Spatio-Temporal Graph Neural Networks for Predictive Learning in Urban Computing: A Survey

Guangyin Jin, Yuxuan Liang, *Member, IEEE*, Yuchen Fang, Zezhi Shao, Jincai Huang, Junbo Zhang, *Senior Member, IEEE*, and Yu Zheng, *Fellow, IEEE*

(Survey Paper)

***Abstract*—With recent advances in sensing technologies, a myriad of spatio-temporal data has been generated and recorded in smart cities. Forecasting the evolution patterns of spatio-temporal data is an important yet demanding aspect of urban computing, which can enhance intelligent management decisions in various fields, including transportation, environment, climate, public safety, healthcare, and others. Traditional statistical and deep learning methods struggle to capture complex correlations in urban spatio-temporal data. To this end, Spatio-Temporal Graph Neural Networks (STGNN) have been proposed, achieving great promise in recent years. STGNNs enable the extraction of complex spatio-temporal dependencies by integrating graph neural networks (GNNs) and various temporal learning methods. In this manuscript, we provide a comprehensive survey on recent progress on STGNN technologies for predictive learning in urban computing. Firstly, we provide a brief introduction to the construction methods of spatio-temporal graph data and the prevalent deep-learning architectures used in STGNNs. We then sort out the primary application domains and specific predictive learning tasks based on existing literature. Afterward, we scrutinize the design of STGNNs and their combination with some advanced technologies in recent years. Finally, we conclude the limitations of existing research and suggest potential directions for future work.**

***Index Terms*—Graph neural networks, predictive learning, spatio-temporal data mining, time series, urban computing.**

Manuscript received 8 May 2023; revised 19 September 2023; accepted 11 November 2023. Date of publication 23 November 2023; date of current version 4 October 2024. This work was supported in part by Guangzhou Municiple Science and Technology Project under Grant 2023A03J0011, in part by the National Natural Science Foundation of China under Grants 62172034 and 72242106, and in part by Beijing Natural Science Foundation under Grant 4212021. Recommended for acceptance by W. Zhao. *(Corresponding author: Yuxuan Liang.)*

Guangyin Jin is with the National Innovative Institute of Defense Technology, Beijing 100071, China, and also with the National University of Defense Technology, Changsha 410003, China (e-mail: jinguangyin96@foxmail.com).

Yuxuan Liang is with the Intelligent Transportation Thrust, Hong Kong University of Science and Technology (Guangzhou), Guangzhou 511442, China (e-mail: yuxliang@outlook.com).

Yuchen Fang is with the School of Computer Science and Engineering, University of Electronic Science and Technology of China, Chengdu 611730, China (e-mail: fyclmiss@gmail.com).

Zezhi Shao is with the Institute of Computing Technology, University of Chinese Academy of Sciences, Beijing 100086, China (e-mail: shaozezhi19b@ict.ac.cn).

Jincai Huang is with the College of Systems Engineering, National University of Defense Technology, Changsha 410003, China (e-mail: huangjincai@nudt.edu.cn).

Junbo Zhang and Yu Zheng are with JD Intelligent Cities Research and JD iCity, JD Technology, Beijing 100176, China (e-mail: msjunbozhang@outlook.com; msyuzheng@outlook.com).

Digital Object Identifier 10.1109/TKDE.2023.3333824

I. Introduction

With the rapid advancement of sensing and data stream processing technologies, vast amounts of data in urban systems have been efficiently collected and stored. This has laid the foundation for the era of urban computing, which aims to understand the urban patterns and dynamics from different application domains where the Big Data explodes, such as transportation, environment, climate, etc. Predictive learning is a typical supervised learning paradigm that learns from historical data to forecast future trends. According to urban computing theories [1], predictive learning based on massive urban data is the most important loop, forming the foundation for intelligent decision-making, scheduling, and management in smart cities. In addition, the predictability of urban Big Data can also provide the possibility for the development of some new technologies such as digital twin cities and metaverse [2].

The majority of urban data is spatio-temporal, representing that it not only pertains to spatial locations but also changes over time. Within urban systems, spatio-temporal data exhibits ubiquitous properties of *correlation* and *heterogeneity* [3]. Correlation refers to the data being auto-correlated not only in the temporal dimension, but also in the spatial dimension. Heterogeneity is a property of spatio-temporal data wherein it displays varying patterns across different temporal or spatial ranges.The complex nature of the above characteristics has resulted in an increased difficulty in feature engineering. As a consequence, some methods that performed well in traditional time series forecasting, such as Support Vector Regression (SVR) [4], Random Forest (RF) [5], and Gradient Boosting Decision Tree (GBDT) [6], are less effective in achieving accurate prediction results. In the past decade, the rapid development of deep learning technologies has led to the emergence of hybrid neural networks based on Convolutional Neural Networks (CNN) [7] and Recurrent Neural Networks (RNN) [8]. These hybrid networks (e.g., ConvLSTM [9], PredRNN [10]) have been increasingly applied to predictive learning of urban spatio-temporal data and have shown significant advantages. However, the major limitation of these methods is their inability to learn directly from non-Euclidean data existing in urban systems, such as vehicle flows over road networks, traffic on route networks, and entities in urban knowledge graphs.

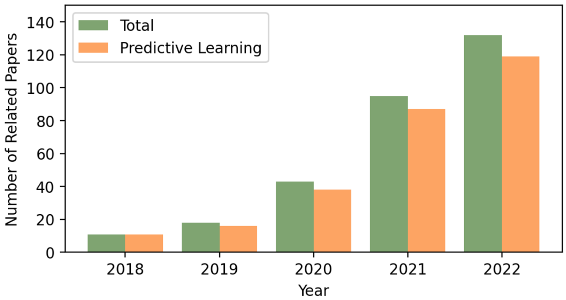


Fig. 1. Publication trend of STGNN-related papers in Google Scholar over the past five years. The blue bars represent the total number of relevant publications and the red bars denote those focusing on predictive learning tasks.

