International Journal of
Network Dynamics and Intelligence (IJNDI)

Scilight

Article

Check for updates

Spatio-Temporal Pattern Discovery Methods in Traffic Data: An Overview

Lei Yang [1], Hao Wu [2,*], and Xin Luo [2,*]

[1] School of Computer Science and Technology, Chongqing University of Posts and Telecommunications, Chongqing 400065, China
[2] College of Computer and Information Science, Southwest University, Chongqing 400715, China
* Correspondence: haowuf@swu.edu.cn (H.W.); luoxin@swu.edu.cn (X.L.)
Received: 8 August 2025; Revised: 24 September 2025; Accepted: 8 June 2026; Published: 17 June 2026

Abstract: As distributed sensor systems expand within extensive urban transportation networks, the substantial volume of traffic data has attracted broad interest from both industrial and academic communities. With the availability of traffic data, the spatio-temporal pattern discovery in traffic data (STPDT) has emerged as a prominent topic. It explores how spatial traffic features evolve over time within a low-dimensional space, thereby discovering desired patterns demonstrating the road network's normal and anomalous traffic states. Based on a thorough investigation into the state-of-the-art STPDT, this survey undertakes the following efforts: (a) categorizing recent advancements in STPDT approaches; (b) providing widely used baseline traffic datasets; (c) comparing the performance of different commonly adopted models on several real-world public datasets; and (d) identifying unique research opportunities and future directions for STPDT. In doing so, this survey seeks to deliver an in-depth and systematic review of current STPDT methods from the perspectives of temporal and spatial dependencies, thereby facilitating future research on this emerging and vital issue.

Keywords: spatio-temporal pattern discovery (STPD); intelligent transportation system; matrix decomposition; tensor decomposition; graph neural network; neural network

1. Introduction

The evolution of modern cities towards smart cities is underway. Rapid urbanization and population growth exert mounting pressure on traffic management within cities. Intelligent transportation systems (ITSs) are thus becoming increasingly essential for managing urban environments effectively. They provide a diverse array of transportation services, such as path optimization, traffic estimation, congestion management, and vehicle dispatching.

Therefore, transportation studies have evolved into advanced intelligent systems known as ITS and have become a major focus of research since the 1970s. While several recent surveys have provided valuable insights into the development of ITS from various perspectives, our work differs in focus and scope. For instance, Gong et al. [1] comprehensively review the integration of edge intelligence within ITS, and Kaffash et al. [2] concentrate on big data-driven models and their applications in ITS. Rahmani et al. [3] focus specifically on the evolution and applications of graph neural networks (GNNs), while Lin et al. [4] investigate the utilization of generative adversarial networks (GANs) in transportation scenarios. Haydari et al. [5] emphasize deep reinforcement learning techniques, particularly in traffic signal control.

While existing surveys have explored the role of specific techniques in ITS, such as edge computing [1], big data [2], graph neural networks [3], and deep reinforcement learning [5], these works tend to focus on individual algorithm families or broader ITS applications without providing a unified view of STPD techniques. In contrast, our work takes a comprehensive and method-oriented perspective by systematically reviewing the diverse families of STPDT that underpin a wide range of ITS tasks. We categorize existing approaches into three major families: low-rank decomposition(a more recent development in capturing higher-order correlations), various deep neural networks, and other emerging methods. As far as we are aware, this is the first survey to offer a unified, structured overview on STPTD in the ITS domain, offering both conceptual clarity and practical insights for future research.

With recent advancements in sensing technologies, spatio-temporal traffic data-including flow, speed, and occupancy-is now collected through various sources including loop detectors, microwave detectors, and GPS probes. The availability of such data presents novel opportunities to analyze non-linear and complex spatio-temporal patterns. The significance of analyzing the traffic data is emphasized for several key reasons:

(a) Traffic data analysis offers substantial benefits to various subsequent applications, encompassing traffic prediction, regulation, and oversight. For instance, accurate traffic flow prediction can significantly mitigate urban congestion. It can also aid traffic signal control systems in developing optimal control strategies.
(b) A thorough analysis of traffic data enables decision-makers to more effectively understand and identify the typical spatial and temporal variations throughout the full extent of road networks. The characterization of traffic dynamics further provides empirical insights essential for long-term traffic planning.

Traffic data provides plentiful spatio-temporal information on traffic states. Therefore, spatio-temporal pattern discovery in traffic data (STPDT) is a crucial issue in traffic data analysis. Over the past few years, STPD methods have demonstrated high efficiency in extracting rich information from data by modeling spatio-temporal dependencies, thus facilitating various tasks, such as imputation and prediction. Overall, STPD is widely adopted for analyzing diverse types of data, including photovoltaic power data, 3D skeleton-based human poses data, and fluid dynamic data.

Traditional STPD methods encompass neighborhood-based methods and regression techniques, such as SVR and KNN.

Gong et al. [6] present a multi-kernel K-Means method with spatial relevance, aiming to discern the latent relationships across diverse views while capturing regional similarities. Subsequently, tensor-based methods were proposed to describe spatio-temporal patterns. Liang et al. [7] present a CP decomposition-based framework that explicitly exploits spatial and temporal correlations by learning from crime data.Chen et al. [8] develop a time-varying reduced-rank vector autoregression (VAR) model that integrates time-varying VAR with tensor factorization to elucidate spatial and temporal patterns within latent spaces. In Ref. [9], a multiscale STGNN is employed to discern the characteristics of human poses across both temporal and spatial dimensions. Karimi et al. [10] leverage spatial and temporal coherence among photovoltaic plants to construct an STGNN for simulating complex non-linear dynamics within a photovoltaic system.

Motivated by the aforementioned successes of STPD, STPDT has garnered widespread attention, leading to a rapidly growing body of related research. However, a review regarding its state-of-the-art remains lacking. This paper presents a comprehensive review of existing STPDT methods. As depicted in Figure 1, the existing STPDT methods are categorized into three branches: a) low-rank decomposition, b) neural networks, and c) other methods. This work aims to make the following contributions:

(1) Summarizing the progress in STPDT by carefully reviewing and categorizing the advanced methods.
(2) Summarizing the typical evaluation metrics and widely adopted datasets for STPDT, along with discussing several empirically validated STPDT methods using four traffic speed datasets to illustrate their performance; and
(3) Discussing the development trends in STPDT.

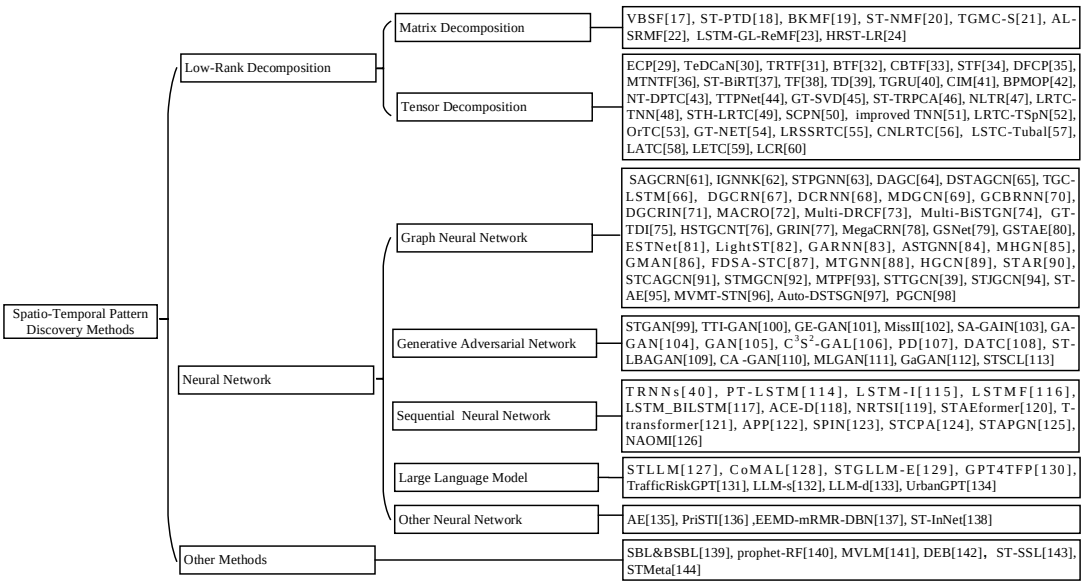


Figure 1. Classification of spatio-temporal pattern discovery methods.

Section 2 provides a detailed background. Section 3 comprehensively reviews state-of-the-art STPDT methods. Section 4 summarizes the typical evaluation metrics and datasets for STPDT and presents empirical studies. Section 5 explores future research directions and potential applications of STPDT. Finally, Section 6 presents the conclusions.

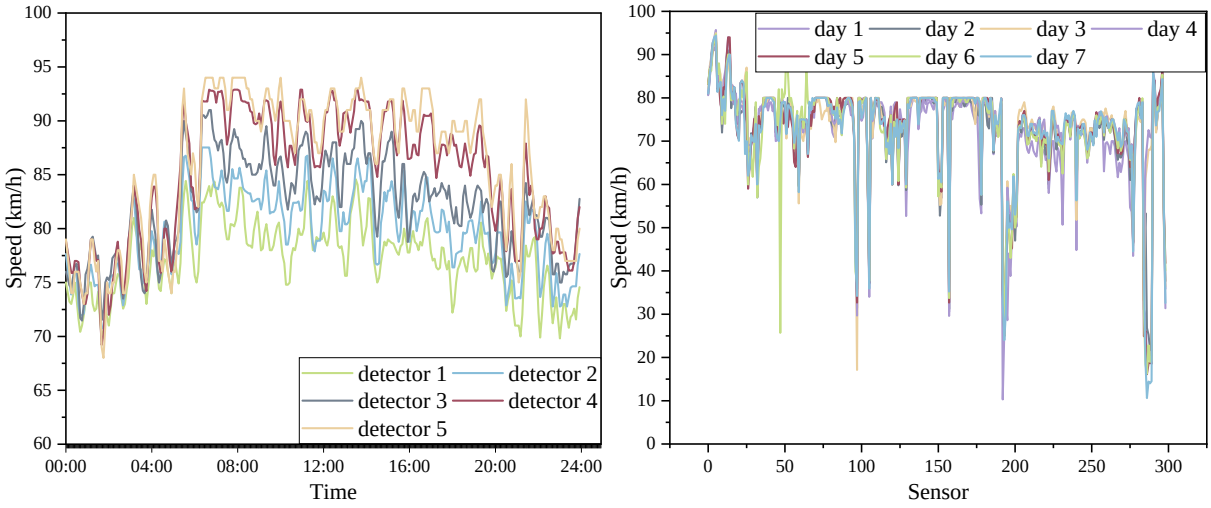
2. Background

This section covers: (1) the introduction to traffic spatiotemporal pattern; and (2) the existing spatio-temporal pattern discovery methods.

2.1. Traffic Spatio-Temporal Pattern

As a representative form of spatiotemporal data, network-wide traffic state data, collected at various times and locations, exhibit complex interrelations. From a spatial perspective, traffic data at one detector are often influenced by those collected at nearby detectors, leading to similarities in traffic state data collected from adjacent detectors. Additionally, a strong correlation can be drawn between the spatial patterns of two non-neighboring detectors situated in functionally similarly regions. For instance, it is unsurprising that two commercial areas, which are far apart, exhibit high similarity in their traffic states. From a temporal viewpoint, traffic data have both short-term proximity and weekly periodicity.

The former refers to the tendency of consecutive days to exhibit similar traffic patterns, while the latter refers to the similarity in traffic patterns observed on the same day of consecutive weeks, primarily because people tend to follow similar routines, such as commuting on weekdays and resting on weekends. Figure 2 provides a straightforward illustration of Beijing speed data, where the multi-dimensional similarities in the traffic spatiotemporal data are clearly visible.



(a) Similarities along the 'location' dimension **(b)** Similarities along the 'day' dimension

Figure 2. Correlation analysis from traffic speed on detectors, days, respectively.

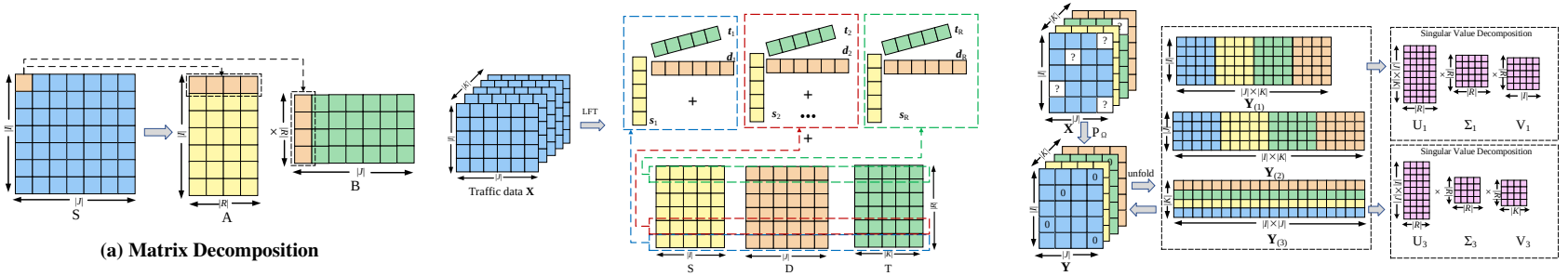
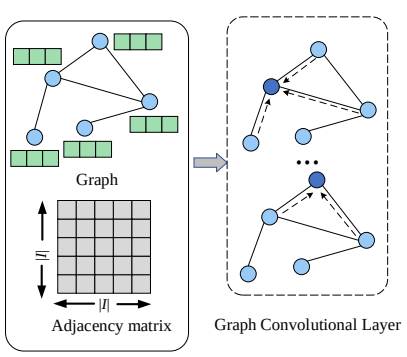
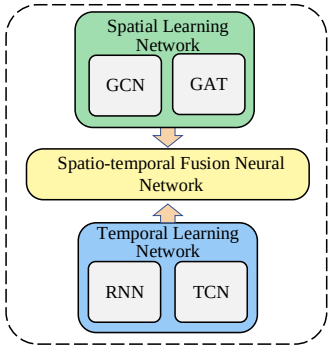
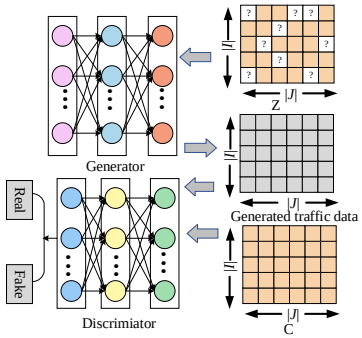
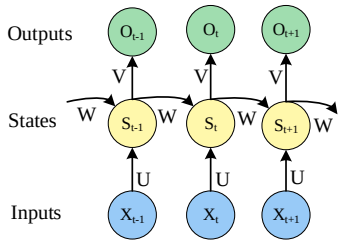
2.2. Spatio-Temporal Pattern Discovery

Recently, research on STPD methods has emerged, highlighting the significance of incorporating spatio-temporal patterns into models to enhance task performance. Yuan et al. [11] present an activity simulation architecture founded on generative adversarial imitation learning, which captures the underlying spatio-temporal variations by modeling activity trajectories using neural differential equations. Luo et al. [12] propose a spatio-temporal attention-enhanced location recommendation model that utilizes self-attention layers along the trajectory to incorporate relative spatio-temporal information from all trajectories. Recent transformer-based solutions have demonstrated significant success in STPD. Tang et al. [13] present a spatio-temporal crisscross attention mechanism that simultaneously models spatial and temporal information. Yan et al. [14] introduce a tracking architecture based on an encoder-decoder transformer, where the encoder models spatio-temporal relationships connecting target objects to search areas, and the decoder forecasts the locations of the target entities. Zhang et al. [15] propose a mixed spatio-temporal encoder that incorporates a temporal transformer for modeling temporal motion and a spatial transformer for learning spatial correlation.

3. STPDT Methods

We review advanced STPDT methods. Furthermore, in Table 1, we summarize the core attributes and classification of existing STPDT methods, and discuss their pros and cons.

Table 1. Summary of STPDT architectures.

Low-Rank Decomposition				
Architecture	 **(a) Matrix Decomposition** **(b) Tensor Decomposition**			
Description	It decomposes the traffic state matrix into a temporal and a spatial pattern matrix.	It considers spatio-temporal traffic data as a three-way tensor, which is then decomposed into a low-rank approximate tensor using either rank minimization or Canonical Polyadic decomposition [16].		
Characteristics	Pros. It effectively utilizes the low-rank structure. Cons. It fails to fully capture the spatiotemporal information.	Pros. (a) It better leverages global patterns spanning various dimensions. (b) It offers excellent interpretability, rapid convergence, high precision, and efficient storage usage. Cons. It models temporal dynamics solely from a numerical perspective.		
Neural Network				
Architecture	 **(c) Graph Convolutional Network**	 **(d) Spatio-Temporal Graph Neural Network**	 **(e) Generative Adversarial Network**	 **(f) Sequential Neural Network**
Description	Node representations are generated by gathering features from the target node as well as its neighboring nodes.	Spatial and temporal learning networks are integrated through merging techniques.	It comprises a generator and a discriminator.	An RNN processes the input from the current time step and the output from the previous time step.
Characteristics	Pros. GCN is employed to model non-Euclidean spatial structures. Cons. It overlooks temporal correlations in STPDT.	Pros. It effectively captures both spatial and temporal interdependencies. Cons. It demands substantial datasets and computational resources.	Pros. GANs do not strictly limit model selection. Cons. They lack support for inference networks.	Pros. It effectively captures temporal patterns. Cons. RNNs struggle to retain information from earlier time steps.

