport systems	2021
ST-SSL[54]	GCN	AAAI 2023	2023
MegaCRN[55]	RNN/GCN	AAAI 2023	2023
D2STGNN[56]	CNN/GCN	VLDB 2022	2022
DDGCRN[57]	GCN	Pattern recognition	2023
RGDAN[58]	GAT	Neural networks	2024
STWave[59]	GAT	ICDE 2023	2023
STAEformer[60]	Transformer	CIKM 2023	2023
PDFormer[61]	Transformer	AAAI 2023	2023
STEP[62]	Pre-training	SIGKDD 2022	2022
STDMAE[63]	Pre-training	IJCAI 2024	2024
STG-NCDE[64]	Differential equations	AAAI 2022	2022
STG-NRDE[65]	Differential equations	ACM transactions on intelligent systems and technology	2023

Data preprocessing

Most machine learning models are implemented in the TensorFlow[2] or PyTorch[67] frameworks. The format of the dataset is usually NPZ (Numpy format) or CSV (Comma-Separated Values) and can be read using Pandas or Numpy. In general, the common dataset's shape is (TimeStep, Node). The first dimension corresponds to time steps, and the second dimension corresponds to nodes. Some datasets may contain multiple channels, such as flow, speed, and density. The shape of such a dataset typically has shape (TimeStep, Node, Channel).

To generate the inputs and labels for the traffic flow prediction model, a pointer t is first initialized. Depending on the history window T and prediction window T', the pointer t traverses the first dimension (time steps) between $[T, End-T']$. At each time step, the traffic flow data of $[t-T+1, t]$ is used as input, and the traffic flow data of $[t+1, t+T']$ is used as the corresponding label. After generating all the input-label pairs, the dataset is generally split into training, validation, and test sets in a 6:2:2 ratio. The dataloader fetches a batch of data pairs at a time according to the configuration, and the training set is shuffled during fetching. Finally, the dataset is fed into the model for training. The shape of a batch of data is generally $[B, N, T, C]$, where B is batch size, N is the number of nodes, T is the length of the time series, and C is the number of channels. Different models may have different processing formats and split ratios, depending on the specific model implementation. The specific data preprocessing process is shown in Fig. 1.

Machine learning for traffic flow prediction

Classic machine learning model

Dougherty et al.[68] was one of the first efforts to use neural networks for traffic flow prediction. Vlahogianni et al.[69] used genetic algorithms (GA) to tune the neural networks. Zheng et al.[70] combined conditional probability and Bayesian estimation and use multiple neural network predictors. Chan et al.[71] employed the hybrid exponential smoothing method and the Levenberg-Marquardt (LM) algorithm to process traffic flow data, aiming to improve the generalization capabilities of the model. This provided opportunities for migratory learning of models. Davis & Nihan[72] proposed a k-nearest neighbor (K-NN) approach to predict traffic flow. Cai et al.[73] updated the original K-NN model for traffic flow prediction which uses a spatiotemporal state matrix instead of a time series in K-NN.

Moreover, some models based on SVR, Support Vector Regression, are also used for traffic flow prediction. Castro-Neto et al.[74] proposed an online support vector machine for regression (OL-SVR). Su et al.[75] proposed an incremental support vector regression (ISVR) which uses an incremental learning approach to update the prediction function in real time.

These classic machine-learning models exhibited stronger performance than old statistical models over the past two decades. However, these models usually use simple structures, so they cannot be trained efficiently.

Deep learning model

About a decade ago, with the fast development of computational power, deep neural networks became more popular and have been applied in many applications[76]. Deep learning learns complicated features extracted from traffic flow information. Deep learning networks have many different network structures. For instance, RNNs can effectively capture the temporal dimension of traffic information, while CNNs can capture the spatial dimension[77]. Compared with classical machine learning models, deep neural network

Table 2. Datasets overview.

Dataset	Nodes	Edges	Time range	Sample rate	Zero value rate	Type
METR-LA	207	1,515	2012/03/01−2012/06/30	5 min	7.581%	Speed
PEMS-BAY	325	2,369	2017/01/01−2017/05/31	5 min	4.906%	Speed
PeMSD7(M)	228	1,664	2012/05/01−2012/06/30	5 min	0.10%	Speed
PeMSD7(L)	1,026	14,534	2012/05/01−2016/06/30	5 min	0.50%	Speed
PEMS03	358	546	2018/09/01−2018/11/30	5 min	5.102%	Flow
PEMS04	307	338	2018/01/01−2018/02/28	5 min	5.730%	Flow
PEMS07	883	865	2017/01/05−2017/08/06	5 min	5.028%	Flow
PEMS08	170	276	2016/07/01−2016/08/31	5 min	5.070%	Flow
CA	8,600	201,363	2017/01/01−2021/12/31	15 min	5.869%	Flow
GLA	3,834	98,703	2017/01/01−2021/12/31	15 min	5.712%	Flow
GBA	2,352	61,246	2017/01/01−2021/12/31	15 min	5.719%	Flow
SD	716	17,319	2017/01/01−2021/12/31	15 min	5.845%	Flow
NYC-taxi	16*12	−	2016/01/01−2016/02/29	30 min	48.63%	Inflow/outflow
NYC-bike	14*8	−	2016/08/01−2016/09/29	30 min	52.55%	Inflow/outflow
SZ-taxi	156	532	2015/01/01−2015/01/31	15 min	26.57%	Speed
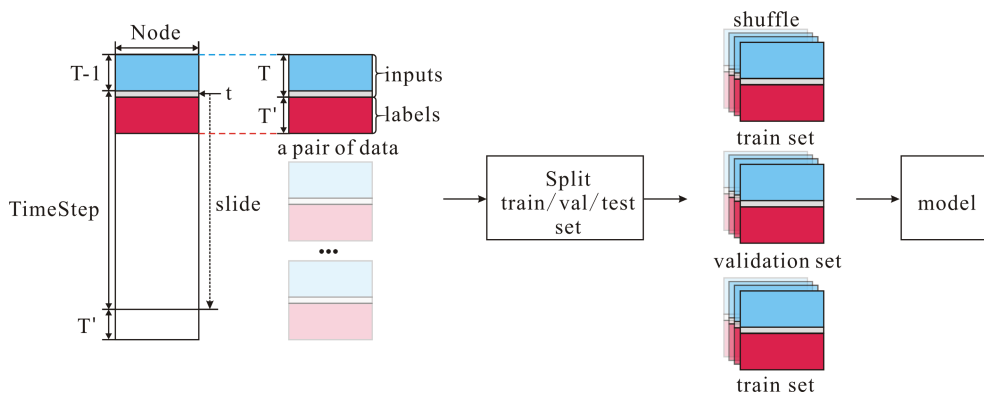
Fig. 1 Data preprocessing steps.

models have several advantages, but they also have some disadvantages. These deep learning models are generalizable. However, in traffic flow prediction, due to the strong spatiotemporal dependencies, specific models need to be designed to capture these dependencies for better prediction. Before the emergence of GNNs, researchers usually transformed urban road networks into Euclidean adjacency matrices and modeled them by combining convolution operations and residual units, which improved the ability of spatiotemporal feature extraction[78].

Graph convolutional network

In practice, the sensor nodes in the road are not distributed in a grid pattern, but irregularly distributed in the road network[79]. Recently, many researchers have begun to focus on graph neural networks[80]. The graph structure can well represent the various relationships in the road network[81]. Therefore, the model's forward propagation better reflects the actual structure of road networks. Graph neural networks overcome the shortcomings of previous studies that ignored the connectivity and globality of the network.

The basic process of graph neural networks is as follows: Consider a graph $G = (V, E)$ with N nodes, with V, representing the nodes, and E, representing the edges. There is an adjacency matrix $A \in \mathbb{R}^{N \times N}$ to represent edges. The traffic flow input for a time step is denoted as $X \in \mathbb{R}^{N \times 1}$. To incorporate the node features themselves when multiplying A and X, it is necessary to add an identity matrix (self-loop) to the adjacency matrix A. Let $\tilde{A} = A + I$. Next, the degree of each node needs to be considered. The contribution of each connected node is inversely proportional to the node's degree. Let $L = D^{-\frac{1}{2}} \tilde{A} D^{-\frac{1}{2}}$, where D is the degree matrix. D is a diagonal matrix recording the total number of outgoing and incoming degrees of $\tilde{A}$.

$D^{-\frac{1}{2}}$ denotes the matrix raised to the power of $-\frac{1}{2}$. $D^{-\frac{1}{2}} \tilde{A} D^{-\frac{1}{2}}$ normalizes the adjacency matrix by the square root of node degrees along each row and column. The final computation yields the normalized Laplacian matrix L.

In forward propagation, the Laplacian matrix L is multiplied by the hidden states and input X. The weight matrix W is applied to the concatenated input X and hidden state H after the graph convolution. The graph convolution is simplified by the formula $X = L[X\|H]W$, where $\|$ denotes concatenation. The complete graph convolution operation is given by the following formula:

$$Y = [D^{-\frac{1}{2}} \tilde{A} D^{-\frac{1}{2}}][X \| H] W \qquad (6)$$

Figure 2 shows the process of graph convolution with a 4×4 adjacency matrix and 4×1 feature vector. Here, the symbol @ represents matrix multiplication.

DCRNN[49], the Diffusion Convolutional Recurrent Neural Network, was one of the first efforts to use graph convolutional networks for traffic prediction. DCRNN uses a diffusion process to model the dynamics of traffic flow and proposes diffusion convolution to capture spatial dependencies. The model leverages recurrent neural networks (RNNs) to capture temporal dependencies. In particular, the matrix multiplication in the Gated Recurrent Unit (GRU) is replaced by graph convolution to capture spatial correlations (Fig. 3).

STGCN[50], the Spatiotemporal Graph Convolutional Network, consists of multiple spatiotemporal convolutional blocks, which alternate between gated sequence convolutional layers and spatial graph convolutional layers. The model employs the graph convolution operator '*G', which is based on spectral graph convolution, filtering the graph signal by the product of the signal and the kernel. The STGCN also employs a gated CNN to extract temporal features,
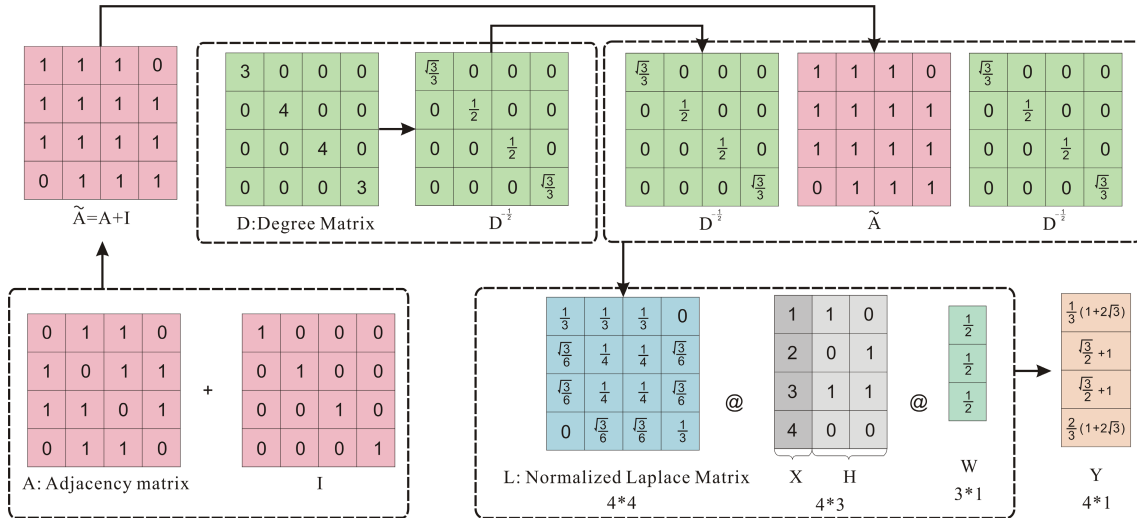
Fig. 2 Graph convolution.

implemented via 1-D causal convolution and gated linear units (GLUs) (Fig. 4).