TABLE I
SUMMARY OF SYMBOL NOTATIONS

Notation	Definition
$\mathcal{G}_t$	Input ST-Graph at time t
A_t	Adjacency matrix of input ST-Graph at time t
$\mathcal{E}_t$	Edge set of input ST-Graph at time t
a_{ij}^t	Connection weight between nodes i and j at time t
$\mathcal{X}_t$	Input node features at time t
x_i^t	Input features of node i at time t
$\boldsymbol{f}(\cdot)$	An arbitrary nonlinear function
H_t	The hidden state of input node features at time t
$\star$	Convolution operator
$\odot$	Element-wise product operator

Over the past few years, there have been significant breakthroughs in representation learning of non-Euclidean data through deep learning techniques, particularly Graph Neural Networks (GNN) [11]. This has paved the way for predictive learning of diverse and intricate urban data. Given the spatio-temporal characteristics of urban data, such as traffic flows, a line of studies integrated GNNs with various temporal learning methods to capture dynamics in both space and time dimension [3]. This type of hybrid neural architecture is generally known as **Spatio-Temporal Graph Neural Network (STGNN)**. Recently, STGNNs have been widely used for predictive learning scenarios in urban computing, including transportation, environment, public safety, health, energy, economy, and other fields. Using the search engine of Google Scholar, we perform meticulous keyword searches and tally the relevant paper publications in the past five years. As depicted in Fig. 1, we can witness a notable surge in the number of relevant papers on STGNN year by year. In 2018, there are fewer than 20 papers, while in 2022, the number reaches nearly 140. This trend of progress indicates that STGNN-related applications have emerged as a highly sought-after research area in recent years. It is worth noting that a majority of these publications concentrated on predictive learning tasks.

Related Surveys: In recent years, there have been a few related surveys on the applications of STGNN-based predictive learning techniques across different fields. Wang et al. [3] conducted a review of deep learning techniques for spatio-temporal data mining involving a series of STGNNs in predictive learning up to 2020. There were also several surveys [12], [13], [14] investigating the blossom of STGNNs in transportation domains. To be specific, [12] analyzed multiple practical problems and revisited related works about prediction, detection, and control problems in urban traffic systems. Bui et al. [13] and Jiang et al. [14] focused on the latest STGNN technologies in traffic forecasting tasks. In recent months, there has been a small number of works [15], [16] surveying applications of STGNNs, but their topics are broad and do not focus on predictive learning in urban computing.

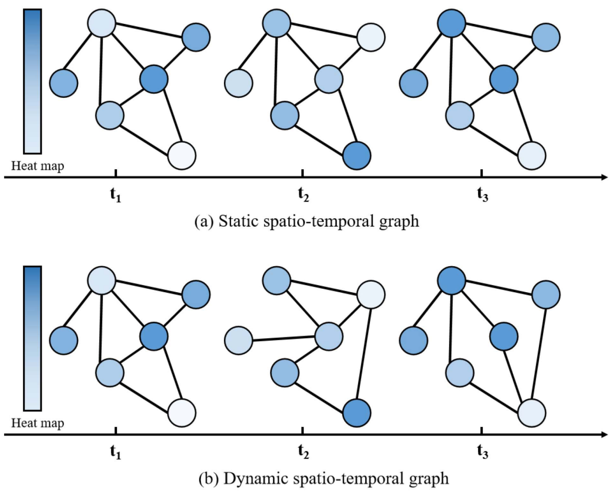
Our Contributions: In contrast to prior surveys, the contributions of our survey lie in four aspects:

- To our knowledge, this is the first comprehensive survey to systematically review recent studies that use STGNNs for predictive learning in urban computing. We scrutinize the progress of STGNN from both application and methodology perspectives based on extensive literature.
- We categorize the primary application domains as well as particular predictive learning tasks of STGNNs in urban computing, and sort out a list of public datasets attached with the previous works on STGNNs.
- We provide an in-depth analysis of existing STGNN methods for temporal learning, spatial learning, and spatio-temporal fusion. We further examine some recently emerging approaches that integrated STGNN with other advanced learning frameworks.
- We summarize the challenges shared by STGNNs for predictive learning tasks in urban computing and suggest future directions for addressing these challenging issues.

Organization: The rest is organized as follows. Section II illustrates how to construct spatio-temporal graphs based on prior knowledge. In Section III, a taxonomy of STGNNs for predictive learning in urban computing is presented. Section IV overviews various predictive learning tasks from different domains that can be addressed by STGNNs. Section V delineates the fundamental deep learning architectures commonly used in STGNNs. Section VI and Section VII delve into an in-depth analysis of the neural architecture design methods of STGNNs and popular advanced techniques that can be combined, respectively. Section VIII further highlights the limitations of existing works and suggests future directions. Finally, we conclude this survey in Section IX. To facilitate a quick understanding of the formulas in the paper, we have also compiled a list of symbols that encompasses the most commonly used symbol notations, as shown in Table I.

II. SPATIO-TEMPORAL GRAPH CONSTRUCTION

Suppose we obtain some observations from sensors, denoted as $\boldsymbol{X} = \{\mathcal{X}_t \in \mathbb{R}^{N\times F} | t = 0, \ldots, T\}$, where N is the number of spatial vertices and F is the number of features. **Spatio-temporal Graph** is an efficient structure to characterize the relationships between different vertices in a certain spatial and temporal range. We can represent a spatio-temporal graph as $\mathcal{G}_t = (\mathcal{V}, \mathcal{E}_t, \boldsymbol{A}_t)$, where $\mathcal{V}$ is the vertices set, $\mathcal{E}_t$ is the edge set, and A_t denotes the adjacency matrix at time t. In most scenarios, the size of $\mathcal{V}$ is static, while the size of $\mathcal{E}_t$ can be time-varying or constant, which indicates that $\boldsymbol{A}_t \in \mathbb{R}^{N\times N}$ also changes with



(a) Static spatio-temporal graph

(b) Dynamic spatio-temporal graph

Fig. 2. Schematic diagram of static and dynamic spatio-temporal graphs. The color shades of the nodes represent the numerical differences in some predictable features.