3.1. Low-Rank Decomposition

3.1.1. Matrix Decomposition

As shown in Table 1a, the initial traffic matrix S is decomposed into two latent factor matrices A and B of rank *R*, satisfying S=AB. Hence, $A^{|I|\times|R|} \in \mathbb{R}$ represents a temporal pattern matrix reflecting the global state's evolution over time, and $B^{|R|\times|J|} \in \mathbb{R}$ denotes a spatial pattern matrix that encodes intrinsic characteristics of road connections. Paliwal et al. [17] propose a variational Bayesian subspace filtering (VBSF) approach for estimating traffic density, where the traffic matrix is specifically treated as a low-rank subspace, and parameters are learned through variational Bayesian formalism. Wang et al. [18] propose a mixed traffic flow recovery model named ST-PTD, which uses time-series analysis to uncover periodic trends and then introduces a bidirectional autoregressive matrix factorization to capture trend information. A frequently employed method for temporal data analysis is the Gaussian process. Lei et al. [19] present a Bayesian kernelized matrix factorization (BKMF) framework to complete traffic values, which introduces Gaussian process kernel regularization into the matrix factorization framework to capture temporal dependencies.Wang et al. [20] introduce a spatio-temporal non-negative matrix factorization (ST-NMF)-based approach for analyzing road networks, incorporating graph Laplacian and Toeplitz matrices as regularization terms for spatial and temporal aspects. Zhang et al. [21] propose a temporal geometric matrix completion with speed information (TGMC-S) model that integrates weighted Laplacian regularization and Toeplitz regularization to estimate traffic flow. Sure et al. [22] develop the augmented lagrangian sparsity regularized matrix factorization (AL-SRMF) and constrained low rank (C-LR) matrix completion approaches for traffic matrix reconstruction, which incorporates a graph adjacency matrix and a Toeplitz matrix as spatial and temporal regularization. Yang et al. [23] propose a spatio-temporal regularized matrix factorization (LSTM-GL-ReMF) model, which replaces an autoregressive temporal regularization term with a long short-term memory unit and also includes a graph Laplacian spatial regularization term. Xu et al. [24] propose a spatio-temporal low rank (HRST-LR) technique with Hessian regularization to impute traffic data, where temporal and spatial regularization terms are built using a Toeplitz matrix and a Hessian matrix.

3.1.2. Tensor Decomposition

As shown in Table 1b, traffic spatiotemporal data collected by *I* sensors at *K* time intervals over *J* days can be represented as a tensor $\mathbf{X}^{|I|\times|J|\times|K|} \in \mathbb{R}$, and assume its rank is *R*. Compared to other popular tensor decomposition methods, the CANDECOMP/PARAFAC (CP) approach is frequently employed to perform STPDT. It expresses the tensor by representing it as a sum of multiple rank-one tensors [25–27], $\mathbf{X}$'s rank-*R* approximation $\hat{\mathbf{X}}$ can be achieved as:

$$\hat{\mathbf{X}} = \sum_{i=1}^{R} \boldsymbol{s}_r \circ \boldsymbol{d}_r \circ \boldsymbol{t}_r, \tag{1}$$

where $\circ$ denotes the outer product [28]. The factor matrices are derived as combinations of vectors from the rank-one tensors: $S^{|I|\times|R|} = [\boldsymbol{s}_1, \boldsymbol{s}_2, \cdots, \boldsymbol{s}_R] \in \mathbb{R}$, $D^{|J|\times|R|} = [\boldsymbol{d}_1, \boldsymbol{d}_2, \cdots \boldsymbol{d}_R] \in \mathbb{R}$, and $T^{|K|\times|R|} = [\boldsymbol{t}_1, \boldsymbol{t}_2, \cdots, \boldsymbol{t}_R] \in \mathbb{R}$. The first two factor matrices, S and D, analyze the spatial patterns across the road networks, while T analyzes the temporal patterns among the temporal embeddings.

The fundamental principle of low-rank spatio-temporal modeling lies in the integration of regularization terms that consider spatio-temporal structures, enhancing the basic low-rank framework. Numerous studies leverage graph-based regularization to represent local spatio-temporal uniformity. A prominent method for capturing spatial structural attributes is through the graph Laplacian matrix. For temporal consistency, a widely used technique to enforce temporal smoothness in traffic data is to use a Toeplitz matrix for temporal regularization. For example, Said et al. [29] present a metropolitan and time-sensitive enhanced CP (ECP) completion model, where the urban context similarity matrix and Toeplitz matrix are incorporated into the CP completion. Bhanu et al. [30] introduce a tensor decomposition method with characteristic network (TeDCaN) constraints for traffic volume imputation, applying a graph Laplacian for spatio-temporal regularization. Baggag et al. [31] present a temporally regularized tensor factorization (TRTF) framework for estimating missing values, which incorporates not only a graph Laplacian spatial regularization term but also an autoregressive temporal regularization term for tensor decomposition. Chen et al. [32] present a Bayesian temporal factorization (BTF) model for imputing missing values, incorporating a vector autoregressive process on the temporal factor matrix to capture temporal dependencies. Yang et al. [33] propose a collaborative Bayesian tensor factorization (CBTF) method for traffic speed prediction that shares a common factor matrix for collaboration. Ahn et al. [34] propose a streaming tensor factorization (STF) approach for missing entry prediction, which leverages attention-based temporal regularization and an online update method to identify temporal patterns. Xing et al. [35] propose a data fusion-based CANDECOMP/PARAFAC (DFCP)

tensor factorization model that constructs a four-way tensor through the integration of vehicle data along with cellphone data to impute missing network volume data. Zhu et al. [36] introduce a multi-task neural tensor factorization (MTNTF) approach to jointly impute traffic speed and volume, utilizing a multi-layer neural network in place of the CP factorization's inner product, effectively capturing the non-linear relationships among shared attributes. Li et al. [37] present a spatio-temporal bi-directional residual optimization (ST-BiRT) framework based on tensor factorization for imputing incomplete traffic data, which combines tensor decomposition with a residual optimization technology to extract spatio-temporal dependencies. Guo et al. [38] developed an approach using a tensor decomposition (TF) algorithm aimed at addressing angle-range estimation challenges in radar systems.

In addition to CP decomposition, other tensor decomposition models, such as tucker decomposition, have been widely used to exploit spatio-temporal correlations in traffic patterns. Li et al. [39] employ Tucker decomposition (TD) to reconstruct the adjacency tensor, which enables the encoding of richer and more global spatio-temporal dependencies. Wu et al. [40] propose a tensor-based GRU (TGRU) that applies Tucker decomposition to compress its weight parameters for traffic flow prediction. Jia et al. [41] present a traffic congestion imputation model (CIM) on a congestion-level tensor that factorizes the day-time interval matrix and the sensor-time interval matrix, incorporating local constraints into the matrix factorization process. Luo et al. [42] establish a Bayesian probabilistic matrix outer product (BPMOP) model for traffic data imputation, introducing an outer product operation involving three matrices, resulting in a rank-1 tensor for tensor decomposition. Chen et al. [43] develop a non-negative temporal dimension-preserving tensor completion (NT-DPTC) framework, which decomposes a tensor into three latent factor tensors and employs a Sigmoid function to ensure the non-negative properties of traffic data. Shen et al. [44] put forward a neural network named TTPNet to estimate travel time, which integrates tensor factorization and graph embedding. Deng et al. [45] propose a graph-tensor singular value decomposition (GT-SVD) model for traffic data imputation, which utilizes the graph Fourier transform to capture spatial features and employs Toeplitz matrices as temporal regularization terms. Feng et al. [46] propose a spatio-temporal tensor robust principal component analysis (ST-TRPCA) method to reconstruct traffic data despite corrupted or missing observations, utilizing a Laplacian matrix for spatial regularization and a Toeplitz matrix for temporal constraint.

To address the challenge of specifying the optimal rank beforehand, rank-minimization-based approaches that employ convex surrogates have been widely applied to SPTD. As shown in Table 1b, the formulation for Rank Minimization Tensor Completion (RMTC) in mathematical terms is presented as follows:

$$\min_{\mathbf{Y}} \operatorname{rank}(\mathbf{Y}), s.t. P_{\Omega}(\mathbf{X}) = P_{\Omega}(\mathbf{Y}), \tag{2}$$

where rank(.) denotes the rank, $\mathbf{X}^{|I|\times|J|\times|K|} \in \mathbb{R}$ represents the original tensor, $\mathbf{Y}^{|I|\times|J|\times|K|} \in \mathbb{R}$ is the estimated tensor to be determined, Ω represents the set of observed entries, and the operator P_{Ω} denotes the orthogonal projection onto Ω, i.e.,

$$[P_{\Omega}(\mathbf{X})]_{ijk} = \begin{cases} x_{ijk}, \text{if } (i,j,k) \in \Omega. \\ 0, \text{otherwise.} \end{cases} \tag{3}$$

The rank minimization problem described above is classified as an NP-hard problem. To address this challenge, many researchers have replaced the rank minimization function with the nuclear norm, which is mathematically expressed as:

$$\min_{\mathbf{Y}} \sum_{k=1}^{3} \alpha_k \left\| \mathbf{Y}_{(k)} \right\|_* \text{s.t.} P_{\Omega}(\mathbf{Y}) = P_{\Omega}(\mathbf{X}), \tag{4}$$

where α_k represents the weight for the *k*-th unfolded matrix of tensor $\mathbf{Y}$ with *k*=1,2,3, and the nuclear norm $||.||_*$ is formulated as:

$$\left\| \mathbf{Y}_{(k)} \right\|_* = \sum\nolimits_{n=1}^{r} \sigma_n \left(\mathbf{Y}_{(k)} \right), \tag{5}$$

where $\sigma_n \left(\mathbf{Y}_{(k)} \right)$ denotes the *n*-th largest singular value of matrix $\mathbf{Y}_{(k)}$.

Unlike decomposition-based methods, nuclear norm-based models avoid computing explicit factor matrices and instead leverage the nuclear norm and its variants to characterize the tensor's low-rank property. Chen et al. [47] introduce a nonconvex low-tubal-rank (NLTR) tensor completion model with temporal regularization to impute missing data by exploiting the underlying physical properties of spatio-temporal traffic data, including similarity and periodicity. According to Chen et al. [48], propose a framework named LRTC-TNN, which utilizes truncated nuclear norm minimization to impute spatio-temporal traffic data. Wang et al. [49] propose a spatio-temporal Hankel low-rank tensor completion (STH-LRTC) method to estimate traffic speed, which approximates the tensor rank through a truncated nuclear norm. Hu et al. [50] develop a tensor completion approach leveraging the tensor

Schatten capped p norm (termed SCPN)'s p-th power as the objective function for recovering traffic data. Dai et al. [51] propose a tensor completion approach based on nuclear norm (improved TNN) that utilizes a similar block matrix technique along with an optimal rank estimation algorithm for imputing traffic flow. Nie et al. [52] define a rank minimization tensor completion framework, known as LRTC-TSpN, designed for imputing traffic volume. This framework is built upon a nonconvex truncated Schatten p-norm as the rank function and employs the alternating direction method of multipliers for optimization. Wang et al. [53] propose an outlier-robust tensor completion (OrTC) model to estimate traffic data, which recovers the traffic data by minimizing the anomaly tensor's $L_{2,1}$-norm and the noise tensor's L_F-norm.

By integrating spatio-temporal constraints, many scalable tensor learning models based on the RMTC framework have been developed. Deng et al. [54] develop a graph-tensor neural network (GT-NET) designed for imputing network traffic, which models traffic data as a low-tubal rank graph-tensor and constructs two Laplacian tensors for temporal regularization. Chen et al. [55] propose a low-rank sparse self-representation tensor completion (LRSSRTC) model, which leverages both rank minimization tensor completion and sparse self-representation to estimate missing values within a cohesive framework. Chen et al. [56] introduce a collaborative nonconvex low-rank tensor completion (CNLRTC) model, which incorporates elastic net self-representation with an auto-regression approach to capture both sample similarity and time series dependency. Chen et al. [57] develop a low-tubal-rank smoothing tensor completion (LSTC-Tubal) framework to recover traffic data, where day-to-day correlations are preserved using a unitary transform matrix. Chen et al. [58] propose a low-rank autoregressive tensor completion (LATC) algorithm for traffic data, which imposes an autoregressive process on each time series to impute traffic data. Nie et al. [59] propose a Laplacian enhanced rank minimization tensor completion (LETC) model for traffic speed estimation, which utilizes temporal graph Fourier transform to encode the temporal dimension, and employs a random-walk Laplacian for spatial regularization. Chen et al. [60] proposed a low-rank Laplacian representation (LCR) model by incorporating a Laplacian Kernel into the circulant matrix nuclear norm for traffic time series imputation. Table 1 summarizes some recent representative low-rank decomposition approaches.

3.2. Neural Network

3.2.1. Graph Neural Network

A Graph Neural Network (GNN)-based model synthesizes node features together with those of neighboring nodes to generate updated representations. As shown in Table 1c,d, the commonly adopted GNN-based methods in STPDT may be classified into two main groups: (i) Graph Convolutional Network (GCN)-based methods as shown in Table 1c; and (ii) Spatio-Temporal GNN (STGNN)-based methods as shown in Table 1d.

Graph Convolutional Network

As shown in Table 1c, a GCN aggregates the representations of neighboring nodes to produce an intermediate representation for the node. In formal terms, GCN's graph convolution operation Gconv(.) can be formulated as follows:

$$\text{Gconv}\left(\mathrm{X}_t, \mathrm{A}; \mathrm{W}\right) = \mathrm{W}\left(\mathrm{I}_N + \mathrm{D}^{-\frac{1}{2}}\mathrm{A}\mathrm{D}^{-\frac{1}{2}}\right)\mathrm{X}_t, \tag{6}$$

where I_N is the identity matrix of size N, $\mathrm{X}_t^{|N|\times|d|} \in \mathbb{R}$ denotes the feature matrix of the graph at time step t, N is the number of nodes, d is the dimension of the node feature vector, A represents the adjacency matrix, D corresponds to the degree matrix, while W denotes a weight matrix that can be trained.

Zhang et al. [61] propose a self-attention graph convolutional residual network (SAGCRN) to impute traffic data, integrating a GCN and a multi-head self-attention mechanism to model both local topology structure and context information. Wu et al. [62] develop an inductive GNN kriging (IGNNK) model, framing traffic data estimation as a kriging problem and utilizing a trained GNN to perform kriging for recovering unsampled sensor data. Kong et al. [63] present a spatio-temporal pivotal GNN (STPGNN), which introduces a pivotal graph convolution module designed to extract the spatio-temporal features centered around key nodes. Fang et al. [64] propose a dilated attentional graph convolution (DAGC) that enhances the graph convolution operator with a flexible receptive field and develops a multi-source spatio-temporal network (MS-Net) aimed at forecasting traffic flow. Zheng et al. [65] present a dynamic spatial-temporal adjacent GCN (DSTAGCN) that applies graph convolution to the reconstructed spatial-temporal graph for learning spatial-temporal features.

Spatio-Temporal Graph Neural Network

As shown in Table 1d, spatial learning networks leverage GCNs and Graph Attention Networks (GATs), while temporal learning networks employ Recurrent Neural Networks (RNNs) and Temporal Convolutional Networks

(TCNs). STGNNs can be broadly categorized into two groups: RNN-based and CNN-based approaches.

(1) RNN-based methods:

As illustrated in Figure 3, RNN-based methods filter inputs and hidden states through graph convolutions, enabling them to capture spatio-temporal dependencies in recurrent units. To exemplify, a basic RNN can be represented as follows:

$$\mathrm{S}_t = \sigma\left(\mathrm{U}\mathrm{X}_t + \mathrm{W}\mathrm{S}_{t-1} + \boldsymbol{b}\right), \tag{7}$$

where $\mathrm{X}_t^{|N|\times|d|} \in \mathbb{R}$ represents the feature matrix at time step t, and $\mathrm{S}_t^{|N|\times|d|} \in \mathbb{R}$ denotes the hidden feature matrix of nodes at time step t. Following the application of graph convolution, formula (7) transforms into the following form:

$$\mathrm{S}_t = \sigma\left(\mathrm{Gconv}\left(\mathrm{X}_t, \mathrm{A}; \mathrm{U}\right) + \mathrm{Gconv}\left(\mathrm{S}_{t-1}, \mathrm{A}; \mathrm{W}\right) + \boldsymbol{b}\right), \tag{8}$$

where Gconv(.) denotes a graph convolutional layer.

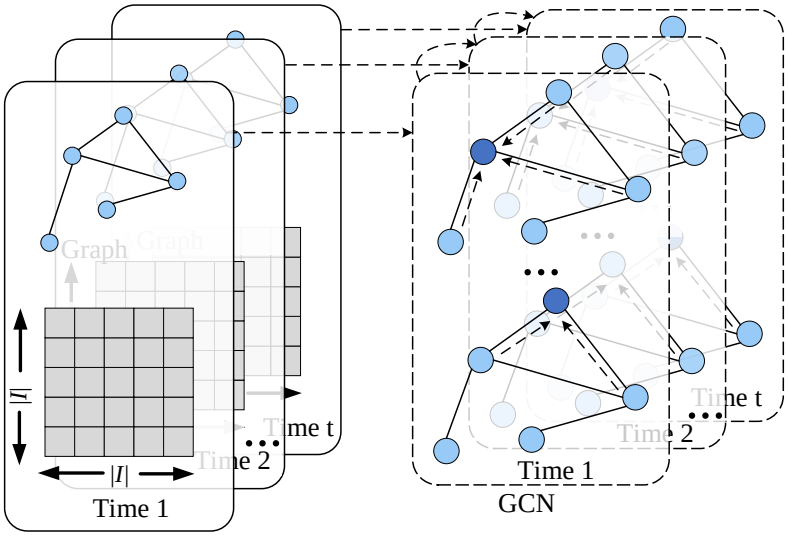


Figure 3. RNN-based STGNNs.