Graph WaveNet[51] learns from node embeddings and proposes a novel adaptive dependency matrix. Such a design can accurately capture hidden spatial dependencies in the data. This design addresses the issue that explicit graph structures do not necessarily reflect the true dependencies and may fail to capture actual relationships (Fig. 5).

STSGCN[52] effectively captures complex local correlations in spatiotemporal network data through a well-designed spatiotemporal synchronization graph convolution mechanism, and includes a multi-module layer to account for data heterogeneity, improving prediction accuracy. The model captures local spatiotemporal correlations by constructing a local spatiotemporal graph using the spatiotemporal simultaneous graph convolution module (STSGCM), and models the heterogeneity across different time periods using a
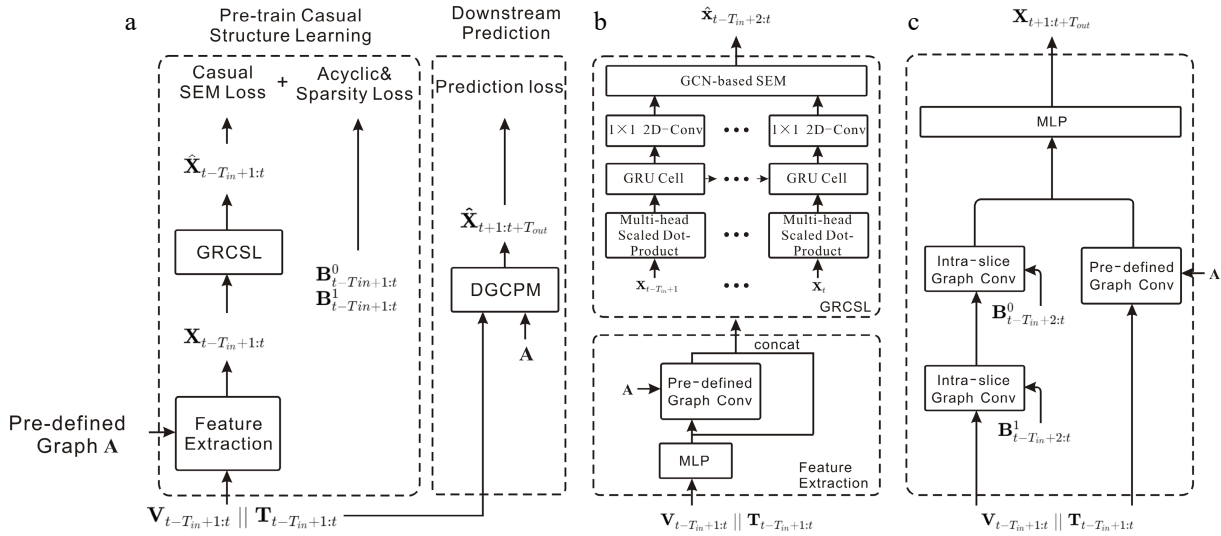
Fig. 3 DCRNN: diffusion convolution for spatial dependencies, GRU for temporal dependencies.
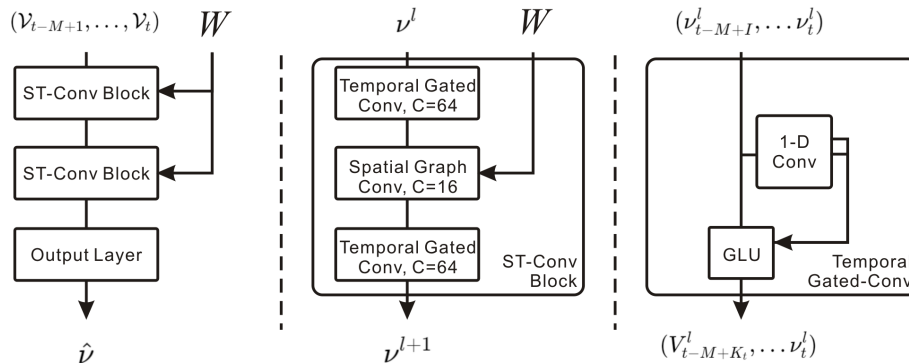
Fig. 4 STGCN: combines spectral graph convolution for spatial features and gated causal CNN for temporal features.

spatiotemporal simultaneous graph convolution layer (STSGCL) composed of multiple STSGCMs (Fig. 6).

STAWnet[53] effectively captures the spatiotemporal dependencies in traffic conditions through a combination of temporal convolutions and attention mechanisms. STAWnet does not require a priori knowledge of the graph structure, instead capturing hidden spatial relationships through self-learned node embeddings. This design improves the flexibility of the model and allows it to be easily extended to other spatiotemporal forecasting tasks. STAWnet also incorporates a dynamic attention mechanism, which adjusts the weights of different nodes according to varying traffic conditions and spatial information (Fig. 7).

ST-SSL[54] integrates temporal and spatial convolutions to effectively encode spatiotemporal traffic patterns and incorporates two self-supervised learning tasks to enhance the main traffic prediction task by capturing spatial and temporal heterogeneity. The innovation of the ST-SSL framework lies in its adaptive augmentation of traffic flow graph data at both the attribute and structure levels. It also introduces a soft clustering paradigm to capture diverse spatial patterns among regions and a temporal self-supervised learning paradigm to maintain dedicated representations of temporal traffic dynamics. These mechanisms enable ST-SSL to effectively overcome the limitations of existing methods in handling spatial and temporal heterogeneity, particularly in scenarios with skewed regional traffic distributions and time-varying traffic patterns (Fig. 8).

MegaCRN[55] integrates a meta-graph learner into a Graph Convolutional Recurrent Network (GCRN). This meta-graph learner is designed to dynamically generate node embeddings from an underlying meta-node library. Each node embedding is a multifaceted representation that encapsulates the temporal and spatial nuances of traffic dynamics. The meta-node library is a repository of memory items, each represented as a vector capturing the features of a typical traffic pattern. The Mega-Graph Learner queries this library to identify the memory item or prototype most similar to the current traffic state represented by the GCRN hidden state (Fig. 9).

D2STGNN[56] addresses the limitations of conventional traffic prediction models by decoupling diffuse and intrinsic signals in traffic data. D2STGNN employs a decoupled spatiotemporal framework (DSTF) that integrates a dynamic graph learning module along with separate models for diffuse and intrinsic signals, thereby enhancing the modeling of spatiotemporal correlations (Fig. 10).

DDGCRN[57] dynamically generates graphs that capture the spatiotemporal dynamics of traffic flows and distinguishes between normal and abnormal traffic signals for a more nuanced understanding of traffic patterns. It adopts a data-driven approach to generate dynamic graphs from time-varying traffic signals, which are subsequently processed by the Dynamic Graph Convolutional Recursive Module (DGCRM) to extract salient spatiotemporal features. Furthermore, the model employs a segmented learning strategy that enhances training efficiency and reduces computational resource consumption during the initial training phase (Fig. 11).

RGDAN[58] integrates the Graph Diffusion Attention module and the Temporal Attention module. It employs a stochastically initialized Graph Attention Network (GAT) that does not rely on predefined node interactions, instead learning the attention weights directly from the data. This stochastic GAT, combined with an adaptive matrix, captures both local and global spatial dependencies, thereby providing a more accurate representation of traffic flow patterns. The Temporal Attention module is adept at identifying and learning from temporal patterns, ensuring that the predictive model is sensitive to temporal variations inherent in traffic flow (Fig. 12).
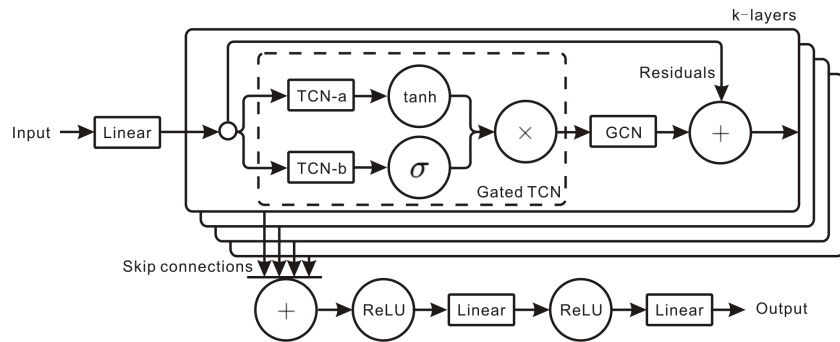

Fig. 5 Graph WaveNet: adaptive adjacency captures spatial relations, while dilated causal convolution models temporal dynamics.
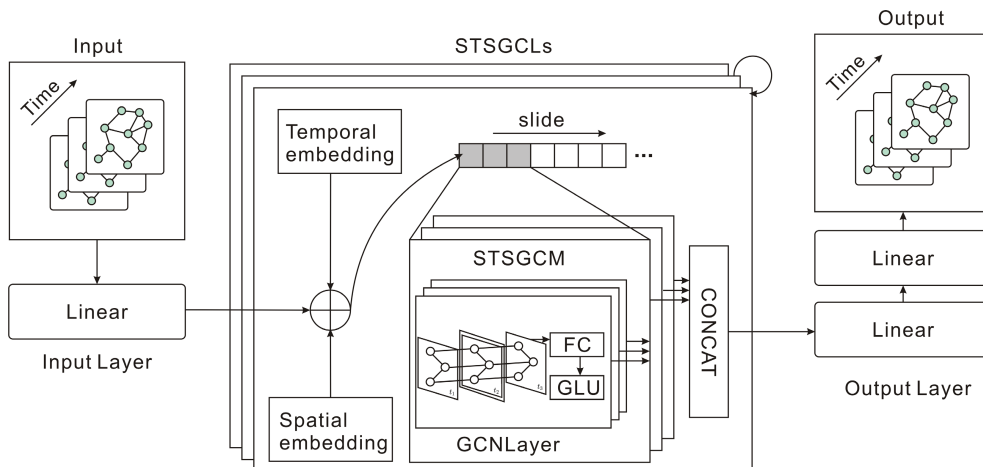