$\mathcal{E}_t$. In terms of connectivity, spatio-temporal graphs can be either directed or undirected, as well as weighted or unweighted. From the perspective of evolution, the structure of spatio-temporal graphs can be either static or dynamic. Fig. 2 illustrates the difference between static and dynamic spatio-temporal graphs. The appropriate type of spatio-temporal graph to construct depends on the task and the given data conditions.

Generally, the construction methods of predefined spatio-temporal graphs in urban computing systems can be divided into four categories: topology-based, distance-based, similarity-based, and interaction-based.

Topology-based graph: In the context of urban systems, topology-based graphs are usually constructed based on given topology structures, such as road networks [17], [18]. The adjacency matrix of a topology-based graph can be formulated as:

$$a_{ij}^t = \begin{cases} 1, & \text{if } v_i \text{ connects to } v_j \\ 0, & \text{otherwise} \end{cases}, \tag{1}$$

where a_{ij}^t denotes an element in adjacency matrix at time t, v_i and v_j are different vertices in the graph. Since the connections in topology structures can be symmetrical or asymmetrical, the topology-based graphs can be directed or undirected. Topology only represents connections in non-Euclidean spaces, thus the topology-based graphs are unweighted. In addition, the topology structures in urban systems are usually fixed for quite a long time, so we can treat them as static graphs.

Distance-based graph: According to first law of geography, *i.e.*, "Everything is related to everything else, but near things are more related to each other", we can construct a distance-based graph when a predefined topology is absent. In most applications, the elements in the adjacency matrix are calculated using a kernel function that takes the distances into account [19], [20], [21]. Gaussian radial basis function and inverted function are two common kernel functions used in previous literature. For example, the adjacency matrix of a distance-based graph with Gaussian radial basis can be computed as:

$$a_{ij}^t = \begin{cases} \frac{\exp(-\left\|d_{ij}^t\right\|_2)}{\sigma}, & \text{if } d_{ij}^t < \epsilon \\ 0, & \text{otherwise} \end{cases}, \tag{2}$$

where d_{ij}^t denotes the distance between node i and node j at time t; ϵ is a predefined threshold to control the sparsity of the adjacency matrix; σ is a hyper-parameter to control the distribution.

Similarity-based graph: Similarity can provide insights into the relations between different entities from a semantic perspective. Similarity-based graphs can be constructed based on either the proximity of time series [22], [23], [24] or similarity of the spatial attribute, e.g., Point of Interest (POI) [25]. In scenarios where additional data is unavailable, similarity-based graphs are typically constructed based on the similarity of time series. Pearson Correlation Coefficient (PCC) and Dynamic Time Wrapping (DTW) are two prevalent methods used to calculate the similarity between time series. For instance, the adjacency matrix of a similarity-based graph computed by PCC is defined as:

$$a_{ij}^t = \begin{cases} \frac{\sum_{i=1}^{n}\left(x_i^{0:t}-x_i^{\overline{0:t}}\right)\left(x_j^{0:t}-x_j^{\overline{0:t}}\right)}{\sqrt{\sum_{i=1}^{n}\left(x_i^{0:t}-x_i^{\overline{0:t}}\right)^2}\sqrt{\sum_{i=1}^{n}\left(x_j^{0:t}-x_j^{\overline{0:t}}\right)^2}} \\ 0, \qquad\qquad \text{otherwise} \end{cases}, \tag{3}$$

where $x_i^{0:t}$ and $x_j^{0:t}$ represent the time series of node i and node j of a given time span t, respectively; $x_i^{\overline{0:t}}$ and $x_j^{\overline{0:t}}$ are the mean value of the time series of node i and node j, and n denotes the number of samples over the time span t.

Interaction-based graph: The interaction between different locations can express their connection from the perspective of information flow [19], [21]. This is especially important when representing the characteristics of mobility, as the proportion of flow between two nodes can indicate the strength of their connection. Hence, the adjacency matrix of an interaction-based graph can be written as:

$$a_{ij}^t = \begin{cases} \frac{F_{ij}^t}{\sum_{m\in N(i)} F_{im}^t}, & \textit{if } F_{ij}^t > 0 \\ 0, & \text{otherwise} \end{cases}, \tag{4}$$

where F_{ij}^t denotes the flow from node i to node j at time t; $N(i)$ indicates the set of nodes that interact with node i; F_{im}^t is the flow from node i to other nodes (e.g., m) in a set $N(i)$ at time t.

In addition to the common predefined graph construction methods mentioned above, many relations in urban systems are implicit and difficult to be directly predefined. Therefore, spatio-temporal graphs based on adaptive learning have been proposed in some recent works. More details about these methods can be found in Section VI-A2.

III. Taxonomy

This section provides a taxonomy of STGNNs for predictive learning in urban computing, which is also a generalization of our follow-up content. As shown in Fig. 3, there are four main parts in our survey that need to be highlighted: main application domains, basic spatio-temporal learning neural architecture, improved spatio-temporal learning methods, and advanced methods combined with STGNNs. We present an overview of