Cui et al. [66] propose a traffic graph convolutional long short-term memory neural network (TGC-LSTM) model, which utilizes a specifically designed traffic graph convolution to extract features and adopts LSTM to predict the traffic dynamics. Li et al. [67] propose a dynamic graph convolution recurrent network (DGCRN) traffic prediction framework, in which the dynamic adjacency matrix is constructed by a hyper-network that synchronizes with the iterations of the RNN. Li et al. [68] present a diffusion convolutional RNN (DCRNN) which employs diffusion convolution and a sequence-based learning architecture to identify spatio-temporal patterns. Liang et al. [69] propose a memory-augmented dynamic GCN (MDGCN) to impute traffic data, where each spatio-temporal block comprises a bidirectional LSTM for extracting temporal information and a dynamic GCN for extracting spatial information. Zhang et al. [70] propose a graph convolutional bidirectional RNN (GCBRNN) for estimating missing entries, which utilizes a GCN and a 1×1 convolution module to learn spatio-temporal correlations. Kong et al. [71] design a dynamic graph convolutional recurrent imputation network (DGCRIN) to impute missing traffic data, which adopts dynamic graph convolution and gate recurrent unit (GRU) to discover spatial and temporal patterns. Ming et al. [72] propose a multi-graph convolutional recurrent network (MACRO) framework to estimate traffic flow, where each LSTM leverages spatial features as input and utilizes an auto-encoder to generate temporal information. Li et al. [73] present a multi-state deep residual collaboration learning algorithm (Multi-DRCF) for traffic data completion, which integrates GCN with LSTM to characterize spatio-temporal interactions. Wang et al. [74] propose a multi-view bidirectional spatio-temporal graph network (Multi-BiSTGN), which fuses the imputation outcomes across three temporal correlation perspectives via a parametric matrix for traffic data imputation. Zhang et al. [75] present the graph transformer traffic data imputation (GT-TDI) framwork, which utilizes GNN to extract spatial features and employs Transformer architectures to model temporal dynamics. Huo et al. [76] propose a hierarchical traffic flow prediction network (HSTGCNT) that combines Transformer and STGNNs to model temporal and spatial patterns. Cini et al. [77] present a graph recurrent imputation network (GRIN) to impute multivariate time sequences, which utilizes a graph-based recurrent neural architecture to analyze spatio-temporal patterns. Jiang et al. [78] propose a meta-graph convolutional recurrent network (MegaCRN) framework by integrating a meta-graph learner into a

graph convolutional recurrent unit for traffic speed prediction. Wang et al. [79] propose a geographical and semantic spatial-temporal network(GSNet) architecture for predicting traffic accident risk, which utilizes multiple graph convolutions, a GRU and an attention mechanism to identify spatio-temporal dependencies. Wang et al. [80] propose a graph-structured spatio-temporal autoencoder (GSTAE) model for anticipating traffic velocities, integrating a GCN and a GRU to discover the intrinsic spatio-temporal dynamics within traffic information. Luo et al. [81] propose an embedded spatio-temporal network (ESTNet) to model traffic flow characteristics, which employs a multi-range GCN to learn static properties and a GRU to learn dynamic properties. Hu et al. [82] propose a simplified spatio-temporal GNN (LightST) for traffic flow forecasting, which introduces stacked temporal linear layers and an adaptive embedding-based graph convolution method to extract temporal and spatial dependencies.

To consider the role of adjacent nodes in capturing spatial features, as shown in Figure 4, a GAT incorporates an attention mechanism within the operation of node aggregation. The input is a set of node features, denoted as $\mathrm{H} =< \boldsymbol{h}_1, \boldsymbol{h}_2, \cdots, \boldsymbol{h}_N >$, where $\boldsymbol{h}_i^{|d|} \in \mathbb{R}$, N represents the number of nodes, and d is the dimension of the node feature vector. This layer generates a new set of node features, $\mathrm{H}' =< \boldsymbol{h}_1', \boldsymbol{h}_2', \cdots, \boldsymbol{h}_N' >$, as output. The attention coefficient is computed using the following equations:

$$\boldsymbol{h}_i^{'} = \sigma\left(\sum_{u \in N(i)} \alpha_{ij} \mathrm{W} \boldsymbol{h}_j\right), \tag{9}$$

$$\alpha_{ij} = \frac{\exp\left(\mathrm{ReLu}\left(\boldsymbol{a}^T\left[\mathrm{W}\boldsymbol{h}_i^{(t-1)} \parallel \mathrm{W}\boldsymbol{h}_j^{(t-1)}\right]\right)\right)}{\sum_{k \in N(i)} \exp\left(\mathrm{ReLu}\left(\boldsymbol{a}^T\left[\mathrm{W}\boldsymbol{h}_i^{(t-1)} \parallel \mathrm{W}\boldsymbol{h}_k^{(t-1)}\right]\right)\right)}. \tag{10}$$

where $N(i)$ is the neighbor node set of a node i, α_{ij} is the attention score from neighboring node j to the central node i, $\mathrm{W}^{|d'|\times|d|} \in \mathbb{R}$ denotes the weight matrix, and $\boldsymbol{a}^{|2d'|} \in \mathbb{R}$ is the weight vector. Shen et al. [83] develop a graph attention RNN (GARNN) to impute traffic data, utilizing LSTM and a GAT to learn temporal and spatial features. Guo et al. [84] introduce an attention-based STGNN (ASTGNN) for traffic forecasting, incorporating multi-head self-attention to analyze temporal patterns and dynamic GCNs to process spatial features. Zhiwen et al. [85] present a multi-view heterogeneous graph network (MHGN) model for missing value imputation, where both spatial and temporal patterns are viewed as graphical structures, and a heterogeneous GAT is developed to analyze spatio-temporal patterns. Zheng et al. [86] propose a graph multi-attention network (GMAN) for predicting traffic conditions, which combines spatio-temporal attention mechanisms combined with gated fusion to effectively model spatio-temporal relationships. Duan et al. [87] propose a fully dynamic self-attention spatio-temporal graph (FDSA-STG) network, utilizing spatio-temporal GAT to capture spatio-temporal correlations within transportation networks.

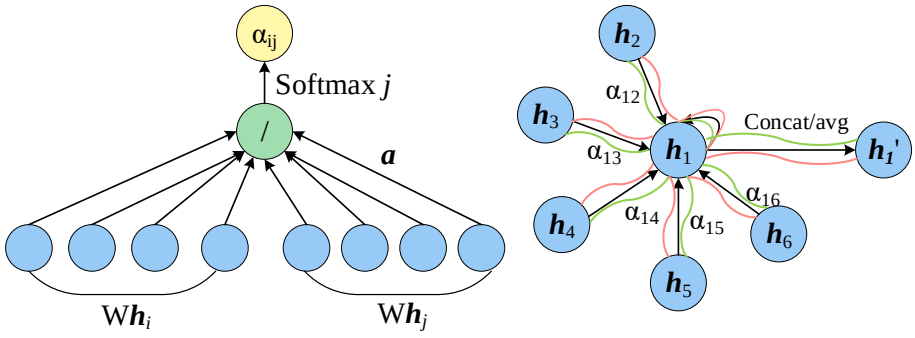


Figure 4. The structure of GAT.

(2) CNN-based methods

As shown in Figure 5, approaches based on CNN combine 1-D-CNN layers and graph convolutional layers to capture temporal and spatial patterns, respectively. Here, consider an STGNN input tensor $\mathbf{X}^{|T|\times|N|\times|d|} \in \mathbb{R}$. The 1-D-CNN layer processes $\mathbf{X}$[:,i,:] along the time dimension to collect temporal data for each node, while the graph convolutional layer acts on $\mathbf{X}$[i,:,:] to gather spatial information across time intervals. The gated TCN integrates a gated mechanism with a pure 1D-CNN architecture, enhancing the temporal learning capability. For temporal features from the input sequence $\mathbf{X}$ are extracted as follows:

$$\mathbf{H}_t = \tanh\left(\mathbf{X} \star \mathbf{\Gamma}_1\right) \odot \sigma\left(\mathbf{X} \star \mathbf{\Gamma}_2\right), \tag{11}$$

where $\mathbf{\Gamma}_1$ and $\mathbf{\Gamma}_2$ denote the trainable parameters for convolution kernels within two separate 1D-CNNs, the symbol $\star$ indicates the convolution operation, while $\odot$ signifies element-wise multiplication.

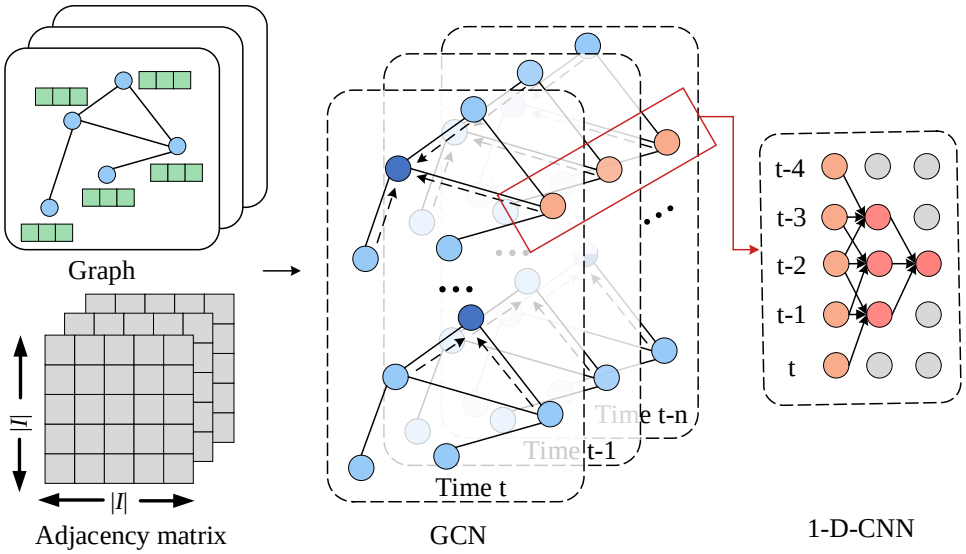


Figure 5. CNN-based STGNNs.

Wu et al. [88] present a multivariate time GNN (MTGNN) framework to predict traffic speed, which utilizes graph convolution and temporal convolution modules to learn spatio-temporal features within the time series. Guo et al. [89] propose a hierarchical GCN (HGCN) tailored for traffic prediction, incorporating gated temporal convolutions, spatial gated graph convolutions, along with a temporal attention module. This design allows it to capture local temporal patterns, identify local spatio-temporal interactions, and model global temporal dependencies. Liang et al. [90] present a spatio-temporal aware data recovery network (STAR) for Cooperative-ITS data imputation, which employs a residual gated TCN as the temporal pattern analysis module, incorporating GCNs and an attention mechanism for spatial pattern analysis. Nie et al. [91] propose a spatio-temporal correlation-adaptive GCN (STCAGCN) model, which integrates a GAT and a gated TCN to estimate traffic volume. Liang et al. [92] present a spatio-temporal multi-GCN (STMGCN)-based method to forecast traffic flow, which utilizes spatial multi-graph convolutional layers and temporal gated convolutional layers to extract spatio-temporal traffic flow patterns.

Furthermore, in order to enhance the extraction of longer-range temporal features, it is feasible to utilize a Causal TCN incorporating dilated factors. As illustrated in Figure 6, for a time series $\boldsymbol{x}^{|T|} \in \mathbb{R}$, the operation of dilated convolution, with a $1 \times k$ kernel filter f and dilated factors d, can be formally represented as:

$$\boldsymbol{z}(s) = \boldsymbol{x} *_d f(s) = \sum_{i=0}^{k-1} f(i) \cdot \boldsymbol{x}(s - d \cdot i), \tag{12}$$

where $\boldsymbol{z}(s)$ denotes the time step at s within the compressed sequence z.

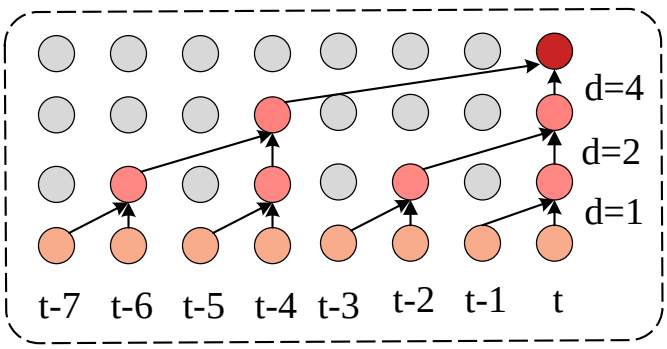


Figure 6. Dilated convolutions.

Qu et al. [93] present a multi-task pretraining and fine-tuning (MTPF) model, in which temporal and spatial dependency modeling is achieved through dilated causal convolution and enhanced adaptive graph convolution. Li

et al. [39] develop a spatial-temporal tensor GCN model (STTGCN), which integrates a tensor decomposition-based graph reconstruction method and dilated convolution modules to learn spatio-temporal correlations for traffic estimation. Zheng et al. [94] present a spatio-temporal joint GCN (STJGCN) which employs a dilated causal spatio-temporal GCN on the spatio-temporal graph to forecast traffic. Liu et al. [95] propose a spatio-temporal autoencoder (ST-AE) for traffic flow prediction that employs TCN and GCN to respectively capture temporal patterns and structural relationships. Wang et al. [96] propose a multi-view multi-task spatio-temporal network (MVMT-STN) architecture to forecast traffic accident risks which utilizes both CNN and GCN to extract spatial and semantic features. Jin et al. [97] present an automated dilated spatio-temporal synchronous graph network (Auto-DSTSGN) for traffic forecast that incorporates a spatio-temporal synchronous GCN and a dilated TCN to extract spatio-temporal features. Shin et al. [98] propose a progressive GCN (PGCN) which adopts a progressive GCN to capture spatial dependencies and uses dilated causal convolution to extract temporal features. Table II summarizes some recent representative GNN-based approaches.

3.2.2. Generative Adversarial Network

As illustrated in Table 1e, a GAN includes two main parts: the generator (G) and the discriminator (D). When provided with an incomplete matrix, the generator outputs a fully imputed matrix without missing values, while the discriminator evaluates whether the imputed matrix is genuine or artificial. The overall training process is formulated as a two-player min-max game, using the objective function outlined below:

$$\begin{aligned} \mathrm{J}^{\mathrm{D}} &= -\mathrm{E}_{C\sim \mathrm{P}} \log \mathrm{D}\left(\mathrm{C}\right) - E_{Z\sim \mathrm{N}} \log\left(1-\mathrm{D}\left(\mathrm{G}\left(\mathrm{Z}\right)\right)\right), \\ \mathrm{J}^{\mathrm{G}} &= -\mathrm{E}_{Z\sim \mathrm{N}} \log \mathrm{D}\left(\mathrm{G}\left(\mathrm{Z}\right)\right), \end{aligned} \tag{13}$$

where G and D denote the generator and discriminator, with J^{D} and J^{G} as their corresponding objective functions. The expectation function is represented by E. The generator takes the incomplete traffic data Z as input, and the discriminator is trained with the complete matrix C, which represents the real data. Lastly, P and Z denote sets of complete and incomplete traffic data matrices, respectively.

Yuan et al. [99] develop a spatio-temporal GAN (STGAN) method for imputing missing traffic data, which employs center loss and dilated convolutions to account for the spatio-temporal correlations among missing entries. Zhang et al. [100] propose a travel times imputation GAN (TTI-GAN) that incorporates a Skip-Gram neural network model to derive road semantic vectors. Xu et al. [101] propose a deep learning architecture based on a graph embedding-GAN (GE-GAN), in which Deep Walk is used to obtain the adjacency traffic state matrix of road networks, thereby aiding the GAN in generating road traffic state data. Hou et al. [102] propose a spatial interaction theory-guided traffic data interpolation model (Miss II) that estimates the traffic flow between two points of interest based on the gravity model. Zhang et al. [103] propose a self-attention generative adversarial imputation network (SA-GAIN) framework to estimate traffic flow, which uses the self-attention mechanism to identify relationships among sensors distributed across different locations. Xu et al. [104] present a graph aggregate GAN (GA-GAN) method for imputing traffic speed, which reconstructs the road networks based on the correlation coefficients of historical road data and aggregates the features of neighboring nodes by graph sample aggregation. Xiao et al. [105] present a GAN-based approach that models missingness patterns and synthesizes data approximating true values, effectively imputing missing counts in a bike-sharing system. Li et al. [106] develop a self-supervised generative adversarial learning framework with conditional cyclical constraints ($\mathrm{C}^3\mathrm{S}^2$-GAL) aimed at imputing missing portions of traffic data. Chen et al. [107] propose an approach that uses parallel data (PD) and GAN to impute traffic flow, where the original and corrupted data are treated as latent codes, and a representation loss is applied to optimize the GAN. Xie et al. [108] propose a deep adversarial tensor completion (DATC) framework that utilizes a 3D convolutional auto-encoder as the generator and a 3D convolutional encoder-decoder as the discriminator to capture non-linear correlations. Yang et al. [109] propose a spatio-temporal learnable bidirectional attention GAN (ST-LBAGAN) for imputing traffic data, where a learnable bidirectional attention map and a multi-channel matrix are employed to capture the spatial and temporal correlations. Han et al. [110] develop a content-aware GAN (CA-GAN) method for estimating traffic speed, which models traffic data as several structured tensors and employs a weighted loss function for GAN training. Zhang et al. [111] present a dynamic multi-level GAN (MLGAN) traffic data interpolation framework that integrates a temporal-spatial feature extractor, consisting of a bidirectional GRU and GCN, into the generator. Zhang et al. [112] develop a traffic speed interpolation model based on a gated attentional GAN (GaGAN), which incorporates spatial and temporal correlation modules into the generator, and utilizes two sub-discriminators to enhance the quality of imputations. Qin et al. [113] develop a spatio-temporal self-interested coalitional learning (STSCL) model to impute the traffic speed, which introduces SCL to foster collaboration between a reconstructor and a discriminator.

3.2.3. Sequential Neural Network

As shown in Table 1f, a sequential neural network is adopted for modeling non-linear temporal dependencies. The algorithmic process of the RNN is expressed in the following formula:

$$\mathrm{O}_t = \mathrm{g}\left(\mathrm{V}\mathrm{H}_t\right), \tag{14}$$

$$\mathrm{H}_t = \mathrm{f}\left(\mathrm{U}\mathrm{X}_t + \mathrm{W}\mathrm{H}_{t-1}\right), \tag{15}$$

where g(.) and f(.) are activation functions, and H_t corresponds to S_t in Table 1f. Due to the issues of vanishing gradients and slow computation in RNNs, researchers have proposed long short-term memory (LSTM)-based approaches for STPDT.