Fig. 6 STSGCM: captures local spatiotemporal correlations via a spatial-temporal synchronous graph convolution module.
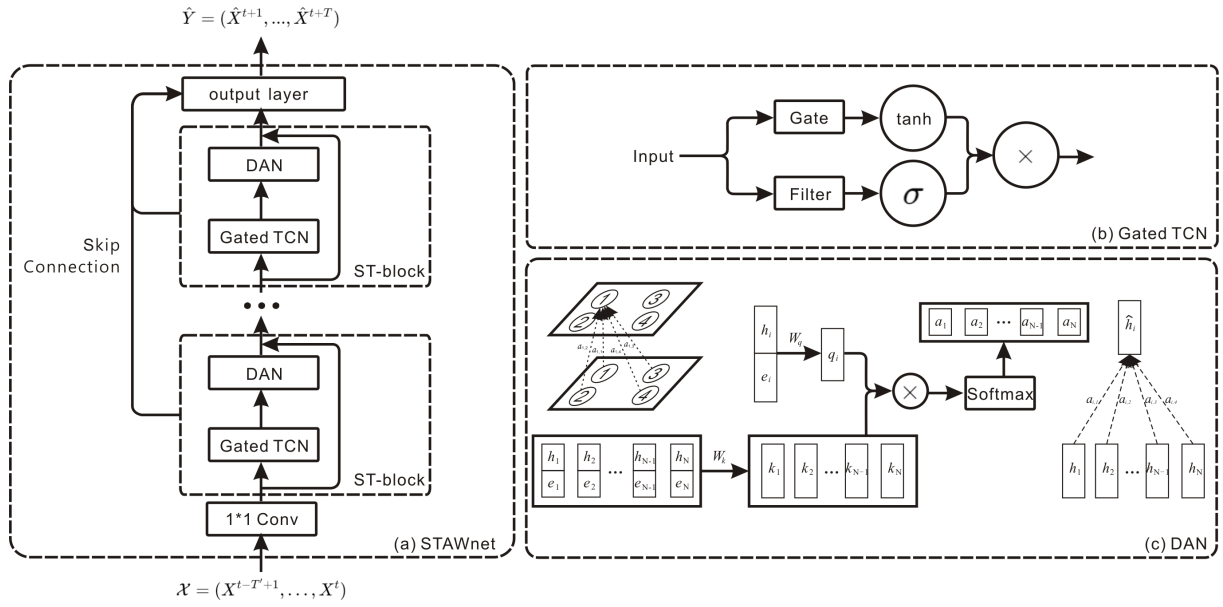
Fig. 7 STAWnet: self-learned node embeddings capture hidden spatial dependencies, while temporal convolution and dynamic attention capture temporal dynamics.
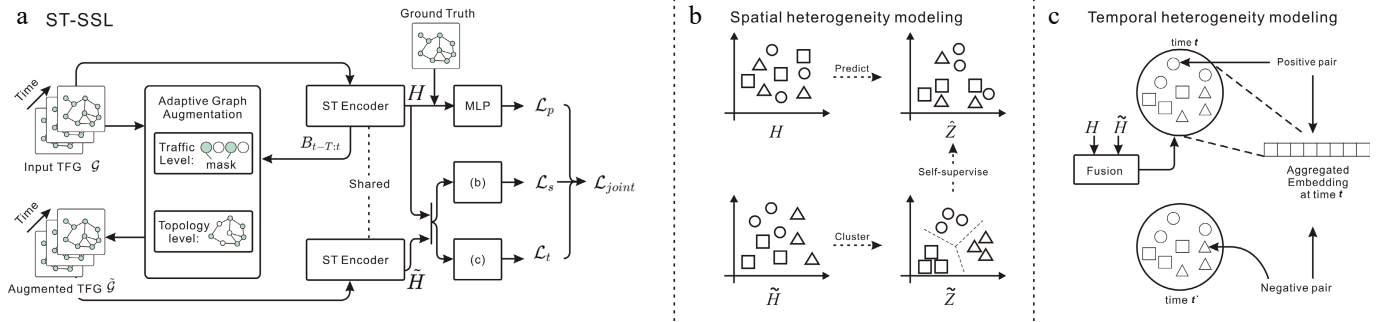

Fig. 8 ST-SSL: integrates temporal and spatial convolutions for spatiotemporal patterns, enhanced by self-supervised tasks for spatial-temporal heterogeneity.
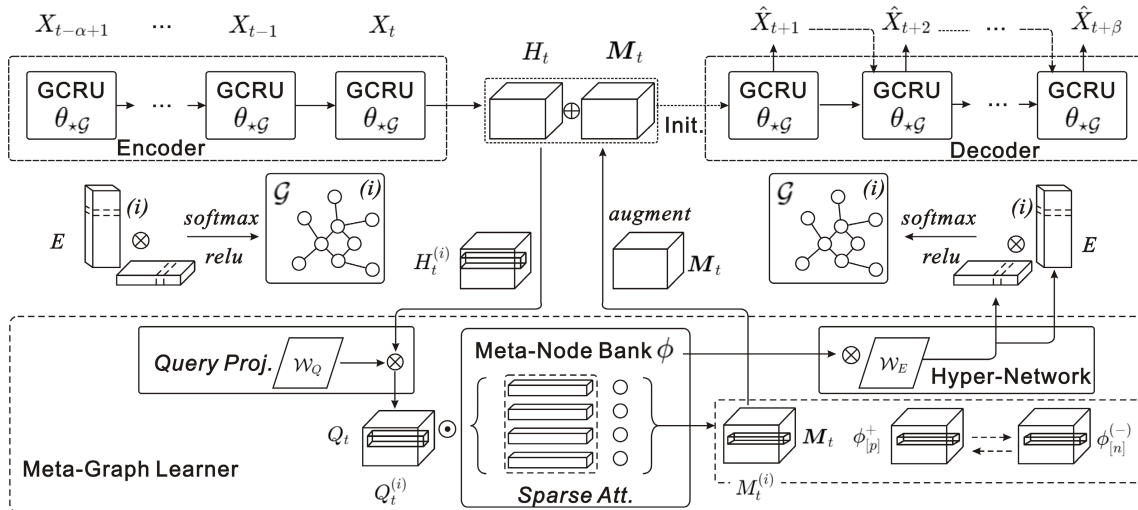

Fig. 9 MegaCRN: meta-graph learner dynamically generates node embeddings capturing spatiotemporal nuances, integrated with a GCRN for prediction.

STWave[59] addresses the challenge of non-stationary traffic time series by adopting a disentangle-and-fuse approach. STWave decouples complex traffic data into stable trends and fluctuating events using the discrete wavelet transform (DWT), thereby mitigating the distribution shift problem. It then employs a dual-channel spatiotemporal network to separately model these components and integrates them for improved future traffic prediction. A key contribution is the introduction of an efficient Spectral Graph
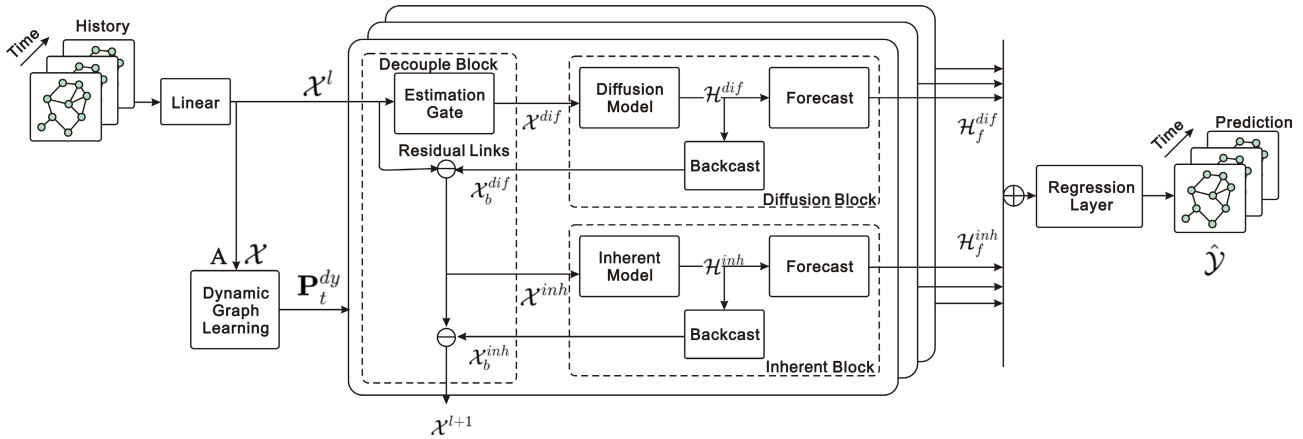
Fig. 10 D2STGNN: decouples diffuse and intrinsic signals via a decoupled spatiotemporal framework with dynamic graph learning.
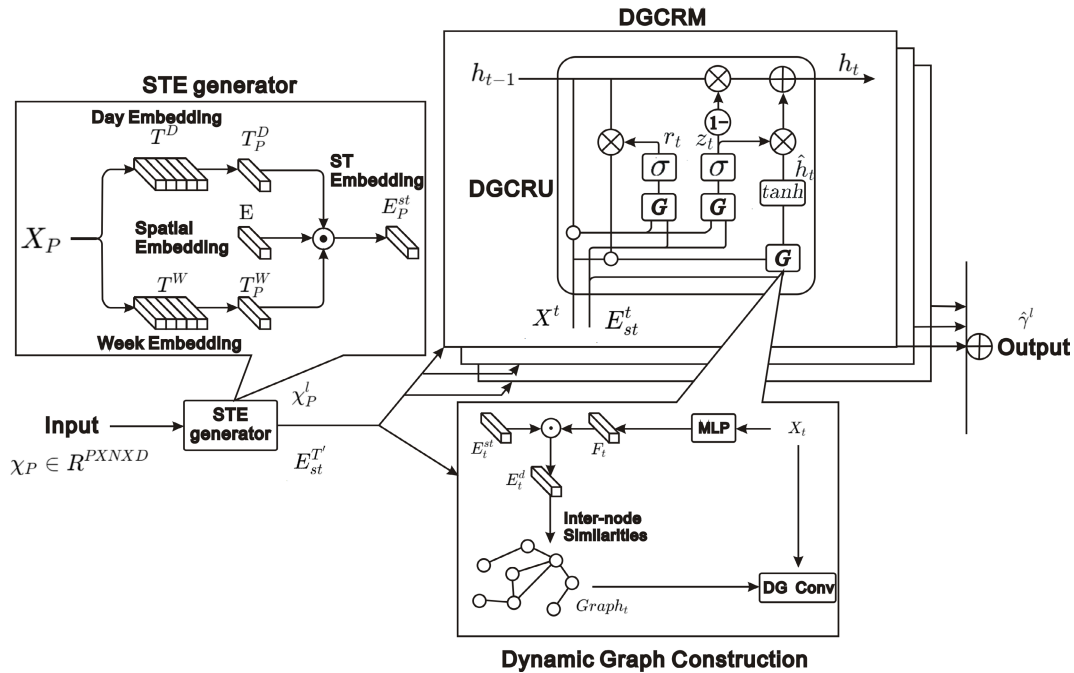

Fig. 11 DDGCRN: dynamically generates graphs from traffic signals and uses DGCRM to capture spatiotemporal features efficiently.

Attention Network (ESGAT) with a novel query sampling strategy and graph wavelet-based positional encoding. This design enhances the modeling of dynamic spatial correlations while reducing computational complexity (Fig. 13).

Transformer-based model

The Transformer model has revolutionized the field of Natural Language Processing (NLP) since it was proposed by Vaswani et al.[82] particularly excelling in tasks such as machine translation and text summarization. The core self-attention mechanism is able to capture long-distance dependencies in sequential data, thereby overcoming the limitations of traditional RNNs in processing long sequences. With the continuous advancement of deep learning, the Transformer architecture has gradually been introduced into other domains, including traffic flow prediction.

This subsection focuses on two types of Transformer-based models: the STAEFormer and the PDFormer. Both models have achieved remarkable results in the field of traffic flow prediction and have improved and extended the Transformer architecture from different aspects to better adapt to the characteristics of traffic data.

STAEFormer[60] leverages a simple yet effective spatiotemporal adaptive embedding technique to enhance the capabilities of vanilla Transformer models, achieving state-of-the-art performance. The key innovation lies in the model's ability to capture the complex spatiotemporal dynamics and chronological information inherent in traffic time series data, which has traditionally been a challenge for forecasting models. This work focuses on improving the representation of the input data rather than complicating the model architecture (Fig. 14).