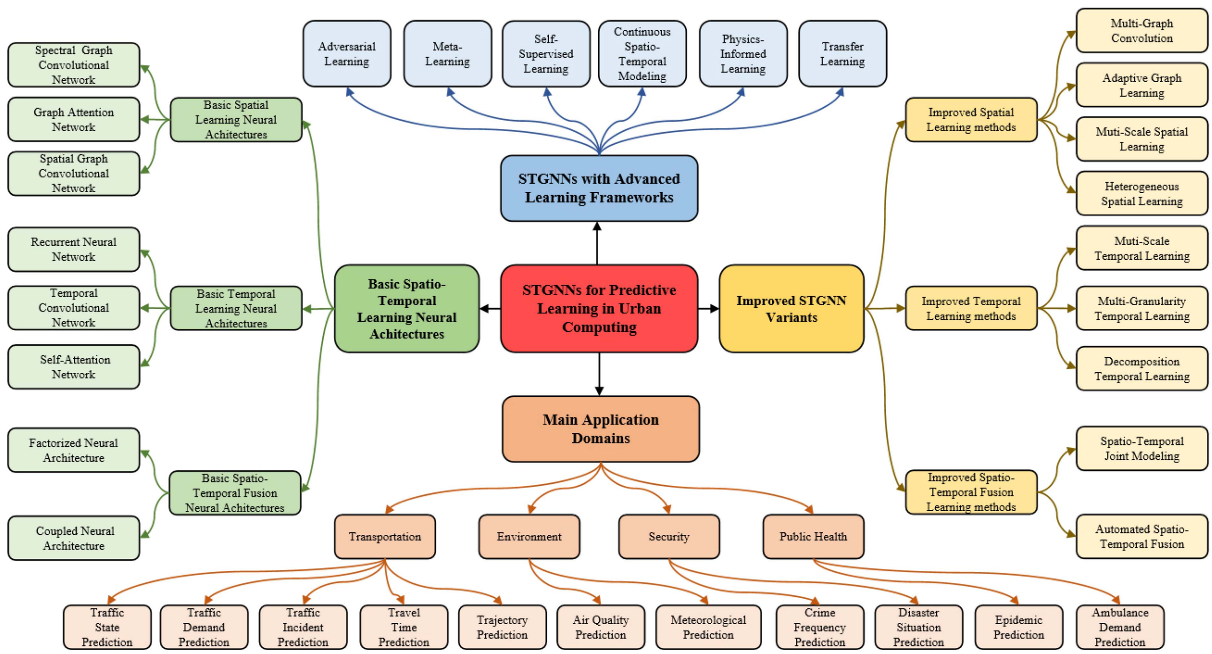


Fig. 3. Taxonomy for STGNN in our survey.

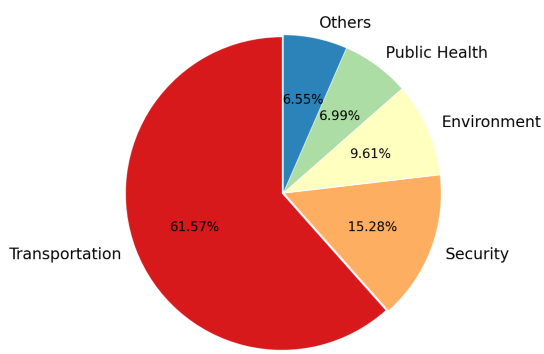


Fig. 4. Summary of the different application domains of STGNN in urban computing.

specific predictive learning tasks based on major application domains in Section IV. In Section V, we review the fundamental neural architectures of STGNNs from three perspectives: spatial learning, temporal learning and spatio-temporal fusion. Subsequently, in Section VI, we examine the enhanced spatio-temporal dependency learning methods from the same perspectives as in Section V. Finally, we discuss the advanced techniques combined with STGNNs in Section VII.

IV. Application Domains & Task Description

This section delves into the primary application domains and specific predictive learning tasks in urban computing. Based on the available literature in recent years, we conducted a statistical analysis of the various application domains of STGNN in urban computing. Fig. 4 illustrates the main application domains of STGNN, which encompass transportation, safety, environment, and public health. Among these, transportation is the most widely studied application domain of STGNN, constituting over 60% of the existing literature.

A. Transportation

Modern urban systems have numerous sensors distributed across traffic road networks and critical regions to monitor changing traffic states, such as flow and speed. The objective of traffic state prediction is to forecast future traffic states based on historical traffic states within a particular spatial range. Traffic state prediction can be divided into two main categories:

- *Network-based prediction:* The object of network-wide prediction is usually the traffic flow or speed on the given road networks [17], [20], [26], [27], [28], [29]. The basic graph structures can be directly converted from road networks in most previous works.
- *Region-based prediction:* This task aims to forecast the traffic (e.g., crowd flow) in urban areas [30], [31], [32], [33]. In this case, the whole urban area is partitioned into irregular or regular regions, and a spatio-temporal graph can be constructed based on the distances, connectivity, semantic correlations between different regions, etc.

In general, traffic state prediction tasks can be summarized in the following form:

$$\left[\mathcal{X}_{(t-T'+1)}, \ldots, \mathcal{X}_{(t)}; \mathcal{G}\right] \xrightarrow{f(\cdot)} \left[\mathcal{X}_{(t+1)}, \ldots, \mathcal{X}_{(t+T)}\right], \quad (5)$$

where $\mathcal{X}_{(t)} \in \mathbb{R}^{N \times d}$ denotes the traffic states of N vertices at time step t, $\mathcal{G}$ is the constructed graph structure, $f(\cdot)$ is the corresponding STGNN model for making predictions.

1) Traffic Demand Prediction: Accurately predicting urban traffic demand patterns (e.g., taxi demands, rail transit passenger demands, and bike-sharing demands) in various regions can facilitate traffic scheduling to alleviate congestion during rush hours. Demands can be broadly categorized into three main types: origin demands, destination demands, and origin-destination (OD) demands. Predicting origin and destination demands is similar to region-based traffic state prediction, *i.e.*, forecasting future demands based on historical demands in N regions [21], [25], [34], [35]. However, OD demand prediction is somewhat distinct, requiring prediction of future origin-destination matrices using historical OD matrices [36], [37], [38], [39], [40], [41]. To be specific, the outputs of OD demand prediction are a series of matrices with size $N \times N$, which can characterize the flow demand among these region pairs.

2) Traffic Incident Prediction: With the dramatic increase in the number of vehicles, more and more traffic incidents such as congestion and accidents have occurred, placing significant pressure on urban traffic management. The aim of the traffic incident prediction task is to predict some important properties (e.g., occurrence probability, occurrence time) of these incidents that may occur on road networks [42], [43], [44], [45], [46]. In addition to differences in the objects being predicted, similar to the traffic state prediction task, accurate traffic incident prediction also requires capturing spatio-temporal dependencies on road networks by building STGNN models.

3) Travel Time Prediction: Travel time prediction is highly valued in industries, especially in online map navigation and ride-hailing software, which can significantly enhance the user experience. This task aims to predict the travel time of a given trajectory based on historical traffic states on road networks. To predict travel time more accurately, not only trajectory characteristics need to be considered, but also the spatio-temporal dynamics (e.g., flow, speed) attached to road networks should be addressed. Under this circumstance, spatio-temporal graphs are established based on road networks. So far, large technology companies such as Baidu [47], [48], Google [49], and DiDi [50] have developed practical travel time prediction functions on their online platforms.