As shown in Figure 7, the forget gate f_t, input gate i_t, the output gate o_t, and the input cell state $\overline{\mathrm{C}}_\mathrm{t}$ at time step *t* can be described as follows:

$$\mathrm{f}_t = \sigma\left(\mathrm{W}_f \cdot \mathrm{X}_t + \mathrm{U}_f \cdot \mathrm{H}_{t-1} + \boldsymbol{b}_f\right), \tag{16}$$

$$\mathrm{i}_t = \sigma\left(\mathrm{W}_i \cdot \mathrm{X}_t + \mathrm{U}_i \cdot \mathrm{H}_{t-1} + \boldsymbol{b}_i\right), \tag{17}$$

$$\mathrm{o}_t = \sigma\left(\mathrm{W}_o \cdot \mathrm{X}_t + \mathrm{U}_o \cdot \mathrm{H}_{t-1} + \boldsymbol{b}_o\right), \tag{18}$$

$$\bar{C}_t = \tanh\left(\mathrm{W}_c \cdot \mathrm{X}_t + \mathrm{U}_c \cdot \mathrm{H}_{t-1} + \boldsymbol{b}_c\right), \tag{19}$$

where . denotes matrix multiplication. The weight matrices $\mathrm{W}_f, \mathrm{W}_i, \mathrm{W}_o$ and $\mathrm{W}_c \in \mathbb{R}^{|N|\times|N|}$ project the input to the three gates and the cell state. Additionally, $\mathrm{U}_f, \mathrm{U}_i, \mathrm{U}_o$ and $\mathrm{U}_c \in \mathbb{R}^{|N|\times|N|}$ represent the weight matrices for the preceding hidden state. Bias vectors are denoted as $\boldsymbol{b}_f, \boldsymbol{b}_i, \boldsymbol{b}_o$ and $\boldsymbol{b}_c$. The gate activation function, represented by σ, is commonly a sigmoid function, while tanh denotes the hyperbolic tangent function.

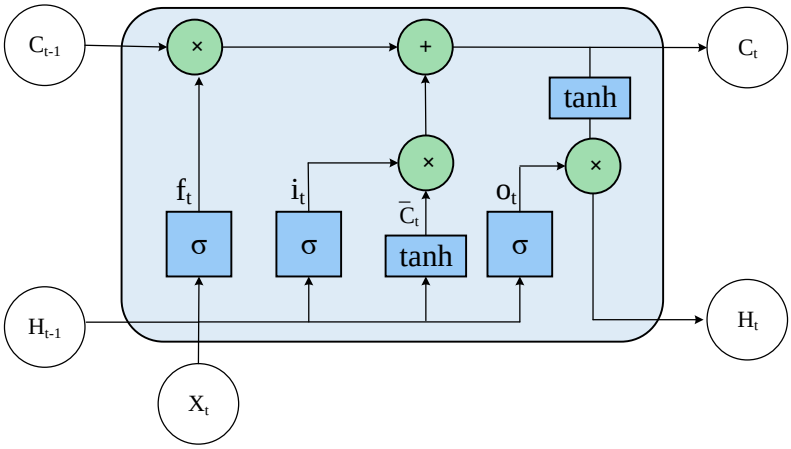


Figure 7. The architecture of LSTM.

The calculation of the ultimate cell state and hidden state proceeds as follows:

$$\mathrm{C}_t = \mathrm{f}_t \times \mathrm{C}_{t-1} + \mathrm{i}_t \times \bar{C}_t, \tag{20}$$

$$\mathrm{H}_t = \mathrm{o}_t \times \tanh\left(\mathrm{C}_t\right) \tag{21}$$

Wu et al. [40] introduce a sequence of tensor-enhanced RNN architectures(TRNNs), which consist of the tensor-based vanilla RNN, tensor-based LSTM, and tensor-based GRU. They further introduce two compact models: TT-based GRU and Tucker-based GRU, for traffic flow prediction. Kwon et al. [114] propose an algorithm for imputing traffic speed data based on a parameter- transferred LSTM (PT-LSTM) architecture, where parameters are transferred from one LSTM layer to its adjacent LSTM to capture spatial features. Cui et al. [115] propose an LSTM-based data imputation method (LSTM-I) that estimates current time step missing values by using earlier LSTM cell and hidden states. Know et al. [116] propose an LSTM-based spatio-temporal data imputation algorithm (LSTMF), which factorizes LSTM model parameters into irreducible components and transfers parameters among these phases to reduce the excessive number of parameters. Ma et al. [117] propose a model (LSTM_BILSTM) for predicting traffic flow on city streets, leveraging time series analysis alongside LSTM optimization. Wang et al. [118] develop an attention calibration encoder-decoder (ACE-D) model, which utilizes LSTM to capture temporal patterns and seq2seq for traffic flow estimation.

Many researchers have adopted the Transformer architecture to capture temporal dependencies in time series data. The Transformer employs an attention mechanism and fully connected neural networks to model sequences. The multi-head attention mechanism is a core component of Transformers, where the basic operation is the scaled dot-product attention, defined as follows:

$$\text{Attention}\left(\mathrm{Q},\mathrm{K},\mathrm{V}\right)=\text{softmax}\left(\frac{\mathrm{QK}^{\mathrm{T}}}{\sqrt{d_{model}}}\right)\mathrm{V}, \tag{22}$$

where Q, K, V and d_{model} stand for the query, key, and value matrices, and key dimension, respectively. As shown in Figure 8, the multi-head attention mechanism divides these matrices into several groups, then applies the attention function (Equation (22)) in parallel. The outputs from each group are concatenated and projected, leading to the final result as follows:

$$\text{MultiHead}\left(\mathrm{Q},\mathrm{K},\mathrm{V}\right)=\text{Concat}\left(\text{head}_1,...,\text{head}_h\right)\mathrm{W}^o, \tag{23}$$

$$\text{head}_j=\text{Attention}\left(\mathrm{QW}_j^{\mathrm{Q}},\mathrm{KW}_j^{\mathrm{K}},\mathrm{VW}_j^{\mathrm{V}}\right), \tag{24}$$

where h represents the number of attention heads, $\mathrm{W}_j^{\mathrm{Q}}$, $\mathrm{W}_j^{\mathrm{K}}$, $\mathrm{W}_j^{\mathrm{V}}$ as the projection matrices used for Q, K, V, while W^o is the ultimate output projection matrix.

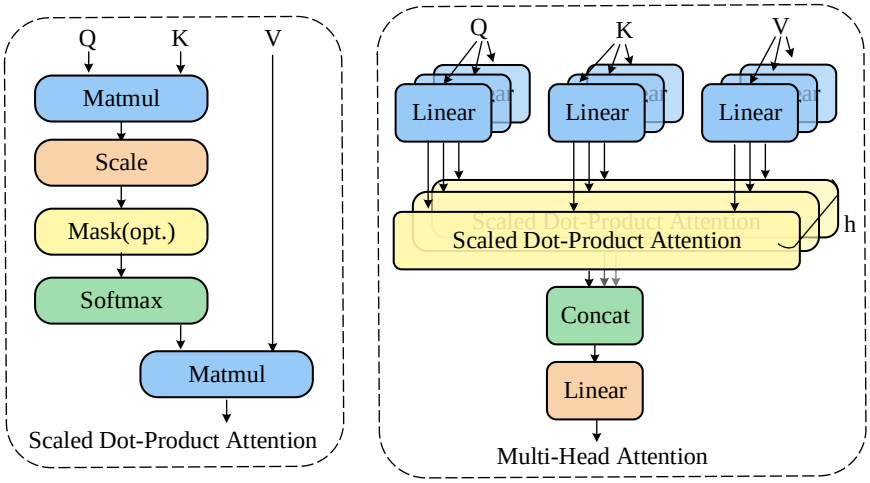


Figure 8. Multi-head Attention.

Shan et al. [119] develop a non-recurrent time series imputation framework (NRTSI) that adopts the self-attention mechanism in transformers to impute data for a series of specified time points. Liu et al. [120] present a spatio-temporal adaptive embedding transformer (STAEformer) that applies vanilla transformers on both the time and spatial axes to model intrinsic spatio-temporal relationships. Yan et al. [121] propose a traffic transformer (T-transformer) algorithm to forecast traffic flow by modeling traffic as a dynamic hierarchical graph. Zhu et al. [122] propose a spatio-temporal attention point process (APP) model that integrates point processes with an attention mechanism to model traffic congestion events. Marisca et al. [123] propose the spatio-temporal point inference network (SPIN) architecture, in which each sparse spatio-temporal attention layer leverages an attention mechanism to capture spatio-temporal correlations. Xu et al. [124] design a spatio-temporal cycle-perceptual attention (STCPA) framework to impute traffic speed, which uses an attention mechanism to analyze spatio-temporal dynamics and a cycle consistency constraint to provide reliable supervisions. Cirstea et al. [125] propose a spatio-temporal aware parameter generation network (STAPGN) to produce model parameters and integrate it with attention mechanisms to forecast traffic time series. Liu et al. [126] present a non-autoregressive multiresolution imputation (NAOMI) deep generative model that employs a divide-and-conquer strategy to iteratively populate the absent values.

3.2.4. Large Language Model

In recent studies, large language models (LLMs) have demonstrated remarkable performance in time series analysis. Liu et al. [127] propose STLLM, a spatio-temporal LLM that learns spatio-temporal representations through embedding construction and integrates them via fusion convolution for traffic forecasting. In an effort to improve traffic efficiency, Yao et al. [128] introduce CoMAL, a collaborative multi-agent LLM-based architecture tailored for mixed-autonomy traffic scenarios. It consists of four key modules: perception, collaboration, reasoning,

and execution. Rong et al. [129] propose STGLLM-E, a spatio-temporal generative large language model designed for edge computing. It integrates spatio-temporal learning with generative LLM, enabling localized, low-latency inference on edge devices. Xu et al. [130] present GPT4TFP, a spatio-temporal fusion model based on LLM for predicting traffic flow. The architecture comprises four components: a spatio-temporal embedding layer, a fusion layer, a frozen pre-trained LLM layer, and a output linear layer. Zhong et al. [131] proposed a TrafficRiskGPT-based framework designed for reasoning about traffic risk. It is build upon the LLaMA3-8B architecture and has been fine-tuned using Low-Rank Adaptation with extensive traffic risk datasets. Hu et al. [132] design a LLM-selectable misbehavior detection framework to verify the authenticity of traffic signs and the accuracy of vehicles motion. Movahedi et al. [133] propose an LLM-driven agent framework for adaptive traffic management, integrating a Zero-Shot Chain-of-Thought approach with a Generally Capable Agent architecture. To tackle the challenge of data scarcity in traffic flow prediction, Li et al. [134] propose UrbanGPT , which combines a spatio-temporal dependency encoding mechanism with instruction-tuning strategies to enhance model performance.

3.2.5. Other Neural Network

Hu et al. [135] introduce an auto-encoder (AE) model for traffic information imputation, leveraging prior traffic state knowledge generated by a preceding Markov model as input to the AE. Liu et al. [136] introduce a conditional diffusion model called PriSTI, tailored for spatio-temporal completion tasks. This framework incorporates a module for conditional feature extraction and a module for noise estimation, enhancing the prior modeling capabilities for accurately imputing traffic speed. Liu et al. [137] propose an ensemble approach based on a deep belief network termed EEMD-mRMR-DBN, which integrates empirical mode decomposition, feature selection with minimal redundancy maximal relevance (mRMR), and a deep belief network to estimate traffic flow. Dai et al. [138] present a deep spatio-temporal inception network (ST-InNet) for collective traffic flow estimation, employing dual inception networks to analyze spatial and temporal dynamics.

3.3. Other Methods

Babu et al. [139] evaluate two relevance vector machine-based approaches, specifically sparse Bayesian learning (SBL) and block SBL, in which appropriate kernel matrix formulations more effectively capture spatio-temporal correlations. To tackle the problem of incomplete traffic data, Li et al. [140] propose a prophet-random forest (prophet-RF) framework that integrates the time-series decomposition method "prophet" for temporal imputation and an iterative random forest method for capturing spatial residual information. Li et al. [141] propose a multi-view learning method (MVLM), which employs LSTM to model time series data, support vector regression to capture spatial dependencies, and collaborative filtering techniques to model interactions between sensors. Kaur et al. [142] introduce the dual-staged error-corrected boosting (DEB) predictor, an error correction technique employed in gradient boosting ensembles, aimed at completing missing records in traffic observations. Ji et al. [143] propose a spatio-temporal self-supervised learning (ST-SSL) traffic prediction algorithm that integrates spatio-temporal convolutions to encode spatio-temporal traffic features. Wang et al. [144] design a spatio-temporal meta-framework (ST Meta) capable of flexibly integrating temporal and spatial patterns identified by a spatio-temporal analytic framework.

3.4. Summary

We summarize the progress in STPDT research, comprehensively reviewing and categorizing advanced techniques. The key characteristics of existing studies, including proposed models, tasks, datasets, and evaluation metrics, are summarized in Tables 2–4. STPDT models are predominantly associated with tensor decomposition and spatio-temporal GNNs.

Given the limitations of current STPDT methods,

(a) They primarily employ attention mechanisms and spatio-temporal regularization terms to capture spatio-temporal patterns, while other auxiliary methods are relatively scarce and warrant further investigation.
(b) Most existing STPDT studies are conducted within the intuition-friendly Euclidean domain, characterized by its flatness and zero curvature. However, these approaches may not offer the most meaningful geometric representations for traffic data in non-Euclidean spaces [145]. Thus, there is an urgent need to advance research on non-Euclidean spaces, such as hyperbolic space, for STPDT.

Table 2. Summary of low-rank decomposition methods.

Ref.	Model	Task	Baseline	Metric	Dataset
[17]	VBSF	speed imputation	KNN, GPR, VAR, LSTM, LRTC	MRE	Delhi speed
[18]	ST-PTD	flow imputation	ST-IDW, ST-KNN, ST-2SMR, ST-ISE, TRMF, BTMF, BGCP, LRTC_TNN	MAE, RMSE	AVI Data (volume), GPS Data (speed)
[19]	BKMF	speed imputation	GSMF, GLTL, BPTF, TRMF, GPFA	MAE, RMSE	Seattle speed, METR-LA speed
[20]	ST-NMF	speed imputation	SRSVD, SRMF, LMaFit, TAS-LR, NMF	NMAE, RMSE	Beijing speed, Portland speed
[21]	TGMC-S	volume imputation	SA, GCF, GMC, CGMC	MAPE, SSI	Xuancheng, Anhui volume, Synthetic dataset
[22]	LR-MC	speed imputation	Baseline approximation, Baseline ODS, ODS with neighbor, SRSVD, SRMF, STGM, Ensemble correlation, TASLR	NMSE, RMSE	PeMS
[23]	LSTM-GL ReMF	speed imputation	TRMF, BTMF/BTTF, LSTM, GRU-D, SGMN, DCRNN, GCGRNN, STGCN, ASTGCN	MAPE, RMSE	Seattle speed, Metr-LA speed, PeMSD8 speed
[24]	HRST-LR	data imputation	NMF, SRSVD, SRMF, LMaFit, TAS-LR, GRIN, AL-SRMF	MAPE, RMSE, MAE	Guangzhou speed, Seattle speed, Portland volume, Portland occupancy
[34]	STF	tensor stream imputation	SOFIA, Online SCP	NRE, ART	Redar Traffic, Chicago Taxi
[35]	DFCP	volume imputation	HaLRTC, CP-ALS-NR, CP-WOPT-CR	MAPE, RMSE	Cellphone location, License plate
[32]	BTMF, BTTF	data imputation, prediction	BTRMF/BTRTF, TRMF, BPMF/BPTF, BGCP, BATF, HaLRTC	MAPE, RMSE	Guangzhou speed, Hangzhou metro passenger flow, Seattle speed, London Speed, NYC Taxi
[29]	Enhanced CP	flow completion	CP_ALS, CP_ARLS, CP_OPT, CP_WOPT, CP_APR, GAIN, MIDA, MIWAE	RE	Porto Taxi, T-drive
[30]	TeDCaN	volume imputation	DTC, CP-ALS, LE, LSTM, ST-MGCN, HA, Persistence Model, TeDCaN Variations	RMSE, P@K, MAP	Green-Taxi of New York City, Private taxi-data of Thessolaniki City
[31]	TRTF	speed imputation, prediction	LSM-RN, TRMF, CP-WOPT	MAPE, ND, RMSE, NRMSE	Aarhus, Doha
[37]	ST-BiRT	speed imputation	HaLRTC, GAIN, BGCP, LRTC-TNN, FNNTEL	MAPE, RMSE, MAE	Guangzhou speed
[38]	TF	target location estimation	Liu's algorithm, MUSIC, ESPRIT-based, ESPRIT-like, CRB	RMSE_θ, RMSE_r, RMSE	Simulation experiments

Table 2. *Cont.*

Ref.	Model	Task	Baseline	Metric	Dataset
[44]	TTPNet	Travel time prediction	TEMP, XGBoost, DeepTTE, Deep Travel, STDR, TTPNet_S	MAE, MAPE, SR	Beijing taxi trajectories, Shanghai taxi trajectories
[45]	GT-SVD	volume imputation	BATF, FPCA, TenALS, ADMM, HaLRTC	RE, MAE, RMSE, MAPE, R^2	PeMS08 volume, PeMS04 volume
[46]	ST-TRPAC	volume imputation	RPCA, RSTD, TRPCA, IRTPCA, HaLRTC, RTC	RE, MAE, RMSE, MAPE	PeMS04 volume
[41]	CIM	congestion data imputation	NNI, MF, LRTI, TDI, BGCP, MLP, PCNN	RMSE, MAE, MRE	Jinan vehicles passage records
[42]	BPMOP	speed imputation	HaLRTC, BGCP, BATF	MAPE, RMSE	PeMSD7 speed, Guangzhou speed
[43]	NT-DPTC	data imputation	BPMF, HaLRTC, BTRMF, BATF, LSTC	RMSE, NMAE, MAE	Guangzhou speed, Seattle speed, Hangzhou metro passenger flow, Birmingham parking
[36]	MTNTF	speed and volume imputation	KNN, DSAE, HaLRTC, BGCP, NTF, GAIN, ASNTF	RMSE, MAPE, MAE	Shanghai speed, volume
[48]	LRTC-TNN	data imputation	BTMF, TRMF, BGCP, BATF, HaLRTC	MAPE, RMSE	Guangzhou speed, Birmingham parking,Hangzhou passenger flow, Seattle speed
[50]	RSCPN	data recovery	BPMF, BTMF, BGCP, HaLRTC, LRTC-TNN, TRPCA	MAPE, RMSE	PEMSD7 speed, PEMSD8 volume, Seattle speed, Guangzhou speed
[51]	Improved TNN	flow completion	LRCQT, SiLRTC, HaLRTC, TMac-ins, TMac-dec	PSNR, SSIM, FSIM	Vehicle video data
[52]	LRTC-TSpN	data imputation	HaLRTC, LRTC-TNN, TNN-DCT, SpBCD	MAE, RMSE	Xuhui flow, Guangzhou speed, Portland volume, Birmingham parking occupancy
[53]	OrTC-AM	data estimation, anomaly detection	DSTC, Online-SGD, TRPCA, TensorDet, BRCP-TC	ER, SRR, TPR, FPR	Abilene
[54]	GT-NET	data imputation	NTC, Autoencoder, SRMF, TNN-ADMM, GLRTC, CP-ALS, Tucker-ALS	RSE, MAE	Abilene, GÉANT volume
[55]	LRSSRTC	data imputation	LRMC, PPCA, MissForest, MICE, LLS, KSR-EN, HaLRTC	RMSE	Interstate 205 volume, PeMS speed
[56]	CNLRTC	speed recovery	HaLRTC, LRTC-TNN, LSTC-Tubal, LRSSRTC, LATC, CNLRTC	RMSE, MAE, MAPE	PEMS speed, Seattle speed
[57]	LSTC-Tubal	speed imputation	BPMF, BGCP, BATF, HaLRTC, LRTC-TNN, LSTC-DCT	MAPE, RMSE	PeMS-4W speed, PeMS-8W speed, London-1M speed, Guangzhou speed

Table 3. Summary of graph neural network methods.