PDFormer[61] introduces a novel propagation delay-aware dynamic long-range Transformer specifically designed to address the complex spatial-temporal dependencies in traffic flow prediction. The model innovatively integrates a spatial self-attention module to capture dynamic spatial relationships, employs graph masking matrices to emphasize both short- and long-range spatial dependencies, and introduces a delay-aware feature transformation module to account for the time delay in traffic condition propagation. This comprehensive approach effectively models the temporal dynamics and spatial heterogeneities in traffic data, leading to improved accuracy and interpretability in traffic flow predictions (Fig. 15).
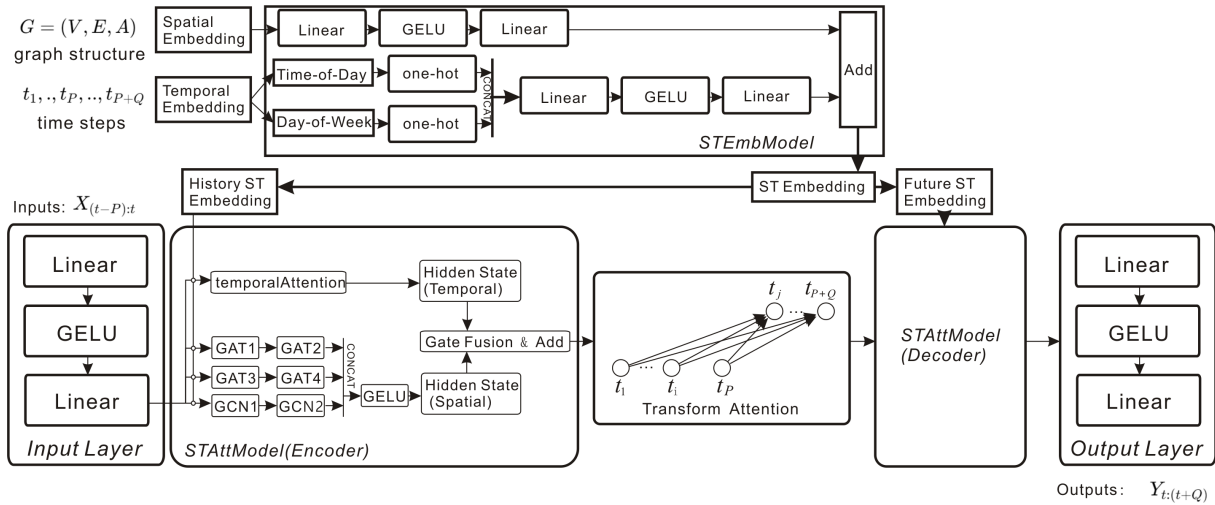
Fig. 12 RGDAN: random GAT with adaptive matrix captures spatial dependencies, while temporal attention captures time-related patterns.
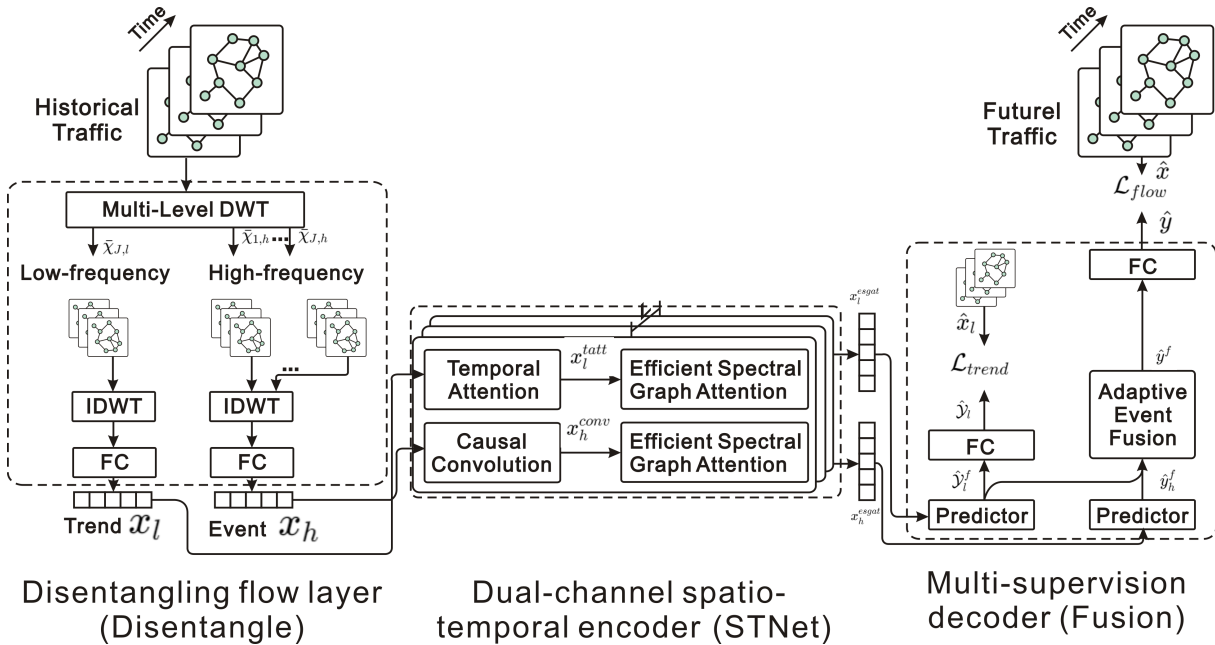

Fig. 13 STWave: decouples traffic series into trends and events with DWT, and uses dual-channel spatiotemporal network and ESGAT for modeling.
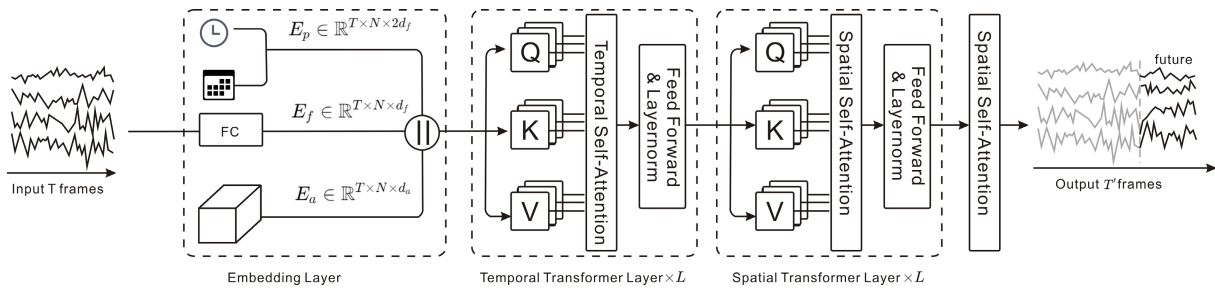

Fig. 14 STAEFormer: spatiotemporal adaptive embedding enhances Transformer to capture complex spatiotemporal dynamics in traffic data.

Pre-training model

STEP[62], STGNN is enhanced by a scalable time series pre-training model, which is the first traffic flow prediction model to employ pre-training techniques. The motivation of STEP is based on STGNN. Although STGNN shows good performance in MTS prediction, it only considers short-term historical MTS data due to model complexity. To leverage long-term historical MTS data for analyzing temporal and spatial patterns, STEP proposes a novel framework. It introduces a pre-training model to efficiently learn segment-level representations from long-term historical MTS data, using these representations as context for STGNN (Fig. 16).

STD-MAE[63], short for Spatial-Temporal-Decoupled Masked Pre-training, shares the same motivation as STEP. The method aims to overcome the limitation of input series length. It employs two decoupled masked autoencoders to reconstruct spatiotemporal series along the spatial and temporal dimensions. Compared with
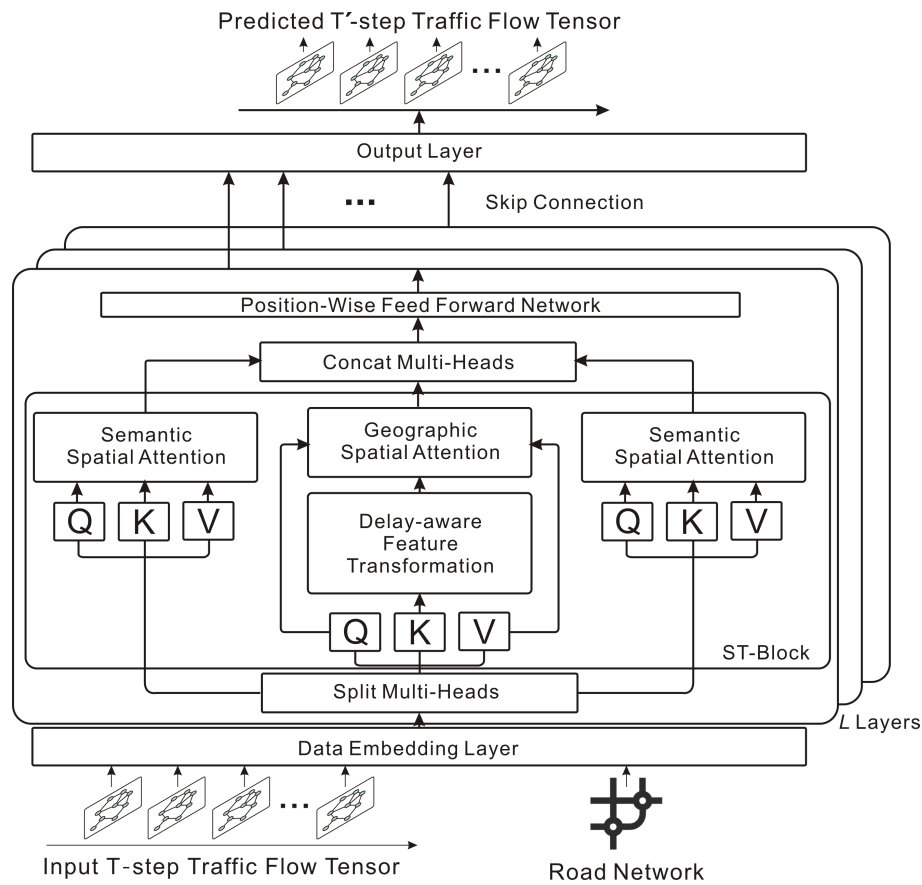
Fig. 15 PDFormer: spatial self-attention, graph masking, and delay-aware transformation model complex spatiotemporal dependencies.
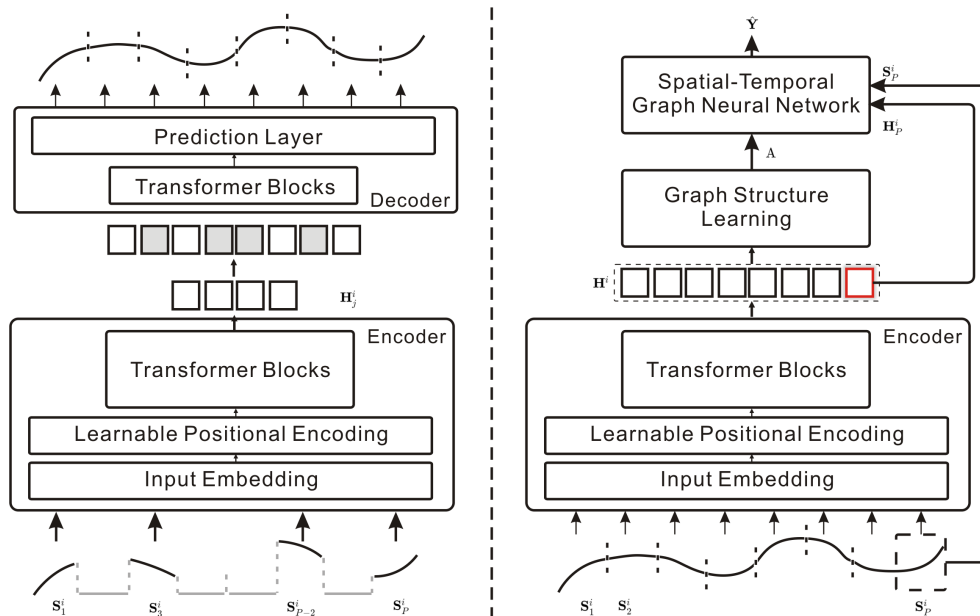

Fig. 16 STEP: Pre-training on long-term traffic sequences provides segment-level representations to enhance STGNN predictions.