4) Trajectory Prediction: Trajectory prediction is a crucial task for comprehending the intricate group dynamics of humans and vehicles [78], [80], [82], [106], [107], [108], which fosters advancements in autonomous driving and urban monitoring technologies. There are some correlations or interactions in the movement patterns of agents in the group, thus we can build spatio-temporal graphs based on the relations between different agents within a group. Upon creating these spatio-temporal graphs, STGNNs can be devised to predict the coordinates that agents may occupy in the future, considering their historical traversal coordinates, thereby facilitating the predictions of future trajectories.

B. Environment

1) Air Quality Prediction: Air quality has become a pressing issue that needs immediate attention and improvement. Accurate air quality prediction can not only assist governments in formulating energy-saving and emission-reduction policies but also provide guidance for residents' outdoor activities. Air quality index (AQI), PM2.5, and emissions are the indicators we are among the most significant indicators of concern. The related data are collected by city-level or national-level monitoring stations [85], [109]. Due to the fluidity of the air, monitoring stations that are geospatially close or sharing the same wind direction may collect correlated results [110], [111], [112]. Hence, utilizing STGNN models can not only establish such spatial dependencies but also capture the time-varying dynamics of air quality.

2) Meteorological Prediction: Meteorological forecasting is another research topic intimately connected to the environment and human society. Similar to air quality data, meteorological data are also collected by distributed monitoring stations. However, the correlations between different stations could be more complex and susceptible to a greater number of factors. In recent years, STGNN-based approaches have been progressively applied in various meteorological prediction scenarios such as temperature prediction [86], [113], [114], frost prediction [115] and wind prediction [87], [88], [116], showcasing their superior performance in practice.

C. Public Safety

1) Crime Frequency Prediction: Effectively combating and preventing crime is the foundation for ensuring urban safety. Accurate prediction of crime frequency can assist governments in understanding real-time crime dynamics and allocating police resources rationally. Most existing work in this research line focus on crime frequency prediction in urban areas. Given that different urban regions have distinct functions, POI, and other characteristics, these factors could contribute to varying crime types and trends. However, regions with similar characteristics or close distances may exhibit latent correlations in crime incidents [89], [117], [118]. Consequently, many previous studies [89], [91], [93], [98], [117], [119], [120], [121] have introduced a series of STGNN to capture these correlations to reduce the prediction errors.

2) Disaster Situation Prediction: Natural disasters, e.g., earthquakes, have posed big challenges to the safety of human society since ancient times. Disaster situation prediction can enable governments to implement disaster prevention measures, allocate disaster relief materials, and evacuate residents in a timely manner. STGNNs have been a fruitful approach in this task to model correlated and heterogeneous features across geographical locations. Currently, literature has introduced the STGNN models into scenarios such as flood prediction [122], [123], fire prediction [94], [124], typhoon forecasting [95], [96], [125] and earthquake prediction [97], [126], [127].

D. Public Health

1) Epidemic Prediction: Epidemics are one of the greatest challenges to the public health systems, especially the novel coronavirus that has been prevalent in recent years, which has caused more than six million deaths worldwide. Therefore, accurately predicting the spread of epidemics is an important but challenging task, which can provide data support for the strengthening strategy of the urban public health systems. Some recent existing works have employed STGNN models to address the national-level [100], [102], [103], [128], [129], [130] or

TABLE II
PUBLIC DATASETS FOR MAIN APPLICATION DOMAINS

Domain	Dataset	Link	Reference
Transportation	California-PEMS	http://pems.dot.ca.gov/	[17], [23], [51]–[54]
	METR-LA	https://www.metro.net/	[20], [28], [55]–[59]
	NYC taxi	https://www1.nyc.gov/site/tlc/about/tlc-trip-record-data.page	[32], [33], [35], [60], [61]
	San Francisco taxi	https://crawdad.org/ crawdad/epfl/mobility/20090224/	[26], [62]
	T-Drive Taxi	https://www.microsoft.com/en-us/research/publication/t-drive-trajectory-data-sample/	[63], [64]
	NYC bike	https://www.citibikenyc.com/sytem-data	[19], [31], [61], [64]–[66]
	Chicago bike	https://www.divvybikes.com/system-data	[19], [67], [68]
	NYC accident	https://data.cityofnewyork.us/	[42], [45], [69], [70]
	Chicago accident	https://data.cityofchicago.org/	[42], [71]
	Chengdu taxi trajectory	http://www.dcjingsai.com./	[72]–[76]
	Porto taxi trajectory	https://www.kaggle.com/crailtap/taxi-trajectory.	[74]–[77]
	ETH walking pedestrians	https://data.vision.ee.ethz.ch/cvl/aem/ewap_dataset_full.tgz	[78]–[83]
	UCY walking pedestrians	https://graphics.cs.ucy.ac.cy/research/downloads/crowd-data	[78]–[80], [83]
Environment	Beijing air quality	https://biendata.com/competition/kdd_2018/data/	[84], [85]
	Shanghai air quality	http://www.cnemc.cn/en/	[85]
	WeatherBench	https://mediatum.ub.tum.de/1524895	[86]
	Denmark wind speed	https://sites.google.com/view/siamak-mehrkanoon/code-data	[87], [88]
	Dutch wind speed	https://github.com/HansBambel/multidim_conv	[87]
Public safety	NYC crime	https://data.cityofnewyork.us/	[89]–[91]
	Chicago crime	https://data.cityofchicago.org/	[89]–[93]
	San Francisco crime	https://datasf.org/openda	[67], [94]
	San Francisco fire	https://datasf.org/openda	[67], [94]
	Japan typhoon	http://agora.ex.nii.ac.jp/digital-typhoon/	[95], [96]
	California earthquake	https://service.iris.edu/	[97], [98]
Public health	US Covid-19	https://github.com/CSSEGISandData/COVID-19	[99]–[103]
	Italy Covid-19	https://github.com/pcm-dpc/COVID-19	[104]
	Japan-Prefectures ILI	https://tinyurl.com/y5dt7stm	[105]
	US ILI	https://tinyurl.com/y39tog3h	[105]

international-level [131] epidemic prediction tasks. Many of them combine the mathematical formulations of epidemic dynamics and the modeling of spatio-temporal graphs, which have achieved better prediction results than traditional methods [104], [130], [132], [133].