Ref.	Model	Task	Baseline	Metric	Dataset
[61]	SAGCRN	speed completion	HaLRTC, TNN-DCT, LATC-Tubal, LSTM, GRU, GCRN, STGCN	RMSE, MAE, NMAE	METR-LA speed, PEMS-BAY speed
[62]	IGNNK	speed recovery	kNN, OKriging, KPMF, GLTL	RMSE, MAE, R^2, MAPE	METR-LA speed, Seattle speed, PeMS-Bay speed
[66]	TGC-LSTM	speed prediction	ARIMA, SVR, FNN, LSTM, DiffGRU, Conv+LSTM, SGC+LSTM, LSGC+LSTM	MAE, MAPE, RMSE	Seattle speed, INRIX speed
[63]	STPGNN	flow forecast	ARIMA, DCRNN, GWNet, STSGCN, MTGNN, DMSTGCN, DSTAGNN, TPGNN, SGP	MAE, MAPE, RMSE	PEMS-03 flow, PEMS-04 flow, PEMS-07 flow, PEMS-08 flow, England flow,TaxiBJ, PEMS-BAY speed
[64]	DAGC	flow prediction	HA, GRU, GAT, ChebNet, DCRNN, STGCN, STGCNAction, GSTNet	MAE, MAPE, RMSE	Beijing subway transaction, Beijing bus transaction, Beijing taxi trajectories
[65]	DSTAGCN	traffic prediction	VAR, SVR, LSTM, FC-LSTM, STGCN, DCRNN, T-GCN, STSGCN	MAE, MAPE, RMSE	PEMS-03 flow, PEMS-04 flow, PEMS-07 flow, PEMS-08 flow, METR-LA speed, PEMS-BAY speed
[69]	MDGCN	speed imputation	DAE, BiLSTM, BRITS, CNN-BiLSTM, GACN, STGCN, GWNET, GMAN	MAE, RMSE, MAPE	METR-LA speed, INRIX-SEA speed
[70]	GCBRNN	data imputation, prediction	DA, CP, CP-S, BGCP, LRTC-TNN, BTMF, BLSTM-I	MAE, MAPE	Beijing speed, UTD19 flow
[71]	DGCRIN	speed imputation	HA, BGCP, BTTF, LRTC-TNN, PC-GAIN, BLSTM-I, GCBRNN	MAPE, RMSE, MAE	PeMS08 speed, PeMS04 speed
[72]	MACRO	flow imputation	Mean, ARIMA, KNN, LRDMD, LATC, GRIN, BRITS	MAE, RMSE	Real-world dataset
[73]	Multi-DRCF	speed imputation	GAIN, BTMF, BGCP, HaLRTC, ST-CRMF, DVGAE-GAN, NoTMF, LATC	MAPE, RMSE, MAE	Seattle speed, METR-LA speed
[74]	BiSTGN	flow imputation	HA, SES, ST-KNN, ST2SMR, ST-ISE, TRMF, BTMF, LRTC-TNN, BTTF, BGCP	MAE, RMSE, MAPE	Wuhan speed, Wuhan volume
[75]	GT-TDI	data imputation	KNN, PPCA, LSTC-Tubal, LRTC-TNN, DAE, GCN-GRU	MAAPE, RMSE	PeMSD7 flow, PeMSD7 speed
[76]	HSTGCNT	flow prediction	HA, LSVR, FNN, FC-LSTM, STGCN, DCRNN, GWN, STSGCN, STFGNN	MAE, MAPE, RMSE	PeMS-BAY speed, PeMSD7(M) volume, Beijing passenger flow
[77]	GRIN	speed imputation	MEAN, KNN, MICE, MF, VAR, rGAIN, BRITS, MPGRU	MAE, MSE, MRE	PEMS-BAY speed, METR-LA speed
[78]	MegaCRN	speed prediction	HA, STGCN, DCRNN, GW-Net, STTN, GMAN, MTGNN, StemGNN, AGCRN, CCRNN, GTS, PM-MemNet	MAE, RMSE, MAPE	METR-LA speed, PEMS-BAY speed, EXPY-TKY speed
[79]	GS Net	Traffic accident risk forecast	HA, XGBoost, MLP, GRU, SDCAE, ConvLSTM, Hetero-Conv LSTM, Graph Wave Net	RMSE, Recall, MAP	NYC, Chicago

Table 3. *Cont.*

Ref.	Model	Task	Baseline	Metric	Dataset
[80]	GSTAE	speed imputation, prediction	BRITS, BTMF, BTMF (50) GCBRNN, RIHGCN, DCRNN, ASTGCN, GMAN, FC-GAGA	RMSE, MAE, MAPE	PEMSD7 speed, METR-LA speed
[81]	EST Net	speed prediction	VAR, ARIMA, FNN, FC-LSTM, STGCN, DCRNN, Graph Wave Net, MRA-BGCN, GMAN, STGNN, SLCNN, STSGCN	MAE, RMSE, MAqE	METR-LA speed, PEMS-BAY speed
[83]	GARNN	flow imputation	HA, KNN, BRITS, BGCP, PC-GAIN, Impute RNN	MAPE, RMSE, MAE	PeMS04 flow, PeMS08 flow
[84]	ASTGCN	flow prediction	VAR, SVR, LSTM, DCRNN, STGCN, ASTGCN, Graph WaveNet, STSGCN	MAE, RMSE, MAPE	PEMS03 flow, PEMS04 flow, PEMS07 flow, PEMS08 flow, Hangzhou metro crowd flow
[85]	MHGN	trajectory data imputation	FM, BPTF, TRMF, LSTC, NMF, LRTC-TNN, HaLRTC, GAIN, MIWAE, MCFlow, ST-SCL	MAPE, RMSE	Xi'an taxi dataset, Road network dataset
[86]	GMAN	traffic prediction	ARIMA, SVR, FNN, FC-LSTM, STGCN, DCRNN, Graph Wave Net	MAE, RMSE, MAPE	Xiamen volume, PeMS-Bay speed
[87]	FDSA-STG	flow prediction	HA, VAR, ARIMA, LSTM, GeoMAN, ASTGCN	MAE, RMSE	PeMSD4 flow, speed, occupancy rate; PeMSD8 flow, speed, occupancy rate
[89]	HGCN	speed prediction	HA, ARIMA, LSTM, GUR, GCRN, Gated-STGCN, OGCRNN, OTSGGCN, GWNET	MAE, MAPE, RMSE	Jinan speed, Xian speed
[90]	STAR	speed recovery	Average, 2D-Krige, GCN, IGNNK, STAR	MAE, RMSE, MAPE	METR-LA speed, PEMS-BAY speed
[91]	STCAGCN	volume estimation	KNN, ST-SSL, TGMF-S, IGNNK, STAR, LETC	MAE, RMSE, MAPE, WMAPE	PeMS-Bay volume, PeMS-Bay speed
[92]	STMGCN	flow prediction	HA, SVR, BPNN, LSTM, GRU, T-GCN, DCRNN	RMSE, R^2, MAE, RE	Yangtze river vessel trajectories
[93]	MTPF	flow imputation, prediction	DA, KNN, GRU-I, LSTM-I, GCRINT	MAE, RMSE	PEMS04 flow, PEMS08 flow
[39]	STTGCN	flow prediction	SVR, LSTM, DCRNN, STGCN, ASTGCN(r), Graph Wave Net, STSGCN, STFGNN	MAE, MAPE, RMSE	PEMS03 flow, PEMS04 flow, PEMS07 flow, PEMS08 flow
[94]	STJGCN	traffic prediction	VAR, FC-LSTM, SVR, DCRNN, STGCN, ASTGCN, Graph Wave Net, STSGCN, AGCRN, GMAN, Z-GCNETs	MAE, RMSE, MAPE	PEMS03 flow, PEMS07 flow, PEMS04 flow, speed, occupancy; PEMS08 flow, speed, occupancy; Seattle speed
[95]	ST-AE	flow prediction	XG Boost, FC-LSTM, DCRNN, STGCN, Graph Wave Net, STSGCN, AGCRN, STFGNN	MAE, MAPE, RMSE	PEMS03 flow, PEMS04 flow, PEMS07 flow, PEMS08 flow
[96]	MVMT-STN	traffic accident risk prediction	ARIMA, LSTM, MLP, Conv LSTM, Hetero-Conv LSTM, STGCN, T-GCN, GSNet	RMSE, Acc@L, MAP	NYC accidents, taxi trips, POIs, weathers, road network; Chicago accidents, taxi trips, weathers, road network
[97]	Auto-DSTSG	flow prediction	FC-LSTM, DCRNN, STGCN, Graph Wave Net, ASTGCN, STSGCN, STFGNN, STGODE, AutoSTG	MAE, RMSE, MAPE	PEMS03 flow, PEMS04 flow, PEMS07 flow, PEMS08 flow
[98]	PGCN	traffic prediction	HA, FC-LSTM, DCRNN, Graph WaveNet, TGC-GRU, DMSTGCN, STSGCN, AGCRN, Z-GCNETs	MAE, RMSE, MAPE	PeMS-BAY speed, METR-LA speed, Urban-core speed, Seattle speed, PEMS04 speed, flow, occupancy, PEMS08 speed, flow, occupancy, PEMS07 flow

Table 4. Summary of remaining methods.

Ref.	Model	Task	Baseline	Metric	Dataset
[99]	STGAN	data imputation	GAIN, MICE, TFIPDGAN, ANN, ImageCNN, SpaceGAN	NMAE, RMSE, MAE+	Beijing speed, Beijing subway swipe
[100]	TTI-GAN	travel times imputation	ARIMA, k-NN, PPCA, DSAE	RMSE	Chengdu trajectory
[101]	GE-GAN	traffic state estimation	KNN, BP, LSTM, BGCP, Deeptrend2.0	MAE, RMSE, MAPE	PeMS07 volume, Seattle speed
[102]	MissII	flow imputation	NMF, Tensor-CP, GRUI, GAN-2-Stage, ST-LBAGAN, SA-GAIN, DGCRIN	MAE, RMSE, Cosine Similarity	Beijing taxi traces
[103]	SA-GAIN	flow imputation	HA, KNN, GAIN, SAGAN	MAE, MMD, RMSE	Seattle volume
[104]	GA-GAN	speed imputation	BGCP, DSAE, MIWAE, GAIN	MAE, RMSE, MAPE	PEMS-BAY speed, Seattle speed
[105]	GAN	traffic counts imputation	Historical average, KNN, MissForest, EM, MICE	RMSE	Blue-bikes counts
[106]	C^3S^2-GAL	speed imputation	HA, ARIMA, KNN, IterativeSVD, HaLRTC, BTMF, BGCP, NoTMF, GAIN, DA-GAN	MAPE, RMSE, MAE	Guangzhou speed, Seattle speed
[107]	PD	flow imputation	SVR, HA, DSAE	MAE, RMSE, MRE	PeMS05 flow
[108]	DATC	volume completion	CPals, CPnmu, CPwopt, TKals, NTC, NTM, NTF, CoSTCo	NMAE, NRMSE, KL, NDCG@k	Abilene volume, GÉANT volume
[109]	ST-LBAGAN	flow imputation	AdaBoost Regression, Linear Regression, Ridge Regression, HA, TDI, ST-ResNet, TF-3DNet, Deep STN, MDL, GAN	RMSE, MAE	Beijing trajectory, New York trajectory
[110]	CA-GAN	speed completion	HaLRTC	MRE	PeMS03, 04, 05, 07, 08, 10, 11, SR99, 10, Stanislaus speed
[111]	MLGAN	flow imputation	LSTM, SSIM, DSAE, GAN	RMSE, MAE, R^2, MAPE	PeMS04 flow
[112]	GaGAN	speed imputation	GCN, BTTF, HGC-LSTM, ASTGCN, P-GAN	MAE, RMSE, MAPE	Hangzhou speed
[113]	ST-SCL	speed imputation	GAIN, MIWAE, MCFlow, MF	MSE, RMSE	Jinan taxi trajectories
[40]	T-RNNs	flow prediction	MSE-GRU, Tucker-MSE-GRU, T-GRU, TT-GRU, Tucker-GRU	MAE	Hangzhou metro flow, Kaggle competition volume
[114]	PT-LSTM	speed imputation	CNN-LSTM, Bi-LSTM, GRU, G-LSTM	RMSE	Road side units speed
[115]	LSTM-I	speed prediction	BDLSTM-I+BDLSTM, BGCP	MAE, MAPE, RMSE	Seattle speed, PEMS-BAY speed
[116]	LSTMF	speed imputation	Preg, k-NN, MICE, LSTM, GRU, ConvLSTM, BRITS, autoencoder, BTMF, GL-ReMF, GA-GAN, GCN-M	RMSE, MAE	PeMS-BAY speed, Seattle speed
[117]	LSTM_BILSTM	flow prediction	LSTM, BILSTM, CNN, KNN, ConvLSTM	MAE, RMSE, R^2	Ningbo flow
[118]	ACE-D	flow prediction	SVR, LSTM, Encoder-Decoder, AE-D, ACE-D, ATTEN, EK-NN, Ss2s	RMSE, SMAPE, TRMSE, TSMAPE	PeMS04 flow, PeMS11 flow
[119]	NRTSI	traffic time imputation	Linear, GRUI, KNN, MaskGAN, ANP, CSDI, mTAN, SingleRes, NAOMI	Avg MSE, min MSE	PEMS-SF occupancy rate

Table 4. *Cont.*

Ref.	Model	Task	Baseline	Metric	Dataset
[120]	STAEformer	traffic forecast	HI, GWNet, DCRNN, AGCRN, STGCN, GTS, MTGNN, STNorm, GMAN, PDFormer, STID	MAE, RMSE, MAPE	METR-LA speed, PEMS-BAY speed, PEMS04 flow, PEMS07 flow, PEMS08 flow
[121]	Traffic transformer	traffic prediction	ARIMA, FC-LSTM, DCRNN, STGCN, GWN, GMAN	MAE, MAPE, RMSE	METR-LA speed, Urban-BJ, Ring-BJ
[122]	APP	congestion prediction	LSTM, RMTPP, NHP, SAHP, HP	Log-likelihood, accuracy, MSE	Georgia department of transportation congestion dataset, the 911 calls-for-service dataset, the atlanta traffic network dataset
[123]	SPIN	speed reconstruction	GRIN, SAITS, BRITS, rGAIN, MEAN, KNN, MF, MICE, VAR	MAE	PEMS-BAY speed, METR-LA speed
[124]	STCPA	speed imputation	MI, GAIN, GCWC, GCNPer, MIWAE, MCFlow, ST-SCL, MIRACLE	MSE, RMSE	Chengdu speed, New York speed
[125]	STAPGN	flow prediction	LongFormer, DCRNN, STGCN, STG2Seq, GWN, STSGCN, ASTGNN, STFGNN, EnhanceNet, AGCRIN, meta-LSTM	MAE, MAPE, RMSE	PEMS03 flow, PEMS04 flow, PEMS07 flow, PEMS08 flow
[126]	NAOMI	sequence imputation	Linear, KNN, GRUI, MaskGAN, SingleRes	L2Loss	PEMS-SF occupancy
[135]	AEM	traffic prediction, imputation	BTMF, LATC, LAMC, LCR, CNN, MTSVAE	MSE	PeMSD04 volume, occupancy, speed; PeMSD08 volume, occupancy, speed
[136]	PriSTI	speed imputation	MEAN, DA, KNN, Lin-ITP, KF, MICE, VAR, Kalman, TRMF, BATM, BRITS, GRIN, V-RIN, GP-VAE, rGAIN, CSDI, GRIN	MAE, MSE, CRPS	METR-LA speed, PEMS-BAY speed
[137]	EEMD-mRMR-DBN	flow prediction	ARIMA, BP, SVM, DBN, LSTM, EEMD-BP, EEMD-LSTM, EEMD-DBN	RMSE, MAPE	Portland flow
[138]	ST-InNet	flow prediction	HA, CNN, ST-ResNet, ATFM, T-GCN	RMSE, MAE	Nanjing flow
[139]	SBL, BSBL	speed estimation	SVD, MLR, SBL+K, KBSBL	NMAE	PeMS_I605 speed

4. Metric and Dataset

4.1. Metric

In this chapter, we provide a summary of the commonly adopted evaluation metrics for STPDT models. As illustrated in Tables 2–4, the majority of existing studies concentrate on traffic data imputation and prediction, and the evaluation metrics used in these studies are as follows:

(a) MAPE: The mean absolute percentage error assesses the ratio of the estimation error to the actual traffic state values. A lower MAPE indicates higher accuracy and is calculated as:

$$MAPE = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{y_i - \hat{y}_i}{y_i}\right| \times 100\%, \tag{25}$$

where n represents the quantity of samples, y_i is the ground truth, and $\hat{y}_i$ is the estimated value generated by the learning model.