STEP, STD-MAE introduces an additional pre-training step along the spatial dimension. By applying masking separately along the two dimensions, the model can effectively capture long-range heterogeneity in MTS data (Fig. 17).

These two pre-trained models have achieved good performance. The pre-training is designed to remove the limitation imposed by the length of the input sequence. More importantly, they are not constrained by the limitations of data sources. A wider range of data can be utilized for pre-training, rather than being restricted to traffic flow data from a single area.

Differential equations model

Differential equations offer a natural framework for representing traffic flow as a continuous function of both time and space. The combination of neural networks and differential equations has
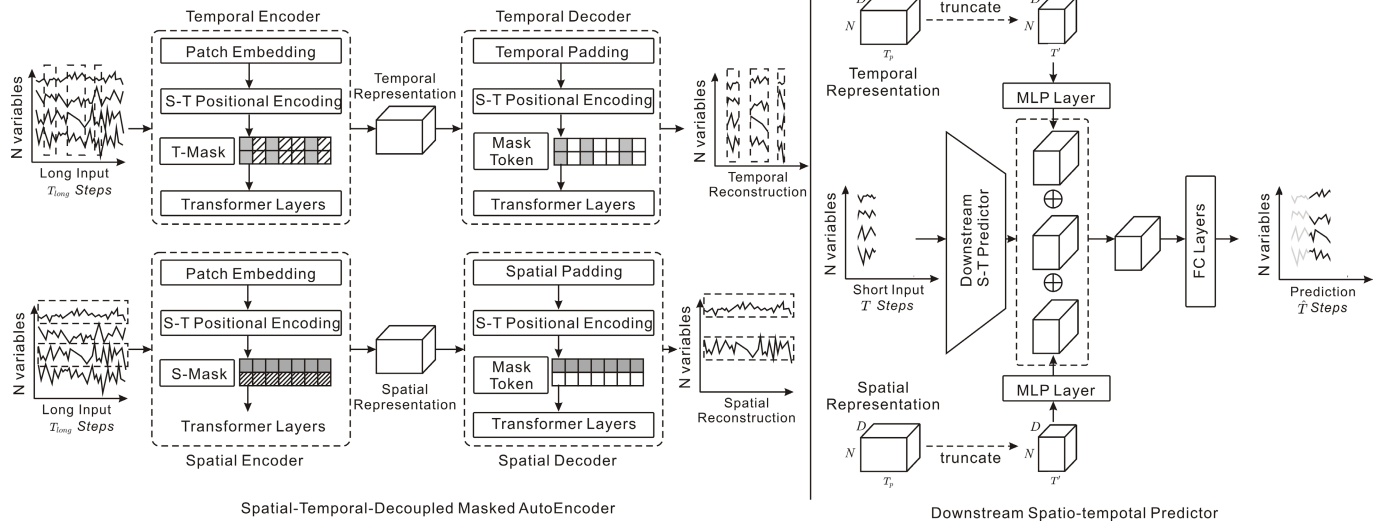
Fig. 17 STD-MAE: Decoupled spatial-temporal masked autoencoders reconstruct sequences to capture long-range heterogeneity in traffic data.

given rise to a new class of models that integrate the representational power of deep learning with the theoretical rigor of differential equation modeling. Often referred to as Neural Controlled Differential Equations (NCDEs) and Neural Rough Differential Equations (NRDEs), these models have achieved breakthroughs in sequential data modeling by treating sequential data as observations of continuous-time dynamics (Fig. 18).

STG-NCDE[64] integrates two distinct Neural Controlled Differential Equations (NCDEs) within a unified framework. One NCDE focuses on the temporal dimension, capturing the evolution of traffic patterns over time, while the other NCDE addresses the spatial dimension, modeling the complex interactions and dependencies among different locations in a road network.

STG-NRDE[65] utilizes Neural Rough Difference Equations (NRDEs) to process sequential data and extends the concept to temporal and spatial dimensions. The authors designed two NRDEs, one for each dimension, and integrated them into a cohesive framework to capture the complex dynamics of traffic flows. The key innovation of this work is its ability to handle irregular time-series data and model the intricate interdependencies in traffic data more effectively than traditional approaches.

Empirical analysis

To establish a fair benchmarking protocol and provide practical guidance for model selection in real-world applications, a systematic evaluation was conducted on the PEMS04 dataset. Specifically, a diverse set of representative spatiotemporal forecasting models were re-implemented, including the following: DCRNN[49], STGCN[50], GraphWaveNet[51], STSGCN[52], STAWnet[53], ST-SSL[54], MegaCRN[55], D2STGNN[56], DDGCRN[57], RGDAN[58], STWave[59], STAEFormer[60], PDFormer[61], STEP[62], STD-MAE[63], STG-NCDE[64], and STG-NRDE[65]. The prediction task is formulated as forecasting the next 12 time steps using a 12-step lookback window, with each step corresponding to 5 min, so that 12 steps cover 1 h in the PEMS04 dataset.

To ensure methodological rigor and comparability, all models were trained and evaluated using the same hyperparameter settings as reported in the original papers. In addition, the computational environment was unified across all experiments (see Table 3) to minimize potential biases arising from hardware or software variations.

During benchmarking, various aspects of model performance were systematically tracked, including GPU memory footprint, CPU memory usage, parameter size, predictive accuracy, and runtime
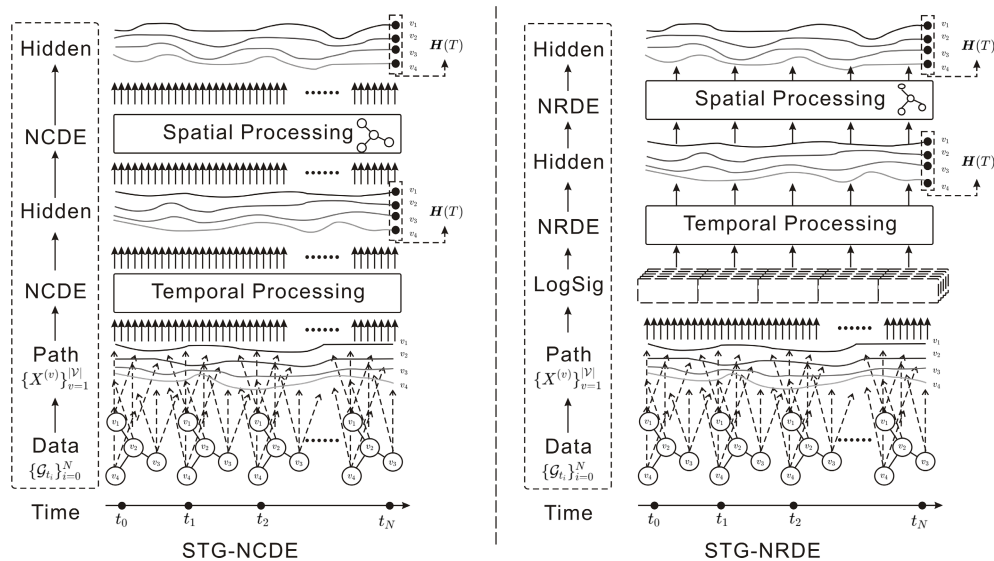

Fig. 18 STG-NCDE and STG-NRDE: Temporal and spatial NCDEs/NRDEs jointly model continuous-time dynamics and intricate patterns in traffic flows.

efficiency. Figure 19 provides an overview of each model's predictive performance, peak GPU memory usage, and parameter size, providing a clear reference for model selection in practical applications. Table 4 presents detailed runtime statistics for all evaluated models. Note that HA, ARIMA, and VAR are not deep learning models and do not involve a training process; therefore, some metrics, including GPU memory usage, are omitted.

Figure 20 presents the relationship between training time and MAE, illustrating the model's convergence speed toward a target performance level. This curve provides insights into the model's training efficiency, indicating not only the total computational cost required to achieve a given accuracy but also its convergence rate relative to other models. Such analysis facilitates evaluation of both effectiveness and practical applicability in real-world scenarios where training time is often a critical factor.

Future directions and challenges

Deep neural networks have driven the development of traffic flow prediction. However, there are still some challenges in this field. In this section, future directions and challenges are discussed.

Few-shot learning and transfer learning

Effective traffic flow prediction remains challenging under insufficient data conditions, especially in the early stages of building new roads or traffic networks. Some cities may lack sufficient sensing infrastructure, resulting in limited available data[83]. Even in cities with a sufficient number of sensors, the amount of usable data in practical deployments can be severely limited due to the presence of noisy or missing measurements. This issue is particularly

Table 3. Device information.

Device	Info
CPU	Intel(R) Xeon(R) Gold 6133 CPU @ 2.50GHz
GPU	NVIDIA GeForce RTX 4090
RAM	32 GB
VRAM	24 GB
Python	3.10.16
CUDA	12.3
PyTorch	2.6.0

pronounced when considering multiple nodes simultaneously over a continuous time interval, highlighting data scarcity as a critical challenge. For instance, a case study is conducted using the PEMS04 dataset and the GraphWaveNet model. The model is evaluated with varying proportions of the training set—10%, 30%, 50%, 70%, and 100%—while keeping the validation and test sets unchanged. All models are trained with identical hyperparameter settings. Table 5 presents the results of this experiment.

Notably, it was observed that using only 10% of the training set outperforms models trained with 30%, 50%, or 70% of the data. This counterintuitive result may be attributed to sampling bias or noise distribution in the training subsets. Nonetheless, overall, model performance generally improves as the training data volume increases. This indicates that data quantity remains a critical factor affecting model effectiveness. Consequently, few-shot learning and transfer learning techniques have become key research focuses[84–86].

Few-shot learning and transfer learning are two effective approaches for addressing data scarcity[87]. Few-shot learning aims to achieve effective model training and generalization with very limited samples[88]. It improves the generalization ability of models with few samples through methods such as metric learning and data augmentation. Transfer learning enhances the performance of target domain models by transferring pre-trained models or knowledge from one domain (source domain) to another domain (target domain)[89]. It can effectively predict traffic conditions in newly built roads or regions with sparse traffic data using techniques such as model fine-tuning, feature transfer, and cross-domain adaptation.