2) Ambulance Demand Prediction: In today's aging society, the allocation of ambulance resources is a challenging task that needs careful consideration. Accurate ambulance demand prediction can effectively alleviate the burden on the urban healthcare systems. Since there could be time-varying correlations in public medical resources, traffic conditions, and demand patterns among different regions of the social systems, STGNN-based methods have increasingly been exploited to learn these multi-view spatial correlations in recent years [134], [135], [136].

E. Other Application Domains

In addition to the four main application domains mentioned above, other scenarios where the spatio-temporal graph structures can be established based on the intrinsic relations of data are potential areas for the development of STGNN-based predictive learning models. In recent years, STGNN-based predictive learning models have also been promoted to other domains such as energy, economy, finance, and production. In the energy domain, STGNN models have been utilized in wind power prediction [137], [138] and photovoltaic power prediction [139]. In economy, a typical application is nation-level regional economy prediction, where researchers have explored the usage of STGNN models [140], [141].

F. Open Datasets and Benchmarks

As depicted in Table II, we have compiled a list of some of the most frequently used public datasets from previous works in the primary application domains, including their details such as source links and related publications. These datasets are widely used in the research field of transportation and urban mobility, particularly for traffic state prediction. Due to their high granularity, realistic nature, and real-world applicability, they serve as valuable resources for researchers working on spatio-temporal forecasting and traffic modeling.

There are also some well-known benchmarks in these primary application domains, especially in traffic prediction, such as BasicTS,[1] Traffic-Benchmark,[2] DL-Traff[3] and LargeST.[4] For other application domains, there are fewer benchmarks, but there are some as follows, e.g., weatherbench2[5] for meteorological forecasting and PM2.5-GNN[6] for air quality forecasting.

V. BASIC NEURAL ARCHITECTURES

Here we introduce basic neural architectures for STGNNs. As shown in Fig. 5, the basic framework of STGNNs for predictive learning contains three main modules: Data Processing Module (DPM), Spatio-Temporal Graph Learning Module (STGLM), and Task-Aware Prediction Module (TPM). For predictive learning tasks in urban computing, DPM is responsible for constructing the spatio-temporal graph data from the raw data; STGLM seeks to capture hidden spatio-temporal dependencies from complex social systems, while TPM aims to map the spatio-temporal hidden representations from STGLM into the space of downstream prediction tasks. STGLM serves as the most vital component of STGNNs, which usually combines

[1] https://github.com/zezhishao/BasicTS
[2] https://github.com/tsinghua-fib-lab/Traffic-Benchmark
[3] https://github.com/deepkashiwa20
[4] https://github.com/liuxu77/LargeST
[5] https://github.com/google-research/weatherbench2
[6] https://github.com/shuowang-ai/PM2.5-GNN

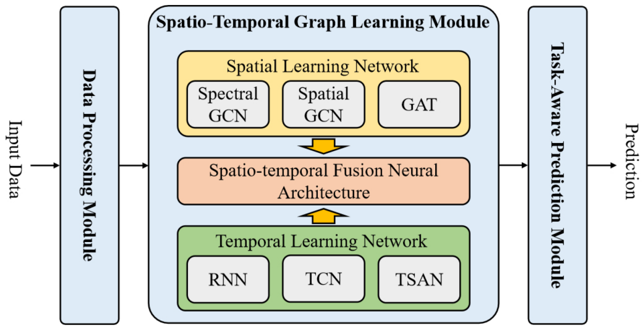


Fig. 5. Basic framework of STGNNs.

spatial learning networks and temporal learning networks organically through a certain spatio-temporal fusion method. Spatial learning networks may utilize spectral graph convolutional networks (Spectral GCNs) [142], spatial graph convolutional networks (Spatial GCNs) [11], [143], and graph attention networks (GATs) [144] as potential options. Temporal learning networks, on the other hand, may incorporate recurrent neural networks (RNNs), temporal convolutional networks (TCNs), or temporal self-attention networks (TSANs). Compared with STGLM, TPM is a relatively simple neural network, thus the majority of existing research focuses on the design of the neural architectures in STGLM.

A. Graph Neural Networks

Graph neural networks (GNNs) are fruitful tools for learning spatial dependencies in non-Euclidean space. In recent years, popular GNNs can be divided into three categories: spectral GCNs, spatial GCNs and GATs.

1) Spectral Graph Convolutional Network: Initially, most GNNs were based on the Fourier transform, which converts the graph signal in the spatial domain into the spectral domain to conduct convolution calculations [145]. The notation $\odot$ is the convolution operator, $\mathbf{U}$ denotes the matrix of eigenvectors of the normalized graph Laplacian and $\mathbf{U}^T\mathbf{g}$ is the filter in the spectral domain. The graph convolution operation is defined as:

$$\boldsymbol{g}_w \star \boldsymbol{x} = \boldsymbol{U}\boldsymbol{g}_w\boldsymbol{U}^T\boldsymbol{x}. \tag{6}$$

Most of the subsequent GNNs based on the spectral domain mainly improve the calculation method of $\boldsymbol{g}_w$. For example, ChebNet [142] is one of the most popular Spectral GNN methods. According to the theory that $\boldsymbol{g}_w$ can be approximated by a truncated expansion of Chebyshev polynomials [146].

2) Spatial Graph Convolutional Network: While spectral graph convolutional networks (GCNs) have made significant advancements, their primary limitation lies in their dependence on the graph Laplacian matrix. Whenever there is a change in the underlying graph structure, the graph Laplacian matrix must be recomputed, rendering spectral GCNs better suited to scenarios where the graph structure remains constant. To overcome the dependency on the graph Laplacian matrix, Kipf et al. simplify the graph convolution operation [11] by performing message passing in the spatial domain. We called this new form as spatial GCNs, which is defined as:

$$\boldsymbol{g}_w \star \boldsymbol{x} = w\left(\boldsymbol{I}_N + \boldsymbol{D}^{-\frac{1}{2}}\boldsymbol{A}\boldsymbol{D}^{-\frac{1}{2}}\right)\boldsymbol{x}, \tag{7}$$

where $\boldsymbol{A}$ is the adjacency matrix; $\boldsymbol{D}$ is the degree matrix; w is the learnable parameters in the spatial GCN.