(b) RMSE: The root mean squared error quantifies the difference between the actual traffic state values and those predicted by the STPDT model, i.e.,

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

(c) MAE: The mean absolute error measures the absolute difference between the actual traffic state values and those predicted by the STPDT model,i.e.,

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i|. \tag{27}$$

4.2. Datasets

The increasing availability of traffic-related datasets offers new perspectives for investigating the STPDT problem. The most commonly used datasets are summarized as follows:

(1) METR-LA [69,136]: This dataset includes traffic speed information gathered from 207 detectors over a period of four months (from 1 March to 30 June 2012) with 5-min intervals in Los Angeles County, USA.
(2) PeMS-BAY [136]: This dataset contains traffic speed information gathered from 325 detectors over a period of six months (from 1 January to 30 June 2017) with 5-min intervals in the San Francisco Bay Area, California.
(3) Seattle Traffic Speed [32,58]: This dataset contains freeway traffic speed data gathered from 323 sensors over the entire year of 2015, with 5-min intervals in Seattle, USA.
(4) Seattle20 [142]: This dataset contains traffic volume data gathered from 400 sensors over a period of two months (from 1 May to 27 June 2011) with 20-second intervals in Washington State.
(5) Beijing Traffic Speed [146]: This dataset includes traffic speed information gathered from 3126 detectors over a period of 75 days (from 12 May to 25 July 2022) with 5-min intervals in Beijing, China.
(6) Guangzhou Traffic Speed [31,57]: This dataset consists of traffic speed information gathered from 214 sensors over a period of two months (61 days, from 1 August to 30 September 2016) with 10-min intervals in Guangzhou, China.
(7) Hangzhou Passenger Flow [31,57]: This dataset offers incoming passenger flow data from 80 subway stations spanning 25 days (from 1 January to 25 January 2019) at 10-min intervals in Hangzhou, China.
(8) Portland Traffic Volume [57]: This dataset contains traffic volume and occupancy data from 1156 loop detectors over a period of 31 days (from 1 January to 31 January 2021) with 15-min intervals in the Portland-Vancouver metropolitan area.
(9) INRIX-SEA [69]: This dataset includes freeway traffic speed information from 745 road links, with a 5-min sampling rate, covering the entire year of 2012 in the downtown Seattle area, USA.
(10) PeMSD4 [83,135]: This dataset contains traffic volume, occupancy, and speed information gathered from 307 detectors over a period of 59 days (from 1 January to 28 February 2018) with 5-min sampling rate in the San Francisco Bay Area.
(11) PeMSD7 [42]: This dataset includes traffic speed data collected from 228 sensors over a period of 44 days (the weekdays of May and June of 2012) with a 5-min time slots per day in the state of California.
(12) PeMSD8 [83,135]: This dataset contains traffic flow, occupancy and speed from 170 sensors over 62 days (from 1 July to 31 August 2016) with 5-min sampling rate in San Bernardino.

(13) Birmingham Parking Occupancy [52]: This dataset records the occupancy of 30 car parks over 77 days (from 4 October to 19 December 2016) at 30-min intervals (8:00 a.m. to 5:00 p.m.) in Birmingham.
(14) Aarhus in Denmark Speed [31]: This dataset measures speed at 449 observation points over 117 days (from 13 February to 9 June 2014) at 5-min intervals in Aarhus, Denmark.
(15) Shanghai Traffic Data [36]: This dataset includes traffic speed and flow data for 18 segments over 28 days (form 2 December to 29 December 2019) at 5-min intervals in the Changning District, Shanghai.
(16) PEMS-SF [119]: This dataset provides occupancy rates from 963 sensors over 15 months (from 2 January 2008 to 30 March 2009) at 10-min intervals in the San Francisco Bay Area.
(17) New York City (NYC) Taxi [30]: This dataset records traffic volume between origins and destinations at various time intervals.
(18) Chicago Taxi [34]: This dataset includes pickup counts between origins and destinations at different hours.
(19) BikeNYC [109]: This dataset contains trip records from 2013 to April 2024.
(20) NGSIM [49]: This dataset provides traffic trajectories from the Next Generation Simulation program on US Highway 101.

These public datasets are divided into speed, volume, occupancy, and trajectories according to different application tasks. Table 5 summarizes the commonly adopted traffic datasets.

Table 5. Summary of traffic datasets.

Application Task	Dataset	Sensors	Time Period	Intervals	Ref.
Speed	METR_LA [1]	207	2012.3.1–2012.6.30	5 min	[136]
	PeMS-BAY [1]	325	2017.1.1–2017.6.30	5 min	[136]
	Seattle [2]	323	2015.1.1–2015.12.31	5 min	[32]
	Beijing [4]	3126	2022.5.12–2022.7.25	5 min	[146]
	Guangzhou [5]	214	2016.8.1–2016.9.30	10 min	[31,57]
	INRIX-SEA [8]	745	2012.1.1–2012.12.31	5 min	[69]
	Aarhus [12]	449	2014.2.13–2014.6.9	5 min	[31]
	Shanghai [13]	18	2019.12.2–2019.12.29	5 min	[36]
Volume	Hangzhou [6]	80	2019.1.1-2019.1.25	10 min	[31,57]
	Portland [7]	1156	2021.1.1–2021.1.31	15 min	[57]
	Seattle20 [3]	400	2011.5.1–2011.6.27	20 s	[142]
	PeMSD7 [9]	883	2017.5.1–2017.8.31	5 min	[42]
	PeMSD4 [10]	307	2018.1.1–2018.2.28	5 min	[83,135]
	PeMSD8 [10]	170	2016.7.1–2016.8.31	5 min	[83,135]
	NYC-taxi [15]	-	2013.1.1–2013.12.31	-	[30]
	Chicago [16]	-	2016.2.1–2016.9.30	-	[34]
Occupancy	Birmingham [11]	30	2016.10.4–2016.12.19	30 min	[52]
	PEMS-SF [14]	963	2008.1.1–2009.3.30	10 min	[119]
Trajectories	BikeNYC [17]	-	2013–2024.4	-	[109]
	NGSIM [18]	-	-	-	[49]

[1,8] https://drive.google.com/drive/folders/1Y0wBUF0vBS9KeG5VjRAJy_gySBJqmA-P
[2] https://github.com/zhiyongc/Seattle-Loop-Data
[3] https://data.transportation.gov/Automobiles/Seattle-20-Second-Freeway/ixg2-6cni
[4] https://drive.google.com/drive/folders/1xqJY4S58Jy60jU3kErcezfILAHzDBKhE
[5] https://zenodo.org/records/1205229 and http://www.openits.cn/openData2/792.jhtml
[6] https://zenodo.org/records/3145404#.Y411xxRBw2wg
[7] https://portal.its.pdx.edu/home
[9] https://github.com/VeritasYin/STGCN_IJCAI-18
[10] https://github.com/guoshnBJTU/ASTGNN/tree/main/data
[11] https://archive.ics.uci.edu/dataset/482/parking+birmingham
[12] http://iot.ee.surrey.ac.uk:8080/datasets.html
[13] https://github.com/sysuits/High-dimensional-traffic-data-analysis
[14] http://archive.ics.uci.edu/ml/datasets/PEMS-SF
[15] https://www.nyc.gov/site/tlc/about/tlc-trip-record-data.page
[15] https://opendata.cityofnewyork.us/
[16] https://www.divvybikes.com/system-data
[16] https://data.cityofchicago.org/
[17] http://s3.amazonaws.com/trip data/index.html
[18] https://www.fhwa.dot.gov/publications/research/operations/07030/index.cfm

4.3. Experiments and Results

This section assesses the success of the different advanced traffic speed imputation and prediction methods. Table 6 shows that M1–M5 are completion models, whereas M6–M9 are prediction models. Table 7 presents the imputation results, and Table 8 provides the prediction results for the next 15 min. M1–M9 all adopt the default hyperparameter settings reported in previous studies.

The experiments are conducted on four real-world traffic speed datasets: PeMS-BAY(D1), METR-LA(D2), SEATTLE (D3) and INRIX-SEA (D4). For this analysis, 70% of the data is primarily dedicated to training, 20% is set aside for testing, and the last 10% is specifically kept for validation.

Table 6. Compared models.

No.	Name	Description
M1	LCR	A LCR model that introduces a Laplacian kernel into the RMTC framework [60].
M2	NT-DPTC	An NT-DPTC model that decomposes a tensor into three latent factor tensors [43].
M3	BTMF	A BTMF model that integrates a low-rank MF and a vector autoregressive (VAR) process [32].
M4	BTTF	A BTTF model that integrates low-rank TF and a VAR process [32].
M5	IGNNK	An IGNNK model in which GNN performs kriging through reconstruction [62].
M6	DCRNN	A DCRNN model that uses diffusion convolution and an RNN to model spatio-temporal dependencies [68].
M7	ASTGNN	An ASTGNN model that combines a self-attention mechanism with a dynamic graph convolution module to capture the temporal and spatial dynamics [84].
M8	DGCRN	A DGCRN model that models spatial and temporal dependencies using a GNN and an RNN, respectively [67].
M9	MTGNN	An MTGNN model that integrates graph learning, graph convolution, and temporal convolution modules in an end-to-end architecture [88].

Table 7. Performance comparison among M1–M5.

No.	Metric	PEMS-BAY	METR-LA	SEATTLE	INRIX-SEA
LCR	RMSE	**1.891**	**4.868**	8.708	**1.846**
	MAE	**1.267**	**3.457**	6.085	1.128
	MAPE	**0.024**	**0.081**	0.199	0.042
NT-DPTC	RMSE	6.018	9.798	8.761	3.262
	MAE	3.335	5.741	5.798	1.582
	MAPE	0.079	0.143	0.187	0.055
BTMF	RMSE	4.0932	7.7056	7.739	2.396
	MAE	2.327	4.614	5.214	**1.085**
	MAPE	0.050	0.112	0.154	**0.038**
BTTF	RMSE	4.147	8.380	**6.358**	2.434
	MAE	2.229	5.547	**3.992**	1.181
	MAPE	0.049	0.140	**0.122**	0.042
IGNNK	RMSE	6.615	10.350	7.064	9.834
	MAE	3.888	6.833	4.446	7.861
	MAPE	0.090	0.197	0.150	0.337

Bold numbers indicate the best accuracy.

Table 8. Performance comparison among M6–M9.

No.	Metric	PEMS-BAY	METR-LA	SEATTLE	INRIX-SEA
DCRNN	RMSE	2.798	5.355	4.817	5.428
	MAE	1.328	2.760	2.873	3.688
	MAPE	0.028	0.071	0.077	0.129
ASTGNN	RMSE	2.818	9.433	4.749	**2.654**
	MAE	1.303	3.400	2.782	**0.929**
	MAPE	0.027	0.074	0.071	**0.035**
DGCRN	RMSE	**2.689**	**5.036**	**4.630**	2.894
	MAE	**1.287**	**2.619**	**2.752**	1.066
	MAPE	**0.027**	**0.066**	**0.072**	0.039
MTGNN	RMSE	2.799	5.192	4.715	2.587
	MAE	1.325	2.688	2.816	0.957
	MAPE	0.028	0.069	0.075	0.035

Bold numbers indicate the best accuracy.

MAPE, RMSE, and MAE are chosen as the evaluation metrics. Tables 7 and 8 compare the performance of STPDT models on D1, D2, D3, and D4, respectively. Table 9 presents a summary of the total time cost across the compared models. Consequently, it is observed that:

(1) For Speed Imputation, the RMTC framework (LCR) achieves the best performance on the PeMS-BAY and METR-LA datasets, while matrix/tensor decomposition methods (BTMF and BTTF) demonstrate superior performance on the SEATTLE and INRIX-SEA datasets. Both approaches consistently outperform the graph neural network-based imputation method (IGNNK): As indicated in Table 7, on dataset D1, the RMSE, MAE and MAPE of M1 are 1.891, 1.267, and 0.024, respectively. These values are 71.41%, 67.41%, and 73.33% lower than M5's corresponding metrics of 6.615, 3.888, and 0.090, respectively.

(2) For Speed Prediction, the results of RNN-based STGNN and CNN-based STGNN are comparable: As presented in Table 8, on dataset PeMS-BAY, the lowest RMSE, MAE, and MAPE of DGCRN are 2.689, 1.287, and 0.027, respectively. These values are 3.93%, 2.87%, and 3.57% lower than MTGNN's corresponding metrics of 2.799, 1.325 and 0.028, respectively.

(3) The neural network model requires significantly more computation time compared to the low-rank decomposition model: As shown in Table 9, the total time costs of IGNNK on datasets PeMS-BAY, METR-LA, SEATTLE, and INRIX-SEA are 11,407, 1832, and 14,549, 3355 respectively, which are 92.21%, 95.69%, 98.35% and 97.44% higher than BTTF's corresponding time costs.

Table 9. Total time cost among M1–M9 (seconds).

No.	PEMS-BAY	METR-LA	SEATTLE	INRIX-SEA
LCR	**260**	**50**	9255	221
NT-DPTC	8693	2792	696	1823
BTMF	149	56	1167	**53**
BTTF	889	79	**240**	86
IGNNK	11,407	1832	14,549	3355
DCRNN	67,854	25,160	69,353	24,530
ASTGNN	88,411	43,293	192,360	35,938
DGCRN	29,367	10,205	47,097	130,055
MTGNN	10,024	2836	8881	1814

Bold numbers indicate the shortest running time.

5. Future Development Trends

Based on the preceding literature review, we outline future development trends in this section.

(1) Spatial dependency models: As demonstrated in Table 2 and Section 3.1 , existing STPDT models predominantly adopt the graph Laplacian as the spatial regularization term. While effective, this approach may be

limited in its ability to capture more complex spatial dependencies. It would therefore be valuable to explore alternative spatial regularization terms, such as adaptive neighbors, total variation, and directed auto-regression. In particular, adaptive neighbors can adaptively construct spatial relationships based on data characteristics, which enables the identification of low-dimensional manifolds embedded in high-dimensional spaces. This property has shown great potential in representation learning for spatio-temporal traffic flow data, and may provide a promising direction for enhancing spatial dependency modeling.

(2) Temporal dependency models: Existing STPDT models predominantly utilize short-term proximity to capture temporal correlations. Future research could benefit from simultaneously exploiting both short-term and long-term proximity, including seasonal trends and weekly periodicity, to enhance model performance.

(3) Automatic parameter tuning: Given the numerous hyper-parameters in the STGNN model, which introduce additional computational burdens during calibration, future research should focus on integrating automatic parameter tuning to streamline the calibration process.

(4) Benchmarking STPDT: Future research will focus on the quantitative evaluation of spatio-temporal pattern characteristics. It is essential to establish a unified standard for evaluating the effectiveness of STPDT and for comparing it with alternative approaches. Additionally, the existing STPDT model primarily analyzes global traffic patterns, leaving the stochastic nature of traffic networks and the impact of anomalous factors underexplored.

(5) Multi-source data: In future studies, we plan to further refine urban traffic prediction by addressing traffic congestion. Additionally, incorporating more external factors (e.g., road features, temperature, adverse weather conditions, and the impact of large-scale events) could enhance the model's prediction accuracy.

6. Conclusions

Due to the fast progress of distributed sensor networks in large-scale urban transportation networks, substantial amounts of traffic data, obtained from loop magnetic detectors and mobile global positioning systems, offer rich spatio-temporal insights into traffic states. The availability of this data has sparked significant interest among researchers, engineers and public agencies in uncovering spatio-temporal patterns to better understand traffic state evolution and enhance urban transportation management. This paper thoroughly reviews the latest research on STPDT. Initially, traffic spatio-temporal patterns and the current state of research in spatio-temporal discovery are presented. Subsequently, the advancements in STPDT are explored, with a comprehensive review, categorization, and discussion of the state-of-the-art. Typical STPDT evaluation metrics and datasets are then summarized. Several prominent STPDT models are empirically validated to demonstrate their performance. Finally, potential opportunities and guidelines for subsequent research are highlighted. This review aims to inspire further research applications of STPDT among researchers and engineers.

Author Contributions: L.Y.: conceptualization, methodology, software, writing—original draft preparation; H.W.: visualization, investigation, supervision; X.L.: writing—reviewing and editing.

Funding: This work is supported in part by the National Natural Science Foundation of China under grants 62302402 and 62272078, in part by the Chongqing Natural Science Foundation under grants CSTB2024TIAD-KPX0018 and CSTB2023NSCO-LZX006, and by the National Key Research and Development Program of China under grant 2024YFF0908200.

Conflicts of Interest: The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Use of AI and AI-Assisted Technologies: No AI tools were utilized for this paper.