In recent years, few-shot learning and transfer learning have shown preliminary progress in traffic flow forecasting tasks. Chen et al.[90] proposed a cross-city transfer learning framework that integrates multi-semantic fusion with hierarchical graph clustering, which enables the model to capture dynamic traffic states, preserve static spatial dependencies, and leverage multi-granular information via simultaneous multi-scale prediction. Ouyang et al.[91] proposed a graph-based domain adversarial framework for cross-city spatiotemporal prediction, leveraging self-adaptive spatiotemporal knowledge, a knowledge attention mechanism, and dual domain discriminators to extract domain-invariant features. Li et al.[92] proposed a physics-guided multi-source transfer learning method
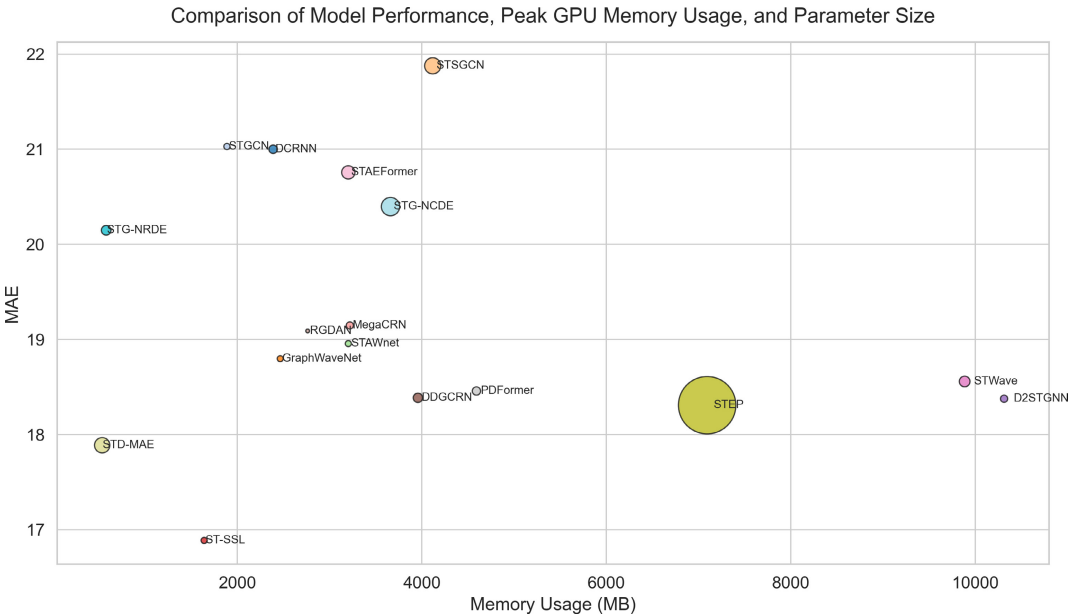

Fig. 19 Model performance vs memory and parameters.

Table 4. Runtime, resource usage, and predictive performance of models on PEMS04.

Model	Params	MAE	RMSE	MAPE	VRAM	RAM	Iter per second
HA	–	38.03	59.24	27.88	–	–	–
ARIMA	–	33.73	48.80	24.18	–	–	–
VAR	–	24.54	38.61	17.24	–	–	–
DCRNN	546,881	21.00	33.16	14.66	2,385.22	1,276.96	3.09
STGCN	297,228	21.03	33.38	15.07	1,886.88	1,177.90	14.30
GraphWaveNet	278,632	18.80	30.87	12.38	2,461.48	1,241.01	5.46
STSGCN	2,024,445	21.88	35.49	15.29	4,114.91	1,028.78	7.09
STAWnet	281,756	18.96	31.35	12.72	3,200.94	1,254.59	5.21
ST-SSL	288,275	16.89	26.98	10.48	3,208.00	1,566.00	16.19
MegaCRN	392,761	19.15	31.36	12.58	3,217.96	1,140.38	7.72
D2STGNN	398,788	18.38	30.10	12.06	10,310.0	5,067.00	6.34
DDGCRN	671,704	18.39	30.73	12.09	3,957.25	1,441.35	8.38
RGDAN	107,285	19.09	32.10	12.99	2,759.20	1,164.68	19.12
STWave	882,558	18.56	30.36	12.81	9,884.15	4,712.99	2.17
STAEFormer	1,354,932	20.76	32.38	20.06	3,200.36	1,217.37	19.88
PDFormer	531,165	18.46	29.97	12.67	4,591.52	5,351.85	1.64
STEP	25,469,726	18.31	29.90	12.49	7,091.23	1,936.99	2.02
STD-MAE	673,384	17.89	29.33	12.11	532.70	3,628.76	99.72
STG-NRDE	716,628	20.15	32.22	13.51	577.83	3,739.68	5.26
STG-NCDE	2,600,532	20.40	32.30	13.95	3,657.37	3,793.33	0.85
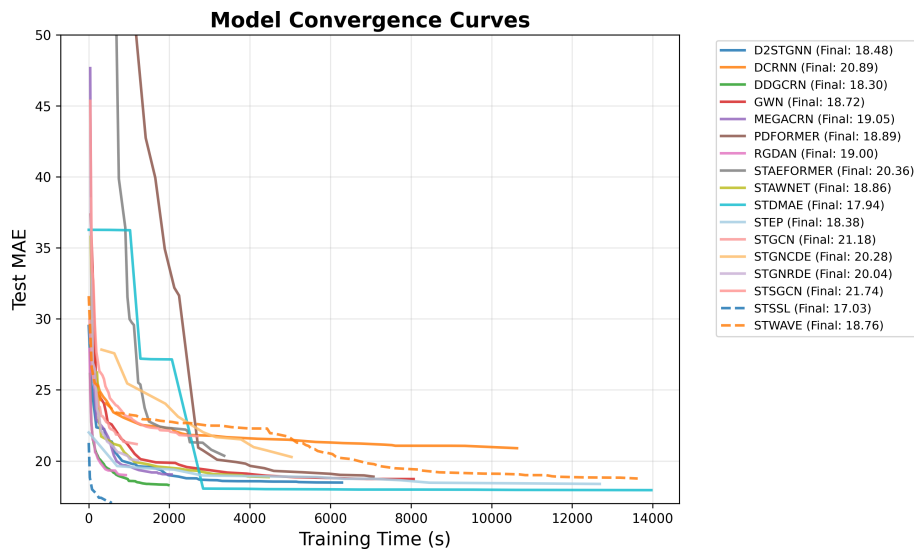
Fig. 20 Training convergence curves of MAE over time.

Table 5. Model performance under different training data proportions.

Training data proportion	MAE	RMSE	MAPE
10%	19.66	31.77	13.08%
30%	20.12	32.54	13.15%
50%	19.87	31.97	13.64%
70%	19.50	31.54	12.66%
100%	18.82	30.78	12.41%

for multi-region traffic flow, which leverages prior knowledge from observations, empirical studies, and traffic network physical properties, and transfers network traffic features using adversarial training combined with an Macroscopic Fundamental Diagram (MFD)-based weighting scheme.

Future research in traffic prediction could further explore cross-city transfer learning and few-shot learning to address challenges posed by data scarcity in newly deployed sensors or regions. One promising direction is the development of dynamic multi-source transfer frameworks that leverage both static and evolving traffic network properties. These frameworks incorporate long-term, high-resolution data and real-time traffic regulation records. In parallel, advanced adaptation techniques are needed to mitigate issues such as overfitting and catastrophic forgetting, ensuring robust knowledge transfer across heterogeneous spatiotemporal domains, such as trajectories, weather, and points of interest. Combining these strategies may enable the development of more generalizable and adaptive models capable of few-shot learning in new cities or traffic scenarios.

Multimodal data in traffic flow prediction

Subtle changes in traffic flow often have a great impact on future traffic flow. Due to the inherent randomness of traffic flow and external noise, such as accidents, weather conditions, manual traffic control, or detector malfunctions, accurately identifying effective changes in traffic flow and filtering noise is challenging[93]. Multi-modal data fusion is an effective approach that integrates multiple data sources and thus improves the accuracy of traffic flow prediction. By integrating multiple data sources, such as traffic surveillance videos, GPS trajectories, meteorological information, and social media data[94], researchers can build more comprehensive

Zhang et al. Digital Transportation and Safety 2025, 4(4): 281−297

Page 293 of 297

and robust traffic flow prediction models. These diverse data sources provide rich information that can help the model better understand the complex dynamics and potential patterns of traffic flow.

Multimodal data fusion techniques aim to extract features from different data sources and effectively combine these features to enhance the model's predictive capability. Multi-view learning and deep fusion strategies are two commonly used methods[95]. Multi-view learning captures the diversity of traffic flow by processing data from different views[96], while deep fusion strategies use deep learning models, such as convolutional neural networks and graph neural networks, to deeply integrate multimodal data[97], thereby making full use of the complementary advantages of different data sources and improving the accuracy and robustness of predictions.

Future research can further explore the application of multimodal data fusion in traffic flow prediction. This can improve model adaptability in diverse data environments and enhance the understanding of complex traffic scenarios by incorporating additional data sources. Meanwhile, as sensor technology and data collection methods continue to advance, multimodal data fusion is expected to be widely applied in real-world traffic scenarios, promoting the development of intelligent transportation systems and enabling more accurate and efficient traffic management.

Development of a lightweight and effective traffic flow prediction model

As the scale of the transportation network expands, a large amount of traffic flow data is being collected. These data provide a wealth of information for traffic flow prediction, but also impose higher demands on model computational capabilities[98]. Currently, models require substantial computational resources to process large-scale road network data[99], making the development of lightweight and effective models an important future research direction.

Future research can further explore the use of efficient graph neural network algorithms and parallel computing techniques in traffic flow prediction. Such approaches can enhance model processing capabilities for large-scale data while reducing computational costs through optimized resource utilization. With ongoing advancements in computing hardware and algorithms, lightweight and efficient models are expected to see broader application in real-world traffic scenarios, thereby supporting the development of intelligent transportation systems and enabling more accurate and efficient traffic management.

Conclusions

In this article, machine learning–based traffic flow prediction is reviewed. First, the background of traffic flow prediction tasks is introduced, and the problem definition of traffic flow prediction is provided. Three common traffic flow metrics and three standard measures of model predictive performance are summarized. Next, the 15 traffic flow datasets used in previous studies are reviewed, with detailed descriptions of how these datasets have been preprocessed.

The discussion begins with classical machine learning models and then explores more recent deep learning models. In particular, graph neural networks and spatiotemporal convolutional networks are discussed, including state-of-the-art pre-trained models and differential equation–based models. Finally, based on this review, three future research directions for traffic flow prediction are proposed. These directions include few-shot and transfer learning, multimodal data integration, and the development of lightweight and efficient models.

Data availability

The data that support the findings of this study are available in the Github repository https://github.com, and listed in the Supplementary Table S1; further details can be found in references [49–65] and [100–102].

Conflict of interest

The authors declare that they have no conflict of interest.