3) Graph Attention Network: To account for the importance of neighbor nodes in learning spatial dependencies, GAT [144] integrates the attention mechanism into the node aggregation operation as:

$$\begin{aligned} \boldsymbol{h}_v^{t+1} &= \rho\left(\sum_{u\in\mathcal{N}_v}\alpha_{vu}\boldsymbol{W}\boldsymbol{h}_u^t\right), \\ \alpha_{vu} &= \frac{\exp\left(\text{LeakyReLU}\left(\boldsymbol{a}^T\left[\boldsymbol{W}\boldsymbol{h}_v\|\boldsymbol{W}\boldsymbol{h}_u\right]\right)\right)}{\sum_{k\in\mathcal{N}_u}\exp\left(\text{LeakyReLU}\left(\boldsymbol{a}^T\left[\boldsymbol{W}\boldsymbol{h}_v\|\boldsymbol{W}\boldsymbol{h}_k\right]\right)\right)}, \end{aligned} \tag{8}$$

where α_{vu} denotes the attention scores of neighbor node u to the central node v, W is the weight matrix associated with the linear transformation for each node, and $\mathbf{a}$ is the weight parameter for attention output.

B. Recurrent Neural Networks

Recurrent neural networks (RNNs) are a class of deep neural networks for sequential learning based on recursive computations, and have found extensive applications in time series modeling. However, the vanilla version of RNNs is subject to a significant limitation – the gradient vanishing or explosion problem during the training process [147]. In response to this challenge, two of the most prominent variants of RNNs, *i.e.*, Long Short-Term Memory (LSTM) [148] and Gated Recurrent Units (GRU) [149], have been proposed. So far, GRU is the most widely used variant because it takes into account both high performance and low computational complexity.

GRU only has two efficient gated computational units: an update gate and a reset gate. $\mathbf{u}_t$ represents the update gate, which determines how to combine the information of the new input time step with the memory of the previous time step. $\mathbf{r}_t$ represents the reset gate, which defines the amount of memory reserved from the previous time step to the current time step. Although the learnable parameters of GRU are streamlined, its performance can be compared with LSTM in previous works, while improving the training and inference efficiency. The calculation process of GRU is defined as follows:

$$\begin{aligned} \boldsymbol{u}_t &= \sigma(\boldsymbol{W}_u \cdot x_t + \boldsymbol{U}_u \cdot \boldsymbol{C}_{t-1} + \boldsymbol{b}_u), \\ \boldsymbol{r}_t &= \sigma(\boldsymbol{W}_r \cdot x_t + \boldsymbol{U}_r \cdot \boldsymbol{C}_{t-1} + \boldsymbol{b}_r), \\ \tilde{\boldsymbol{C}}_t &= \tanh(\boldsymbol{W}_C \cdot x_t + \boldsymbol{U}_C(\boldsymbol{r}_t \odot \boldsymbol{C}_{t-1}) + \boldsymbol{b}_C), \\ \boldsymbol{C}_t &= \boldsymbol{u}_t \odot \boldsymbol{C}_{t-1} + (1-\boldsymbol{u}_t) \odot \tilde{\boldsymbol{C}}_t. \end{aligned} \tag{9}$$

where $\tilde{\mathbf{C}}_t$ represents the candidate state of the current GRU unit after the calculation of the reset gate, $\mathbf{C}_t$ represents the state of the GRU unit after the calculation of the update gate.

C. Temporal Convolutional Networks

RNNs have been extensively applied for temporal learning in many spatio-temporal tasks, but their disadvantage is readily apparent: the recurrent structures necessitate the computation of sequences at every time step, leading to a substantial increase in computational cost and a consequent decrease in model efficiency. In contrast, Temporal Convolutional Networks (TCN) with their parallel 1D-CNN structures can address this problem effectively.

1) Gated Temporal Convolutional Network: Inspired by the gated mechanism in LSTMs and GRUs, we can also integrate it with pure 1D-CNN architecture to enhance the capability of temporal learning. We called this hybrid neural architecture a gated temporal convolutional network (Gated-TCN) [150]. The calculation process of Gated-TCN is defined as follows:

$$F(x) = tanh(\mathbf{\Theta_1} \star x) \odot \sigma(\mathbf{\Theta_2} \star x), \tag{10}$$

where $\mathbf{\Theta_1}$ and $\mathbf{\Theta_2}$ represent the learnable parameters of the convolution kernel in two different 1D-CNNs, respectively; $\star$ denotes the convolution operation; $\odot$ is the element-wise multiplication mechanism; $\sigma(\mathbf{\Theta_2} \star x)$ indicates the gating unit, which is utilized to control the utilization rate of historical information.

2) Causal Temporal Convolutional Network: While TCN is an efficient parallel neural architecture for sequential learning, it violates the temporal order of spatio-temporal graph data. Compared to standard TCNs, Causal TCNs, which are proposed in Wavenet [151], offer the additional benefit of explicitly modeling the causal nature of temporal data. This is achieved by removing connections between future time steps and past time steps, which eliminates the possibility of data leakage from future time steps to past time steps. Furthermore, in order to more effectively capture longer-range temporal dependencies, 1D-CNN with dilated factors [152] can be used. By increasing the dilated factors layer by layer, this model has the capacity to learn temporal dependencies from a short range to a long range. A Causal TCN with dilated factors can be expressed as:

$$F(s) = (\boldsymbol{x} *_d f)(s) = \sum_{i}^{k-1} f(i) \cdot x_{s-d \cdot i}, \tag{11}$$

where s is the input time series; d represents the dilation factor, and the ordinary convolution operator is a special case of the dilated convolution operator when $d = 1$; $s - d \cdot i$ refers to the positioning of certain historical information.