References

1. Gong, T.; Zhu, L.; Yu, F.R.; et al. Edge Intelligence in Intelligent Transportation Systems: A Survey. *IEEE Trans. Intell. Transp. Syst.* **2023**, *24*, 8919–8944.
2. Kaffash, S.; Nguyen, A.T.; Zhu, J. Big Data Algorithms and Applications in Intelligent Transportation System: A Review and Bibliometric Analysis. *Int. J. Prod. Econ.* **2021**, *231*, 107868.
3. Rahmani, S.; Baghbani, A.; Bouguila, N.; et al. Graph Neural Networks for Intelligent Transportation Systems: A Survey. *IEEE Trans. Intell. Transp. Syst.* **2023**, *24*, 8846–8885.
4. Lin, H.; Liu, Y.; Li, S.; et al. How Generative Adversarial Networks Promote the Development of Intelligent Transportation Systems: A Survey. *IEEE/CAA J. Autom. Sin.* **2023**, *10*, 1781–1796.
5. Haydari, A.; Yılmaz, Y. Deep Reinforcement Learning for Intelligent Transportation Systems: A Survey. *IEEE Trans. Intell. Transp. Syst.* **2022**, *23*, 11–32.
6. Gong, Y.; Li, Z.; Zhang, J.; et al. Missing Value Imputation for Multi-View Urban Statistical Data via Spatial Correlation

Learning. *IEEE Trans. Knowl. Data Eng.* **2023**, *35*, 686–698.
7. Liang, W.; Cao, J.; Chen, L.; et al. Crime Prediction with Missing Data via Spatiotemporal Regularized Tensor Decomposition. *IEEE Trans. Big Data* **2023**, *9*, 1392–1407.
8. Chen, X.; Zhang, C.; Chen, X.; et al. Discovering Dynamic Patterns from Spatiotemporal Data with Time-Varying Low-Rank Autoregression. *IEEE Trans. Knowl. Data Eng.* **2024**, *36*, 504–517.
9. Li, M.; Chen, S.; Zhao, Y.; et al. Multiscale Spatio-Temporal Graph Neural Networks for 3D Skeleton-Based Motion Prediction. *IEEE Trans. Image Process.* **2021**, *30*, 7760–7775.
10. Karimi, A.M.; Wu, Y.; Koyuturk, M.; et al. Spatiotemporal Graph Neural Network for Performance Prediction of Photovoltaic Power Systems. In Proceedings of the AAAI Conference on Artificial Intelligence, online, 2–9 February 2021; pp. 15323–15330.
11. Yuan, Y.; Ding, J.; Wang, H.; et al. Activity Trajectory Generation via Modeling Spatiotemporal Dynamics. In Proceedings of the 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining, Washington, DC, USA, 14–18 August 2022; pp. 4752–4762.
12. Luo, Y.; Liu, Q.; Liu, Z. STAN: Spatio-Temporal Attention Network for Next Location Recommendation. In Proceedings of the Web Conference 2021, Ljubljana, Slovenia, 19–23 April 2021; pp. 2177–2185.
13. Tang, Z.; Qiu, Z.; Hao, Y.; et al. 3D Human Pose Estimation with Spatio-Temporal Criss-Cross Attention. In Proceedings of the 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Vancouver, BC, Canada, 18–22 June 2023; pp. 4790–4799.
14. Yan, B.; Peng, H.; Fu, J.; et al. Learning Spatio-Temporal Transformer for Visual Tracking. In Proceedings of the IEEE/CVF International Conference on Computer Vision, online, 11–17 October 2021; pp. 10448–10457.
15. Zhang, J.; Tu, Z.; Yang, J.; et al. MixSTE: Seq2seq Mixed Spatio-Temporal Encoder for 3D Human Pose Estimation in Video. In Proceedings of the 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), New Orleans, Lousiana, 19–24 June 2022; pp. 13232–13242.
16. Wu, H.; Luo, X.; Zhou, M.; et al. A PID-Incorporated Latent Factorization of Tensors Approach to Dynamically Weighted Directed Network Analysis. *IEEE/CAA J. Autom. Sin.* **2022**, *9*, 533–546.
17. Paliwal, C.; Bhatt, U.; Biyani, P.; et al. Traffic Estimation and Prediction via Online Variational Bayesian Subspace Filtering. *IEEE Trans. Intell. Transp. Syst.* **2022**, *23*, 4674–4684.
18. Wang, P.; Hu, T.; Gao, F.; et al. A Hybrid Data-Driven Framework for Spatiotemporal Traffic Flow Data Imputation. *IEEE Internet Things J.* **2022**, *9*, 16343–16352.
19. Lei, M.; Labbe, A.; Wu, Y.; et al. Bayesian Kernelized Matrix Factorization for Spatiotemporal Traffic Data Imputation and Kriging. *IEEE Trans. Intell. Transp. Syst.* **2022**, *23*, 18962–18974.
20. Wang, Y.; Zhang, Y.; Wang, L.; et al. Urban Traffic Pattern Analysis and Applications Based on Spatio-Temporal Non-Negative Matrix Factorization. *IEEE Trans. Intell. Transp. Syst.* **2022**, *23*, 12752–12765.
21. Zhang, Z.; Li, M.; Lin, X.; et al. Network-Wide Traffic Flow Estimation with Insufficient Volume Detection and Crowdsourcing Data. *Transp. Res. Part C Emerg. Technol.* **2020**, *121*, 102870.
22. Sure, P.; Srinivasan, C.P.; Babu, C.N. Spatio-Temporal Constraint-Based Low Rank Matrix Completion Approaches for Road Traffic Networks. *IEEE Trans. Intell. Transp. Syst.* **2022**, *23*, 13452–13462.
23. Yang, J.M.; Peng, Z.R.; Lin, L. Real-Time Spatiotemporal Prediction and Imputation of Traffic Status Based on LSTM and Graph Laplacian Regularized Matrix Factorization. *Transp. Res. Part C Emerg. Technol.* **2021**, *129*, 103228.
24. Xu, X.; Lin, M.; Luo, X.; et al. HRST-LR: A Hessian Regularization Spatio-Temporal Low Rank Algorithm for Traffic Data Imputation. *IEEE Trans. Intell. Transp. Syst.* **2023**, *24*, 11001–11017.
25. Wu, H.; Qiao, Y.; Luo, X. A Fine-Grained Regularization Scheme for Nonnegative Latent Factorization of High-Dimensional and Incomplete Tensors. *IEEE Trans. Serv. Comput.* **2024**, *17*, 3006–3021.
26. Wu, H.; Wu, X.; Luo, X. *Dynamic Network Representation Based on Latent Factorization of Tensors*; Springer: Berlin/Heidelberg, Germany, 2023.
27. Liao, X.; Wu, H.; Luo, X. A Novel Tensor Causal Convolution Network Model for Highly-Accurate Representation to Spatio-Temporal Data. *IEEE Trans. Autom. Sci. Eng.* **2025**, *22*, 19525–19537.
28. Wu, H.; Mi, J. A Cauchy Loss-Incorporated Nonnegative Latent Factorization of Tensors Model for Spatiotemporal Traffic Data Recovery. *Neurocomputing* **2025**, *626*, 129575.
29. Ben Said, A.; Erradi, A. Spatiotemporal Tensor Completion for Improved Urban Traffic Imputation. *IEEE Trans. Intell. Transp. Syst.* **2022**, *23*, 6836–6849.
30. Bhanu, M.; Mendes-Moreira, J.; Chandra, J. Embedding Traffic Network Characteristics Using Tensor for Improved Traffic Prediction. *IEEE Trans. Intell. Transp. Syst.* **2021**, *22*, 3359–3371.
31. Baggag, A.; Abbar, S.; Sharma, A.; et al. Learning Spatiotemporal Latent Factors of Traffic via Regularized Tensor Factorization: Imputing Missing Values and Forecasting. *IEEE Trans. Knowl. Data Eng.* **2021**, *33*, 2573–2587.
32. Chen, X.; Sun, L. Bayesian Temporal Factorization for Multidimensional Time Series Prediction. *IEEE Trans. Pattern Anal. Mach. Intell.* **2022**, *44*, 4659–4673.
33. Yang, Z.; Yang, L.T.; Li, C.; et al. Collaborative Bayesian Tensor Factorization-Based Reliable Traffic Speed Data

Prediction in T-CPS. *IEEE Trans. Intell. Transp. Syst.* **2025**, *26*, 14393–14406.

34. Ahn, D.; Kim, S.; Kang, U. Accurate Online Tensor Factorization for Temporal Tensor Streams with Missing Values. In Proceedings of the 30th ACM International Conference on Information & Knowledge Management, online, 1–5 November 2021; pp. 2822–2826.
35. Xing, J.; Liu, R.; Anish, K.; et al. A Customized Data Fusion Tensor Approach for Interval-Wise Missing Network Volume Imputation. *IEEE Trans. Intell. Transp. Syst.* **2023**, *24*, 12107–12122.
36. Zhu, Y.; Wang, J.; Wang, J.; et al. Multitask Neural Tensor Factorization for Road Traffic Speed-Volume Correlation Pattern Learning and Joint Imputation. *IEEE Trans. Intell. Transp. Syst.* **2022**, *23*, 24550–24560.
37. Li, J.; Xu, L.; Li, R.; et al. Deep Spatial-Temporal Bi-Directional Residual Optimisation Based on Tensor Decomposition for Traffic Data Imputation on Urban Road Network. *Appl. Intell.* **2022**, *52*, 11363–11381.
38. Guo, Y.; Wang, X.; Lan, X.; et al. Traffic Target Location Estimation Based on Tensor Decomposition in Intelligent Transportation System. *IEEE Trans. Intell. Transp. Syst.* **2024**, *25*, 816–828.
39. Li, Q.; Yang, X.; Wang, Y.; et al. Spatial–Temporal Traffic Modeling with a Fusion Graph Reconstructed by Tensor Decomposition. *IEEE Trans. Intell. Transp. Syst.* **2024**, *25*, 1749–1760.
40. Wu, Q.; Jiang, Z.; Hong, K.; et al. Tensor-Based Recurrent Neural Network and Multi-Modal Prediction with Its Applications in Traffic Network Management. *IEEE Trans. Netw. Serv. Manag.* **2021**, *18*, 780–792.
41. Jia, X.; Dong, X.; Chen, M.; et al. Missing Data Imputation for Traffic Congestion Data Based on Joint Matrix Factorization. *Knowl.-Based Syst.* **2021**, *225*, 107114.
42. Luo, Q.; Yang, M.; Li, W.; et al. Multidimensional Data Processing with Bayesian Inference via Structural Block Decomposition. *IEEE Trans. Cybern.* **2024**, *54*, 3132–3145.
43. Chen, H.; Lin, M.; Liu, J.; et al. NT-DPTC: A Non-Negative Temporal Dimension Preserved Tensor Completion Model for Missing Traffic Data Imputation. *Inf. Sci.* **2024**, *653*, 119797.
44. Shen, Y.; Jin, C.; Hua, J.; et al. TTPNet: A Neural Network for Travel Time Prediction Based on Tensor Decomposition and Graph Embedding. *IEEE Trans. Knowl. Data Eng.* **2020**, *34*, 4514–4526.
45. Deng, L.; Liu, X.Y.; Zheng, H.; et al. Graph Spectral Regularized Tensor Completion for Traffic Data Imputation. *IEEE Trans. Intell. Transp. Syst.* **2022**, *23*, 10996–11010.
46. Feng, X.; Zhang, H.; Wang, C.; et al. Traffic Data Recovery from Corrupted and Incomplete Observations via Spatial-Temporal TRPCA. *IEEE Trans. Intell. Transp. Syst.* **2022**, *23*, 17835–17848.
47. Chen, X.; Wang, K.; Ye, Q. Nonconvex Low-Tubal-Rank Tensor Completion with Temporal Regularization for Spatiotemporal Traffic Data Recovery. *IEEE Trans. Emerg. Top. Comput. Intell.* **2025**, *9*, 4066–4079.
48. Chen, X.; Yang, J.; Sun, L. A Nonconvex Low-Rank Tensor Completion Model for Spatiotemporal Traffic Data Imputation. *Transp. Res. Part C Emerg. Technol.* **2020**, *117*, 102673.
49. Wang, X.; Wu, Y.; Zhuang, D.; et al. Low-Rank Hankel Tensor Completion for Traffic Speed Estimation. *IEEE Trans. Intell. Transp. Syst.* **2023**, *24*, 4862–4871.
50. Hu, L.; Jia, Y.; Chen, W.; et al. A Flexible and Robust Tensor Completion Approach for Traffic Data Recovery with Low-Rankness. *IEEE Trans. Intell. Transp. Syst.* **2023**, *25*, 2558–2572.
51. Dai, C.; Zhang, Y.; Zheng, Z. A Nonlocal Similarity Learning-Based Tensor Completion Model with Its Application in Intelligent Transportation System. *IEEE Trans. Intell. Transp. Syst.* **2024**, *25*, 3140–3151.
52. Nie, T.; Qin, G.; Sun, J. Truncated Tensor Schatten p-Norm Based Approach for Spatiotemporal Traffic Data Imputation with Complicated Missing Patterns. *Transp. Res. Part C Emerg. Technol.* **2022**, *141*, 103737.
53. Wang, Q.; Chen, L.; Wang, Q.; et al. Anomaly-Aware Network Traffic Estimation via Outlier-Robust Tensor Completion. *IEEE Trans. Netw. Serv. Manag.* **2020**, *17*, 2677–2689.
54. Deng, L.; Liu, X.Y.; Zheng, H.; et al. Graph-Tensor Neural Networks for Network Traffic Data Imputation. *IEEE/ACM Trans. Netw.* **2023**, *31*, 3010–3024.
55. Chen, X.; Liang, S.; Zhang, Z.; et al. A Novel Spatiotemporal Data Low-Rank Imputation Approach for Traffic Sensor Network. *IEEE Internet Things J.* **2022**, *9*, 20122–20135.
56. Chen, X.; Wang, K.; Li, Z.; et al. A Novel Nonconvex Low-Rank Tensor Completion Approach for Traffic Sensor Data Recovery from Incomplete Measurements. *IEEE Trans. Instrum. Meas.* **2023**, *72*, 2518715.
57. Chen, X.; Chen, Y.; Saunier, N.; et al. Scalable Low-Rank Tensor Learning for Spatiotemporal Traffic Data Imputation. *Transp. Res. Part C Emerg. Technol.* **2021**, *129*, 103226.
58. Chen, X.; Lei, M.; Saunier, N.; et al. Low-Rank Autoregressive Tensor Completion for Spatiotemporal Traffic Data Imputation. *IEEE Trans. Intell. Transp. Syst.* **2022**, *23*, 12301–12310.
59. Nie, T.; Qin, G.; Wang, Y.; et al. Correlating Sparse Sensing for Large-Scale Traffic Speed Estimation: A Laplacian-Enhanced Low-Rank Tensor Kriging Approach. *Transp. Res. Part C Emerg. Technol.* **2023**, *152*, 104190.
60. Chen, X.; Cheng, Z.; Cai, H.; et al. Laplacian Convolutional Representation for Traffic Time Series Imputation. *IEEE Trans. Knowl. Data Eng.* **2024**, *36*, 6490–6502.
61. Zhang, Y.; Wei, X.; Zhang, X.; et al. Self-Attention Graph Convolution Residual Network for Traffic Data Completion. *IEEE Trans. Big Data* **2023**, *9*, 528–541.

62. Wu, Y.; Zhuang, D.; Labbe, A.; et al. Inductive Graph Neural Networks for Spatiotemporal Kriging. In Proceedings of the AAAI Conference on Artificial Intelligence, online, 2–9 February 2021; pp. 4478–4485.
63. Kong, W.; Guo, Z.; Liu, Y. Spatio-Temporal Pivotal Graph Neural Networks for Traffic Flow Forecasting. In Proceedings of the 38th AAAI Conference on Artificial Intelligence, Vancouver, BC, Canada, 20–27 February 2024; pp. 8627–8635.
64. Fang, S.; Prinet, V.; Chang, J.; et al. MS-Net: Multi-Source Spatio-Temporal Network for Traffic Flow Prediction. *IEEE Trans. Intell. Transp. Syst.* **2022**, *23*, 7142–7155.
65. Zheng, Q.; Zhang, Y. DSTAGCN: Dynamic Spatial-Temporal Adjacent Graph Convolutional Network for Traffic Forecasting. *IEEE Trans. Big Data* **2023**, *9*, 241–253.
66. Cui, Z.; Henrickson, K.; Ke, R.; et al. Traffic Graph Convolutional Recurrent Neural Network: A Deep Learning Framework for Network-Scale Traffic Learning and Forecasting. *IEEE Trans. Intell. Transp. Syst.* **2020**, *21*, 4883–4894.
67. Li, F.; Feng, J.; Yan, H.; et al. Dynamic Graph Convolutional Recurrent Network for Traffic Prediction: Benchmark and Solution. *ACM Trans. Knowl. Discov. Data* **2023**, *17*, 1–21.
68. Li, Y.; Yu, R.; Shahabi, C.; et al. Diffusion Convolutional Recurrent Neural Network: Data-Driven Traffic Forecasting. *arXiv* **2017**, arXiv:1707.01926.
69. Liang, Y.; Zhao, Z.; Sun, L. Memory-Augmented Dynamic Graph Convolution Networks for Traffic Data Imputation with Diverse Missing Patterns. *Transp. Res. Part C Emerg. Technol.* **2022**, *143*, 103826.
70. Zhang, Z.; Lin, X.; Li, M.; et al. A Customized Deep Learning Approach to Integrate Network-Scale Online Traffic Data Imputation and Prediction. *Transp. Res. Part C Emerg. Technol.* **2021**, *132*, 103372.
71. Kong, X.; Zhou, W.; Shen, G.; et al. Dynamic Graph Convolutional Recurrent Imputation Network for Spatiotemporal Traffic Missing Data. *Knowl.-Based Syst.* **2023**, *261*, 110188.
72. Ming, J.; Zhang, L.; Fan, W.; et al. Multi-Graph Convolutional Recurrent Network for Fine-Grained Lane-Level Traffic Flow Imputation. In Proceedings of the 2022 IEEE International Conference on Data Mining (ICDM), Orlando, FL, USA, 28 November–1 December 2022; pp. 348–357.
73. Li, J.; Li, R.; Xu, L. Multi-Stage Deep Residual Collaboration Learning Framework for Complex Spatial–Temporal Traffic Data Imputation. *Appl. Soft Comput.* **2023**, *147*, 110814.
74. Wang, P.; Zhang, T.; Zheng, Y.; et al. A Multi-View Bidirectional Spatiotemporal Graph Network for Urban Traffic Flow Imputation. *Int. J. Geogr. Inf. Sci.* **2022**, *36*, 1231–1257.
75. Zhang, K.; Zhou, F.; Wu, L.; et al. Semantic Understanding and Prompt Engineering for Large-Scale Traffic Data Imputation. *Inf. Fusion* **2024**, *102*, 102038.
76. Huo, G.; Zhang, Y.; Wang, B.; et al. Hierarchical Spatio-Temporal Graph Convolutional Networks and Transformer Network for Traffic Flow Forecasting. *IEEE Trans. Intell. Transp. Syst.* **2023**, *24*, 3855–3867.
77. Cini, A.; Marisca, I.; Alippi, C. Filling the G_ap_s: Multivariate Time Series Imputation by Graph Neural Networks. In Proceedings of the Tenth International Conference on Learning Representations, online, 25–29 April 2022; pp. 1–20.
78. Jiang, R.; Wang, Z.; Yong, J.; et al. Spatio-Temporal Meta-Graph Learning for Traffic Forecasting. In Proceedings of the AAAI Conference on Artificial Intelligence, Washington, DC, USA, 7–14 February 2023; pp. 8078–8086.
79. Wang, B.; Lin, Y.; Guo, S.; et al. GSNet: Learning Spatial-Temporal Correlations from Geographical and Semantic Aspects for Traffic Accident Risk Forecasting. In Proceedings of the AAAI Conference on Artificial Intelligence, online, 2–9 February 2021; pp. 4402–4409.
80. Wang, A.; Ye, Y.; Song, X.; et al. Traffic Prediction with Missing Data: A Multi-Task Learning Approach. *IEEE Trans. Intell. Transp. Syst.* **2023**, *24*, 4189–4202.
81. Luo, G.; Zhang, H.; Yuan, Q.; et al. ESTNet: Embedded Spatial-Temporal Network for Modeling Traffic Flow Dynamics. *IEEE Trans. Intell. Transp. Syst.* **2022**, *23*, 19201–19212.
82. Hu, J.; Zheng, T.; Peng, L.; et al. LightST: A Simplifying Spatio-Temporal Graph Neural Network for Traffic Flow Forecasting. *IEEE Trans. Big Data* **2025**, *11*, 2517–2528.
83. Shen, G.; Zhou, W.; Zhang, W.; et al. Bidirectional Spatial-Temporal Traffic Data Imputation via Graph Attention Recurrent Neural Network. *Neurocomputing* **2023**, *531*, 151–162.
84. Guo, S.; Lin, Y.; Wan, H.; et al. Learning Dynamics and Heterogeneity of Spatial-Temporal Graph Data for Traffic Forecasting. *IEEE Trans. Knowl. Data Eng.* **2022**, *34*, 5415–5428.
85. Zhiwen, Z.; Wang, H.; Fan, Z.; et al. Missing Road Condition Imputation Using a Multi-View Heterogeneous Graph Network from GPS Trajectory. *IEEE Trans. Intell. Transp. Syst.* **2023**, *24*, 4917–4931.
86. Zheng, C.; Fan, X.; Wang, C.; et al. GMAN: A Graph Multi-Attention Network for Traffic Prediction. In Proceedings of the 34th AAAI Conference on Artificial Intelligence, New York, NY, USA, 7–12 February 2020; pp. 1234–1241.
87. Duan, Y.; Chen, N.; Shen, S.; et al. FDSA-STG: Fully Dynamic Self-Attention Spatio-Temporal Graph Networks for Intelligent Traffic Flow Prediction. *IEEE Trans. Veh. Technol.* **2022**, *71*, 9250–9260.
88. Wu, Z.; Pan, S.; Long, G.; et al. 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, online, 6–10 July 2020; pp. 753–763.
89. Guo, K.; Hu, Y.; Sun, Y.; et al. Hierarchical Graph Convolution Network for Traffic Forecasting. In Proceedings of the