Supplementary information accompanies this paper at (https://www.maxapress.com/article/doi/10.48130/dts-0025-0027)

References

1. Xu W, Liu J, Yan J, Yang J, Liu H, et al. 2024. Dynamic spatiotemporal graph wavelet network for traffic flow prediction. *IEEE Internet of Things Journal* 11(5):8019−29
2. Abadi M, Barham P, Chen J, Chen Z, Davis A, et al. 2016. TensorFlow: a system for large-scale machine learning. *Proc. 12th USENIX symposium on operating systems design and implementation (OSDI 16), Savannah, GA, USA, November 2−4, 2016.* USA: USENIX Association. pp. 265−83 www.usenix.org/system/files/conference/osdi16/osdi16-abadi.pdf
3. Sayed SA, Abdel-Hamid Y, Hefny HA. 2023. Artificial intelligence-based traffic flow prediction: a comprehensive review. *Journal of Electrical Systems and Information Technology* 10:13
4. Cai D, Chen K, Lin Z, Li D, Zhou T, et al. 2024. JointSTNet: Joint Pre-Training for Spatial-Temporal Traffic Forecasting. *IEEE Transactions on Consumer Electronics* 71(2):6239−52
5. Wang F, Liang Y, Lin Z, Zhou J, Zhou T. 2024. SSA-ELM: a hybrid learning model for short-term traffic flow forecasting. *Mathematics* 12:1895
6. Chai W, Zhang L, Lin Z, Zhou J, Zhou T. 2024. GSA-KELM-KF: a hybrid model for short-term traffic flow forecasting. *Mathematics* 12:103
7. Wen Y, Xu P, Li Z, Xu W, Wang X. 2023. RPConvformer: a novel Transformer-based deep neural networks for traffic flow prediction. *Expert Systems with Applications* 218:119587

8. Cui Z, Huang B, Dou H, Tan G, Zheng S, et al. 2022. Gsa‑elm: a hybrid learning model for short‑term traffic flow forecasting. *IET Intelligent Transport Systems* 16:41−52

9. Abdullah SM, Periyasamy M, Kamaludeen NA, Towfek SK, Marappan R, et al. 2023. Optimizing traffic flow in smart cities: soft GRU-based recurrent neural networks for enhanced congestion prediction using deep learning. *Sustainability* 15:5949

10. Wu K, Xu C, Yan J, Wang F, Lin Z, et al. 2023. Error-distribution-free kernel extreme learning machine for traffic flow forecasting. *Engineering Applications of Artificial Intelligence* 123:106411

11. Chai W, Luo Q, Lin Z, Yan J, Zhou J, et al. 2024. Spatiotemporal dynamic multi-hop network for traffic flow forecasting. *Sustainability* 16:5860

12. Xing Z, Huang M, Peng D. 2023. Overview of machine learning-based traffic flow prediction. *Digital Transportation and Safety* 2:164−75

13. Smith BL, Demetsky MJ. 1997. Traffic flow forecasting: comparison of modeling approaches. *Journal of Transportation Engineering* 123:261−66

14. Williams BM, Hoel LA. 2003. Modeling and forecasting vehicular traffic flow as a seasonal ARIMA process: Theoretical basis and empirical results. *Journal of transportation engineering* 129:664−72

15. Schimbinschi F, Moreira-Matias L, Nguyen VX, Bailey J. 2017. Topology-regularized universal vector autoregression for traffic forecasting in large urban areas. *Expert Systems with Applications* 82:301−16

16. Tan G, Zhou T, Huang B, Dou H, Song Y, et al. 2024. A noise-immune and attention-based multi-modal framework for short-term traffic flow forecasting. *Soft Computing* 28:4775−90

17. Lu H, Ge Z, Song Y, Jiang D, Zhou T, et al. 2021. A temporal-aware LSTM enhanced by loss-switch mechanism for traffic flow forecasting. *Neurocomputing* 427:169−78

18. Liu M, Liu G, Sun L. 2023. Spatial–temporal dependence and similarity aware traffic flow forecasting. *Information Sciences* 625:81−96

19. Yang S, Li H, Luo Y, Li J, Song Y, et al. 2022. Spatiotemporal adaptive fusion graph network for short-term traffic flow forecasting. *Mathematics* 10:1594

20. Lin Z, Wang D, Cao C, Xie H, Zhou T, et al. 2025. GSA-KAN: a hybrid model for short-term traffic forecasting. *Mathematics* 13:1158

21. Fernandes B, Silva F, Alaiz-Moretón H, Novais P, Analide C, et al. 2019. Traffic flow forecasting on data-scarce environments using ARIMA and LSTM networks. In *New Knowledge in Information Systems and Technologies. WorldCIST'19 2019. Advances in Intelligent Systems and Computing*, eds. Rocha Á, Adeli H, Reis L, Costanzo S. Vol 930. Cham: Springer. pp. 273−82 doi: 10.1007/978-3-030-16181-1_26

22. Zhao Z, Chen W, Wu X, Chen PCY, Liu J. 2017. LSTM network: a deep learning approach for short-term traffic forecast. *IET intelligent transport systems* 11:68−75

23. Fu R, Zhang Z, Li L. Using LSTM and GRU neural network methods for traffic flow prediction. *2016 31st Youth Academic Annual Conference of Chinese Association of Automation (YAC), Wuhan, China, 11−13 November 2016*. USA: IEEE. pp. 324−28 doi: 10.1109/YAC.2016.7804912

24. 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. *Transportation Research Part C: Emerging Technologies* 54:187−97

25. Fang W, Li X, Lin Z, Zhou J, Zhou T. 2024. Mixture correntropy with variable center LSTM network for traffic flow forecasting. *Digital Transportation and Safety* 3:264−70

26. Qi Q, Cheng R, Ge H. 2023. Short-term inbound rail transit passenger flow prediction based on BILSTM model and influence factor analysis. *Digital Transportation and Safety* 2:12−22

27. Yang D, Li S, Peng Z, Wang P, Wang J, et al. 2019. MF-CNN: traffic flow prediction using convolutional neural network and multi-features fusion. *IEICE Transactions on Information and Systems* E102.D:1526−36

28. Méndez M, Merayo MG, Núñez M. 2023. Long-term traffic flow forecasting using a hybrid CNN-BiLSTM model. *Engineering Applications of Artificial Intelligence* 121:106041

29. Zhang W, Yu Y, Qi Y, Shu F, Wang Y. 2019. Short-term traffic flow prediction based on spatio-temporal analysis and CNN deep learning. *Transportmetrica A: Transport Science* 15:1688−711

30. Narmadha S, Vijayakumar V. 2023. Spatio-Temporal vehicle traffic flow prediction using multivariate CNN and LSTM model. *Materials Today: Proceedings* 81:826−33

31. Huang B, Dou H, Luo Y, Li J, Wang J, et al. 2022. Adaptive spatiotemporal transformer graph network for traffic flow forecasting by iot loop detectors. *IEEE Internet of Things Journal* 10:1642−53

32. Jiang W, Xiao Y, Liu Y, Liu Q, Li Z. 2022. Bi‑GRCN: a spatio‑temporal traffic flow prediction model based on graph neural network. *Journal of Advanced Transportation* 2022:5221362

33. Li Z, Zhou J, Lin Z, Zhou T. 2024. Dynamic spatial aware graph transformer for spatiotemporal traffic flow forecasting. *Knowledge-based systems* 297:111946

34. Zhang H, Lin Z, Xie H, Zhou J, Song Y, et al. 2025. Two-way heterogeneity model for dynamic spatiotemporal traffic flow prediction. *Knowledge-Based Systems* 320:113635

35. Chai W, Zheng Y, Tian L, Qin J, Zhou T. 2023. GA-KELM: genetic-algorithm-improved kernel extreme learning machine for traffic flow forecasting. *Mathematics* 11:3574

36. Cui Z, Huang B, Dou H, Cheng Y, Guan J, et al. 2022. A two-stage hybrid extreme learning model for short-term traffic flow forecasting. *Mathematics* 10:2087

37. Ou J, Li J, Wang C, Wang Y, Nie Q. 2024. Building trust for traffic flow forecasting components in intelligent transportation systems via interpretable ensemble learning. *Digital Transportation and Safety* 3:126−43

38. Zou G, Lai Z, Wang T, Liu Z, Li Y. 2024. MT-STNet: a novel multi-task spatiotemporal network for highway traffic flow prediction. *IEEE Transactions on Intelligent Transportation Systems* 25(7):8221−36

39. Goyal MTSP, Gulghane A. 2020. A review of speed flow density study of two different road Indian road and their comparison. *International Journal of Scientific Research & Engineering Trends* 6(2):499-504

40. Dorokhin S, Artemov A, Likhachev D, Novikov A, Starkov E. 2020. Traffic simulation: an analytical review. *IOP Conference Series: Materials Science and Engineering* 918:012058

41. Zhang L, Yuan Z, Yang L, Liu Z. 2020. Recent developments in traffic flow modelling using macroscopic fundamental diagram. *Transport Reviews* 40:689−710

42. Bramich DM, Menéndez M, Ambühl L. 2022. Fitting empirical fundamental diagrams of road traffic: A comprehensive review and comparison of models using an extensive data set. *IEEE Transactions on Intelligent Transportation Systems* 23:14104−27

43. Liu J, Wu N, Qiao Y, Li Z. 2021. A scientometric review of research on traffic forecasting in transportation. *IET Intelligent Transport Systems* 15:1−16

44. Kashyap AA, Raviraj S, Devarakonda A, Nayak KSR, K V S, et al . 2022. Traffic flow prediction models – a review of deep learning techniques. *Cogent Engineering* 9:2010510

45. Jarmuż D, Chmiel J . 2020. A review of approaches to the study of weather's effect on road traffic parameters. *Transport Problems* 15:241−51

46. Fang W, Cai W, Fan B, Yan J, Zhou T. 2021. Kalman-LSTM model for short-term traffic flow forecasting. *2021 IEEE 5th Advanced Information Technology, Electronic and Automation Control Conference (IAEAC), Chongqing, China, 12−14 March 2021*. USA: IEEE. pp. 1604−8 doi: 10.1109/IAEAC50856.2021.9390991

47. Jiang W, Luo J. 2022. Graph neural network for traffic forecasting: a survey. *Expert Systems with Applications* 207:117921

48. Chen X, Lu J, Zhao J, Qu Z, Yang Y, et al. 2020. Traffic flow prediction at varied time scales via ensemble empirical mode decomposition and artificial neural network. *Sustainability* 12:3678

49. Li Y, Yu R, Shahabi C, Liu Y. 2017. Diffusion convolutional recurrent neural network: data-driven traffic forecasting. *arXiv* 1707.01926

50. Yu B, Yin H, Zhu Z. 2017. Spatio-temporal graph convolutional networks: a deep learning framework for traffic forecasting. *Proceedings of the Twenty-Seventh International Joint Conference on Artificial Intelligence*. California: International Joint Conferences on Artificial Intelligence (IJCAI). pp. 3634−40 doi: 10.24963/ijcai.2018/505

51. Wu Z, Pan S, Long G, Jiang J, Zhang C. 2019. Graph wavenet for deep spatial-temporal graph modeling. *Proceedings of the 28th International Joint Conference on Artificial Intelligence, Macao, China 2019*. Macao,

China: AAAI Press. pp. 1907−13 https://dl.acm.org/doi/abs/10.5555/3367243.3367303

52. Song C, Lin Y, Guo S, Wan H. 2020. Spatial-temporal synchronous graph convolutional networks: A new framework for spatial-temporal network data forecasting. *Proceedings of the AAAI Conference on Artificial Intelligence* 34:914−21

53. Tian C, Chan WK. 2021. Spatial-temporal attention wavenet: a deep learning framework for traffic prediction considering spatial-temporal dependencies. *IET Intelligent Transport Systems* 15:549−61

54. Ji J, Wang J, Huang C, Wu J, Xu B, et al. 2023. Spatio-temporal self-supervised learning for traffic flow prediction. *Proceedings of the AAAI Conference on Artificial Intelligence* 37:4356−64

55. Jiang R, Wang Z, Yong J, Jeph P, Chen Q, et al. 2023. Spatio-temporal meta-graph learning for traffic forecasting. *Proceedings of the AAAI Conference on Artificial Intelligence* 37:8078−86

56. Shao Z, Zhang Z, Wei W, Wang F, Xu Y, et al. 2022. Decoupled dynamic spatial-temporal graph neural network for traffic forecasting. *arXiv* 2206.09112

57. Weng W, Fan J, Wu H, Hu Y, Tian H, et al. 2023. A decomposition dynamic graph convolutional recurrent network for traffic forecasting. *Pattern Recognition* 142:109670

58. Fan J, Weng W, Tian H, Wu H, Zhu F, et al. 2024. RGDAN: a random graph diffusion attention network for traffic prediction. *Neural Networks* 172:106093