35th AAAI Conference on Artificial Intelligence (AAAI-21), online, 2–9 February 2021; pp. 151–159.
90. Liang, W.; Li, Y.; Xie, K.; et al. Spatial-Temporal Aware Inductive Graph Neural Network for C-ITS Data Recovery. *IEEE Trans. Intell. Transp. Syst.* **2023**, *24*, 8431–8442.
91. Nie, T.; Qin, G.; Wang, Y.; Sun, J. Towards Better Traffic Volume Estimation: Jointly Addressing the Underdetermination and Nonequilibrium Problems with Correlation-Adaptive GNNs. *Transp. Res. Part C Emerg. Technol.* **2023**, *157*, 104402.
92. Liang, M.; Liu, R.W.; Zhan, Y.; et al. Fine-Grained Vessel Traffic Flow Prediction with a Spatio-Temporal Multigraph Convolutional Network. *IEEE Trans. Intell. Transp. Syst.* **2022**, *23*, 23694–23707.
93. Qu, Y.; Li, Z.; Zhao, X.; et al. Towards Real-World Traffic Prediction and Data Imputation: A Multi-Task Pretraining and Fine-Tuning Approach. *Inf. Sci.* **2024**, *657*, 119972.
94. Zheng, C.; Fan, X.; Pan, S.; et al. Spatio-Temporal Joint Graph Convolutional Networks for Traffic Forecasting. *IEEE Trans. Knowl. Data Eng.* **2024**, *36*, 372–385.
95. Liu, M.; Zhu, T.; Ye, J.; et al. Spatio-Temporal Autoencoder for Traffic Flow Prediction. *IEEE Trans. Intell. Transp. Syst.* **2023**, *24*, 5516–5526.
96. Wang, S.; Zhang, J.; Li, J.; et al. Traffic Accident Risk Prediction via Multi-View Multi-Task Spatio-Temporal Networks. *IEEE Trans. Knowl. Data Eng.* **2023**, *35*, 12323–12336.
97. Jin, G.; Li, F.; Zhang, J.; et al. Automated Dilated Spatio-Temporal Synchronous Graph Modeling for Traffic Prediction. *IEEE Trans. Intell. Transp. Syst.* **2023**, *24*, 8820–8830.
98. Shin, Y.; Yoon, Y. PGCN: Progressive Graph Convolutional Networks for Spatial–Temporal Traffic Forecasting. *IEEE Trans. Intell. Transp. Syst.* **2024**, *25*, 7633–7644.
99. Yuan, Y.; Zhang, Y.; Wang, B.; et al. STGAN: Spatio-Temporal Generative Adversarial Network for Traffic Data Imputation. *IEEE Trans. Big Data* **2023**, *9*, 200–211.
100. Zhang, K.; He, Z.; Zheng, L.; et al. A Generative Adversarial Network for Travel Times Imputation Using Trajectory Data. *Comput.-Aided Civ. Infrastruct. Eng.* **2021**, *36*, 197–212.
101. Xu, D.; Wei, C.; Peng, P.; et al. GE-GAN: A Novel Deep Learning Framework for Road Traffic State Estimation. *Transp. Res. Part C Emerg. Technol.* **2020**, *117*, 102635.
102. Hou, M.; Tang, T.; Xia, F.; et al. MISSII: Missing Information Imputation for Traffic Data. *IEEE Trans. Emerg. Top. Comput.* **2024**, *12*, 752–765.
103. Zhang, W.; Zhang, P.; Yu, Y.; et al. Missing Data Repairs for Traffic Flow with Self-Attention Generative Adversarial Imputation Net. *IEEE Trans. Intell. Transp. Syst.* **2022**, *23*, 7919–7930.
104. Xu, D.; Peng, H.; Wei, C.; et al. Traffic State Data Imputation: An Efficient Generating Method Based on the Graph Aggregator. *IEEE Trans. Intell. Transp. Syst.* **2022**, *23*, 13084–13093.
105. Xiao, X.; Zhang, Y.; Yang, S.; et al. Efficient Missing Counts Imputation of a Bike-Sharing System by Generative Adversarial Network. *IEEE Trans. Intell. Transp. Syst.* **2022**, *23*, 13443–13451.
106. Li, J.; Li, R.; Xu, L.; et al. Self-Supervised Generative Adversarial Learning with Conditional Cyclical Constraints Towards Missing Traffic Data Imputation. *Knowl.-Based Syst.* **2024**, *284*, 111233.
107. Chen, Y.; Lv, Y.; Wang, F.Y. Traffic Flow Imputation Using Parallel Data and Generative Adversarial Networks. *IEEE Trans. Intell. Transp. Syst.* **2020**, *21*, 1624–1630.
108. Xie, K.; Ouyang, Y.; Wang, X.; et al. Deep Adversarial Tensor Completion for Accurate Network Traffic Measurement. *IEEE/ACM Trans. Netw.* **2023**, *31*, 2101–2116.
109. Yang, B.; Kang, Y.; Yuan, Y.; et al. ST-LBAGAN: Spatio-Temporal Learnable Bidirectional Attention Generative Adversarial Networks for Missing Traffic Data Imputation. *Knowl.-Based Syst.* **2021**, *215*, 106705.
110. Han, L.; Zheng, K.; Zhao, L.; et al. Content-Aware Traffic Data Completion in ITS Based on Generative Adversarial Nets. *IEEE Trans. Veh. Technol.* **2020**, *69*, 11950–11962.
111. Zhang, B.; Miao, R.; Chen, Z. Spatial-Temporal Traffic Data Imputation Based on Dynamic Multi-Level Generative Adversarial Networks for Urban Governance. *Appl. Soft Comput.* **2024**, *151*, 111128.
112. Zhang, T.; Wang, J.; Liu, J. A Gated Generative Adversarial Imputation Approach for Signalized Road Networks. *IEEE Trans. Intell. Transp. Syst.* **2021**, *23*, 12144–12160.
113. Qin, H.; Zhan, X.; Li, Y.; et al. Network-Wide Traffic States Imputation Using Self-Interested Coalitional Learning. In Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery & Data Minin, online, 14–18 August 2021; pp. 1370–1378.
114. Kwon, J.; Cha, C.; Park, H. Vehicle Speed Data Imputation Based on Parameter Transferred LSTM. In Proceedings of the 34th Conference on Neural Information Processing Systems (NeurIPS 2020), online, 6–12 December 2020.
115. Cui, Z.; Ke, R.; Pu, Z.; et al. Stacked Bidirectional and Unidirectional LSTM Recurrent Neural Network for Forecasting Network-Wide Traffic State with Missing Values. *Transp. Res. Part C Emerg. Technol.* **2020**, *118*, 102674.
116. Kwon, J.; Park, H. Parameter Transferred Irreducible LSTM for Traffic Data Imputation. *IEEE Sens. J.* **2024**, *24*, 22178–22188.
117. Ma, C.; Dai, G.; Zhou, J. Short-Term Traffic Flow Prediction for Urban Road Sections Based on Time Series Analysis and LSTM_BILSTM Method. *IEEE Trans. Intell. Transp. Syst.* **2022**, *23*, 5615–5624.

118. Wang, Z.; Su, X.; Ding, Z. Long-Term Traffic Prediction Based on LSTM Encoder-Decoder Architecture. *IEEE Trans. Intell. Transp. Syst.* **2021**, *22*, 6561–6571.
119. Shan, S.; Li, Y.; Oliva, J.B. NRTSI: Non-Recurrent Time Series Imputation. In Proceedings of the 2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Rhodes Island, Greece, 4–10 June 2023; pp. 1–5.
120. Liu, H.; Dong, Z.; Jiang, R.; et al. Spatio-Temporal Adaptive Embedding Makes Vanilla Transformer SOTA for Traffic Forecasting. In Proceedings of the 32nd ACM International Conference on Information and Knowledge Management (CIKM 2023), Birmingham, UK, 21–25 October 2023; pp. 4125–4129.
121. Yan, H.; Ma, X.; Pu, Z. Learning Dynamic and Hierarchical Traffic Spatiotemporal Features with Transformer. *IEEE Trans. Intell. Transp. Syst.* **2022**, *23*, 22386–22399.
122. Zhu, S.; Ding, R.; Zhang, M.; et al. Spatio-Temporal Point Processes with Attention for Traffic Congestion Event Modeling. *IEEE Trans. Intell. Transp. Syst.* **2022**, *23*, 7298–7309.
123. Marisca, I.; Cini, A.; Alippi, C. Learning to Reconstruct Missing Data from Spatiotemporal Graphs with Sparse Observations. In Proceedings of the 36th Conference on Neural Information Processing Systems (NeurIPS 2022), New Orleans, LA, USA, 28 November–9 December 2022; pp. 32069–32082.
124. Xu, Q.; Ruan, S.; Long, C.; et al. Traffic Speed Imputation with Spatio-Temporal Attentions and Cycle-Perceptual Training. In Proceedings of the 31st ACM International Conference on Information and Knowledge Management (CIKM 2022), Atlanta, GA, USA, 17–21 October 2022; pp. 2280–2289.
125. Cirstea, R.G.; Yang, B.; Guo, C.; et al. Towards Spatio-Temporal Aware Traffic Time Series Forecasting. In Proceedings of the 38th IEEE International Conference on Data Engineering (ICDE 2022), Kuala Lumpur, Malaysia, 9–12 May 2022; pp. 2900–2913.
126. Liu, Y.; Yu, R.; Zheng, S.; et al. Naomi: Non-Autoregressive Multiresolution Sequence Imputation. In Proceedings of the 32nd Conference on Neural Information Processing Systems (NeurIPS 2019), Vancouver, BC, Canada, 8–14 December 2019; p. 32.
127. Liu, C.; Yang, S.; Xu, Q.; et al. Spatial-Temporal Large Language Model for Traffic Prediction. In Proceedings of the 25th IEEE International Conference on Mobile Data Management (MDM 2024), Brussels, Belgium, 24–27 June 2024; pp. 31–40.
128. Yao, H.; Da, L.; Nandam, V.; et al. Comal: Collaborative Multi-Agent Large Language Models for Mixed-Autonomy Traffic. In Proceedings of the 2025 SIAM International Conference on Data Mining (SDM 2025), Alexandria, VA, USA, 1–3 May 2025; pp. 409–418.
129. Rong, Y.; Mao, Y.; Cui, H.; et al. Edge Computing Enabled Large-Scale Traffic Flow Prediction with GPT in Intelligent Autonomous Transport System for 6G Network. *IEEE Trans. Intell. Transp. Syst.* **2024**, *26*, 117321–17338.
130. Xu, Y.; Liu, M. GPT4TFP: Spatio-Temporal Fusion Large Language Model for Traffic Flow Prediction. *Neurocomputing* **2025**, *625*, 129562.
131. Zhong, W.; Huang, J.; Wu, M.; et al. Large Language Model Based System with Causal Inference and Chain-of-Thoughts Reasoning for Traffic Scene Risk Assessment. *Knowl.-Based Syst.* **2025**, *319*, 113630.
132. Hu, Y.; Wang, F.; Ye, D.; et al. LLM-Based Misbehavior Detection Architecture for Enhanced Traffic Safety in Connected Autonomous Vehicles. *IEEE Trans. Veh. Technol.* **2025**, *74*, 12829–12841.
133. Movahedi, M.; Choi, J. The Crossroads of LLM and Traffic Control: A Study on Large Language Models in Adaptive Traffic Signal Control. *IEEE Trans. Intell. Transp. Syst.* **2025**, *26*, 1701–1716.
134. Li, Z.; Xia, L.; Tang, J.; et al. UrbanGPT: Spatio-Temporal Large Language Models. In Proceedings of the 30th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining (KDD 2024), Barcelona, Spain, 25–29 August 2024; pp. 5351–5362.
135. Hu, J.; Hu, C.; Yang, J.; et al. Do Traffic Flow States Follow Markov Properties? A High-Order Spatiotemporal Traffic State Reconstruction Approach for Traffic Prediction and Imputation. *Chaos Solitons Fractals* **2024**, *183*, 114965.
136. Liu, M.; Huang, H.; Feng, H.; et al. Pristi: A Conditional Diffusion Framework for Spatiotemporal Imputation. In Proceedings of the 39th IEEE International Conference on Data Engineering (ICDE 2023), Anaheim, CA, USA, 3–7 April 2023; pp. 1927–1939.
137. Liu, J.; Wu, N.; Qiao, Y.; et al. Short-Term Traffic Flow Forecasting Using Ensemble Approach Based on Deep Belief Networks. *IEEE Trans. Intell. Transp. Syst.* **2022**, *23*, 404–417.
138. Dai, F.; Huang, P.; Mo, Q.; et al. ST-InNet: Deep Spatio-Temporal Inception Networks for Traffic Flow Prediction in Smart Cities. *IEEE Trans. Intell. Transp. Syst.* **2022**, *23*, 19782–19794.
139. Babu, C.N.; Sure, P.; Bhuma, C.M. Sparse Bayesian Learning Assisted Approaches for Road Network Traffic State Estimation. *IEEE Trans. Intell. Transp. Syst.* **2021**, *22*, 1733–1741.
140. Li, H.; Li, M.; Lin, X.; et al. A Spatiotemporal Approach for Traffic Data Imputation with Complicated Missing Patterns. *Transp. Res. Part C Emerg. Technol.* **2020**, *119*, 102730.
141. Li, L.; Zhang, J.; Wang, Y.; et al. Missing Value Imputation for Traffic-Related Time Series Data Based on a Multi-View Learning Method. *IEEE Trans. Intell. Transp. Syst.* **2019**, *20*, 2933–2943.
142. Kaur, M.; Singh, S.; Aggarwal, N. Missing Traffic Data Imputation Using a Dual-Stage Error-Corrected Boosting Regressor

with Uncertainty Estimation. *Inf. Sci.* **2022**, *586*, 344–373.

143. Ji, J.; Wang, J.; Huang, C.; et al. Spatio-Temporal Self-Supervised Learning for Traffic Flow Prediction. In Proceedings of the 37th AAAI Conference on Artificial Intelligence (AAAI-23), Washington, DC, USA, 7–14 February 2023; pp. 4356–4364.
144. Wang, L.; Chai, D.; Liu, X.; et al. Exploring the Generalizability of Spatio-Temporal Traffic Prediction: Meta-Modeling and an Analytic Framework. *IEEE Trans. Knowl. Data Eng.* **2023**, *35*, 3870–3884.
145. Peng, W.; Varanka, T.; Mostafa, A.; et al. Hyperbolic Deep Neural Networks: A Survey. *IEEE Trans. Pattern Anal. Mach. Intell.* **2022**, *44*, 10023–10044.
146. Cai, Z.; Jiang, R.; Yang, X.; et al. MemDA: Forecasting Urban Time Series with Memory-Based Drift Adaptation. In Proceedings of the 32nd ACM International Conference on Information and Knowledge Management (CIKM 2023), Birmingham, UK, 21–25 October 2023; pp. 193–202.