59. Fang Y, Qin Y, Luo H, Zhao F, Xu B, et al. 2023. When spatio-temporal meet wavelets: disentangled traffic forecasting via efficient spectral graph attention networks. *2023 IEEE 39th International Conference on Data Engineering (ICDE), Anaheim, CA, USA, 3−7 April 2023* . USA: IEEE. pp. 517−29 doi: 10.1109/ICDE55515.2023.00046

60. Liu H, Dong Z, Jiang R, Deng J, Deng J, et al.2023. Spatio-temporal adaptive embedding makes vanilla transformer SOTA for traffic forecasting. *CIKM '23: Proceedings of the 32nd ACM International Conference on Information and Knowledge Management. Birmingham United Kingdom, October 21−25, 2023*. New York, United States: Association for Computing Machinery. pp. 4125−29 doi: 10.1145/3583780.361516

61. Jiang J, Han C, Zhao WX, Wang J. 2023. PDFormer: propagation delay-aware dynamic long-range transformer for traffic flow prediction. *Proceedings of the AAAI Conference on Artificial Intelligence* 37:4365−73

62. Shao Z, Zhang Z, Wang F, Xu Y. 2022. Pre-training enhanced spatial-temporal graph neural network for multivariate time series forecasting. *KDD '22: Proceedings of the 28th ACM SIGKDD Conference on Knowledge Discovery and Data Minin. Washington DC, USA, August 14−18, 2022*. New York, United States: Association for Computing Machinery. pp. 1567−77 doi: 10.1145/3534678.3539396

63. Gao H, Jiang R, Dong Z, Deng J, Song X. 2023. Spatio-temporal-decoupled masked pre-training for traffic forecasting. *arXiv* 2312.00516

64. Choi J, Choi H, Hwang J, Park N. 2022. Graph neural controlled differential equations for traffic forecasting. *Proceedings of the AAAI Conference on Artificial Intelligence* 36:6367−74

65. Choi J, Park N. 2023. Graph neural rough differential equations for traffic forecasting. *ACM Transactions on Intelligent Systems and Technology* 14:1−27

66. Ye J, Liu Z, Du B, Sun L, Li W, et al. 2022. Learning the evolutionary and multi-scale graph structure for multivariate time series forecasting. *KDD '22: Proceedings of the 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining. Washington DC, USA, August 14−18, 2022*. New York, United States: Association for Computing Machinery. pp. 2296−306 doi: 10.1145/3534678.3539274

67. Paszke A, Gross S, Massa F, Lerer A, Bradbury J, et al. 2019. PyTorch: an imperative style, high-performance deep learning library. *Proceedings of the 33rd International Conference on Neural Information Processing Systems, Vancouver BC, Canada, December 8−14, 2019*. Red Hook, NY, United States: Curran Associates Inc. pp. 8026−37 doi: 10.5555/3454287.3455008

68. Dougherty MS, Kirby HR, Boyle RD. 1993. The use of neural networks to recognise and predict traffic congestion. *Traffic Engineering & Control* 34:311−14

69. Vlahogianni EI, Karlaftis MG, Golias JC. 2005. Optimized and meta-optimized neural networks for short-term traffic flow prediction: a genetic approach. *Transportation Research Part C: Emerging Technologies* 13:211−34

70. Zheng W, Lee DH, Shi Q. 2006. Short-term freeway traffic flow prediction: bayesian combined neural network approach. *Journal of Transportation Engineering* 132:114−21

71. Chan KY, Dillon TS, Singh J, Chang E. 2012. Neural-network-based models for short-term traffic flow forecasting using a hybrid exponential smoothing and Levenberg−Marquardt algorithm. *IEEE Transactions on Intelligent Transportation Systems* 13:644−54

72. Davis GA, Nihan NL. 1991. Nonparametric regression and short-term freeway traffic forecasting. *Journal of Transportation Engineering* 117:178−88

73. Cai P, Wang Y, Lu G, Chen P, Ding C, et al. 2016. A spatiotemporal correlative *k*-nearest neighbor model for short-term traffic multistep forecasting. *Transportation Research Part C: Emerging Technologies* 62:21−34

74. Castro-Neto M, Jeong YS, Jeong MK, Han LD. 2009. Online-SVR for short-term traffic flow prediction under typical and atypical traffic conditions. *Expert Systems with Applications* 36:6164−73

75. Su H, Zhang L, Yu S. Short-term traffic flow prediction based on incremental support vector regression. *Third International Conference on Natural Computation (ICNC 2007), Haikou, China, 24−27 August 2007*. USA: IEEE. pp. 640−45 doi: 10.1109/ICNC.2007.661

76. Sengupta S, Basak S, Saikia P, Paul S, Tsalavoutis V, et al. 2020. A review of deep learning with special emphasis on architectures, applications and recent trends. *Knowledge-Based Systems* 194:105596

77. Guo J, Liu Y, Yang Q, Wang Y, Fang S. 2021. GPS-based citywide traffic congestion forecasting using CNN-RNN and C3D hybrid model. *Transportmetrica A: Transport Science* 17:190−211

78. Guo S, Lin Y, Li S, Chen Z, Wan H. 2019. Deep spatial–temporal 3D convolutional neural networks for traffic data forecasting. *IEEE Transactions on Intelligent Transportation Systems* 20:3913−26

79. Bao Y, Huang J, Shen Q, Cao Y, Ding W, et al. 2023. Spatial–temporal complex graph convolution network for traffic flow prediction. *Engineering Applications of Artificial Intelligence* 121:106044

80. Rahmani S, Baghbani A, Bouguila N, Patterson Z. 2023. Graph neural networks for intelligent transportation systems: a survey. *IEEE Transactions on Intelligent Transportation Systems* 24:8846−85

81. Guo K, Hu Y, Sun Y, Qian S, Gao J, et al. 2021. Hierarchical graph convolution network for traffic forecasting. *Proceedings of the AAAI Conference on Artificial Intelligence* 35:151−59

82. Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, et al. 2017. Attention is all you need. *NIPS'17: Proceedings of the 31st International Conference on Neural Information Processing Systems, Long Beach, California, USA, December 4−9, 2017*. Red Hook, NY, United States: Curran Associates Inc. pp. 6000−10 doi: 10.5555/3295222.3295349

83. Liu Z, Zheng G, Yu Y. 2023. Cross-city few-shot traffic forecasting via traffic pattern bank. *CIKM '23: Proceedings of the 32nd ACM International Conference on Information and Knowledge Management. Birmingham United Kingdom October 21−25, 2023*. New York, United States: Association for Computing Machinery. pp. 1451−60 doi: 10.1145/3583780.361482

84. Sun M, Ding W, Zhang T, Liu Z, Xing M, et al. 2023. STDA-Meta: a meta-learning framework for few-shot traffic prediction. *2023 IEEE 29th International Conference on Parallel and Distributed Systems (ICPADS). Ocean Flower Island, China, 17−21 December 2023*. USA: IEEE. pp. 534−41 doi: 10.1109/ICPADS60453.2023.00085

85. Wang Y, Yu C, Hou J, Chu S, Zhang Y, Zhu Y. 2022. ARIMA model and few-shot learning for vehicle speed time series analysis and prediction. *Computational Intelligence and Neuroscience* 2022:2526821

86. Yuan Y, Shao C, Ding J, Jin D, Li Y. 2024. Spatio-temporal few-shot learning via diffusive neural network generation. *Proc. The Twelfth International Conference on Learning Representations, Vienna, Austria, 2024*. Austria: ICLR. pp. 1−28 https://openreview.net/forum?id=QyFm3D3Tzi

87. Song Y, Wang T, Cai P, Mondal SK, Sahoo JP. 2023. A comprehensive survey of few-shot learning: evolution, applications, challenges, and opportunities. *ACM Computing Surveys* 55:1−40

88. Wang Y, Yao Q, Kwok JT, Ni LM. 2020. Generalizing from a few examples: a survey on few-shot learning. *ACM Computing Surveys (csur)* 53:1−34

89. Iman M, Arabnia HR, Rasheed K. 2023. A review of deep transfer learning and recent advancements. *Technologies* 11:40

90. Chen K, Liang Y, Han J, Feng S, Zhu M, et al. 2024. Semantic-fused multi-granularity cross-city traffic prediction. *Transportation Research Part C: Emerging Technologies* 162:104604

91. Ouyang X, Yang Y, Zhou W, Zhang Y, Wang H, et al. 2024. CityTrans: domain-adversarial training with knowledge transfer for spatio-temporal prediction across cities. *IEEE Transactions on Knowledge and Data Engineering* 36:62−76

92. Li J, Liao C, Hu S, Chen X, Lee DH. 2024. Physics-guided multi-source transfer learning for network-scale traffic flow prediction. *IEEE Transactions on Intelligent Transportation Systems* 25:17533−46

93. Li K, Bai W, Huang S, Tan G, Zhou T, et al. 2024. Lag-related noise shrinkage stacked LSTM network for short-term traffic flow forecasting. *IET Intelligent Transport Systems* 18:244−57

94. Adetiloye T, Awasthi A. 2019. Multimodal big data fusion for traffic congestion prediction. *Multimodal Analytics for Next-Generation Big Data Technologies and Applications*, eds. Seng K, Ang LM, Liew AC, Gao J. Cham: Springer. pp. 319−35 doi: 10.1007/978-3-319-97598-6_13

95. Zhao J, Xie X, Xu X, Sun S. 2017. Multi-view learning overview: Recent progress and new challenges. *Information Fusion* 38:43−54

96. Huang X, Ye Y, Yang X, Xiong L. 2023. Multi-view dynamic graph convolution neural network for traffic flow prediction. *Expert Systems with Applications* 222:119779

97. Du S, Li T, Gong X, Horng S-J. 2020. A hybrid method for traffic flow forecasting using multimodal deep learning. *International journal of computational intelligence systems* 13:85−97

98. Mohammad AA, Al Nawaiseh HM, Alhajyaseen WK, Dias C, Mehran B. 2023. Lane-based analysis of the saturation flow rate considering traffic composition. *Transportation Planning and Technology* 46:653−71

99. Bhaumik KK, Niloy FF, Mahmud S, Woo SS. 2024. STLGRU: spatio-temporal lightweight graph GRU for traffic flow prediction. *Advances in Knowledge Discovery and Data Mining. PAKDD 2024. Lecture Notes in Computer Science*, eds. Yang DN, Xie X, Tseng VS, Pei J, Huang JW, Lin JCW. Singapore: Springer. pp. 288−99 doi: 10.1007/978-981-97-2266-2_23

100. Liu X, Xia Y, Liang Y, Hu J, Wang Y, et al. 2024. Largest: a benchmark dataset for large-scale traffic forecasting. *Proceedings of the 37th International Conference on Neural Information Processing Systems, New Orleans, LA, USA, December 10−16, 2023*. Red Hook, NY, USA: Curran Associates Inc. pp. 75354−71 doi: 10.5555/3666122.3669415

101. Wang J, Jiang J, Jiang W, Li C, Zhao WX. 2021. LibCity: an open library for traffic prediction. *SIGSPATIAL '21: Proceedings of the 29th International Conference on Advances in Geographic Information Systems, Beijing China, November 2−5, 2021*. New York, United States: Association for Computing Machinery. pp. 145−48 doi: 10.1145/3474717.3483923

102. Shao Z, Wang F, Xu Y, Wei W, Yu C, et al. 2024. Exploring progress in multivariate time series forecasting: comprehensive benchmarking and heterogeneity analysis. *IEEE Transactions on Knowledge and Data Engineering* 37:291−305