D. Temporal Self-Attention Networks

Self-attention networks represent a highly effective approach for capturing long-range temporal relationships among different time steps, with the most prominent example being the Transformer model [153]. The Transformer comprises three primary components: a scaled dot-product attention network, a feed-forward network, and position encodings. The scaled dot-product network is the core part of Transformers, in which the attention calculation is formulated as:

$$\text{Attention}(\boldsymbol{Q}, \boldsymbol{K}, \boldsymbol{V}) = \text{Softmax}\left(\frac{\boldsymbol{Q}\boldsymbol{K}^T}{\sqrt{\boldsymbol{d}_k}}\right)\boldsymbol{V}, \tag{12}$$

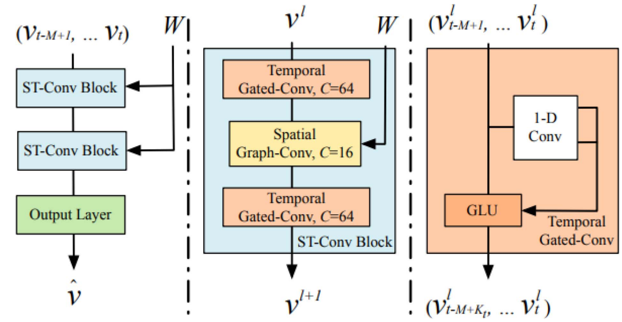


Fig. 6. Overview[7] of STGCN [20].

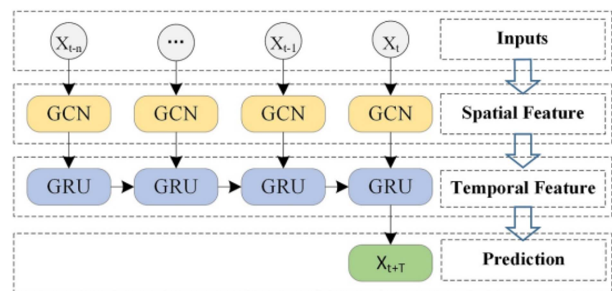


Fig. 7. Overview of T-GCN [26].

where queries $\boldsymbol{Q}$, keys $\boldsymbol{K}$, values $\boldsymbol{V}$ are three basic elements in the self-attention mechanism, which are obtained by non-shared linear transformations from the original input. $\boldsymbol{d}_k$ denotes the scaling factor, whose value is equal to the dimension of the model. Since Transformer contains no recurrence or convolution operator, we have to inject some positional information (eg., trigonometric function-based encoding) about the tokens in the sequence to consider the order of the sequence.

E. Spatio-Temporal Fusion Neural Architecture

In addition to spatial learning networks and temporal learning networks, spatio-temporal fusion neural architecture represents another critical area, as it determines how spatial learning networks and temporal learning networks are integrated into the complete STGNN. Existing fusion neural architectures can be divided into two categories – factorized or coupled neural architecture.

1) Factorized Neural Architecture: In factorized neural architectures, spatial learning networks and temporal learning networks are stacked in parallel or serially like building blocks layer by layer. There are two typical examples for factorized neural architectures in STGNN models, as shown in Figs. 6 and 7, respectively. The first example is STGCN [20], whose temporal learning network is TCN. In each ST-Conv block of STGCN, two TCNs and one GCN are stacked in series, forming a sandwich structure. Since this model learns temporal information through convolutional structures, its spatio-temporal learning method

[7]The figures used in this paper but are not drawn by us (including Figs. 6 to 13) are all with the permission of their authors.

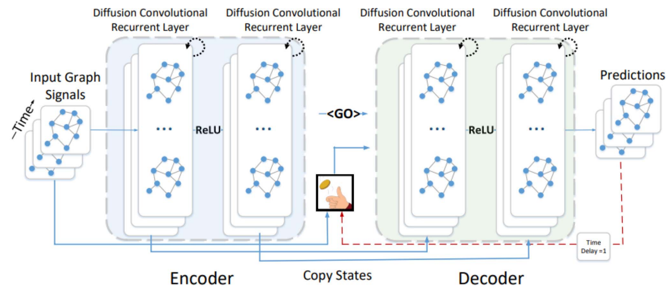


Fig. 8. Overview of DCRNN [28].

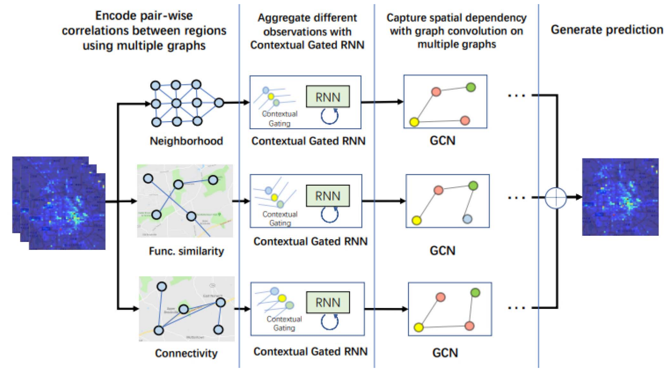


Fig. 9. Overview of STMGCN [25].

is parallelized, *i.e.*, it receives all information of a given time window length as input at the same time. Mathematically, the calculation of each ST-Conv block in this model can be defined as follows::

$$v^{l+1} = \boldsymbol{\Gamma_1^l} *_{\mathcal{T}} \operatorname{ReLU}\left(\boldsymbol{\Theta^l} *_{\mathcal{G}} \left(\Gamma_0^l * \mathcal{T} v^l\right)\right), \tag{13}$$

where Γ_0^l and Γ_1^l denote the upper and lower temporal convolutional kernel within block l, and Θ^l is the spectral kernel of graph convolution.

The second one is T-GCN [26], which utilizes GRUs for temporal learning. This model captures the spatio-temporal dependencies in a recursive manner. For each time step, graph signals are sequentially processed by GCN and GRU to learn spatial and temporal dependencies separately. The whole process of each stacked GCN and GRU in this model can be expressed as:

$$\begin{aligned}
\boldsymbol{f}(\boldsymbol{X}, \boldsymbol{A}) &= \sigma\left(\boldsymbol{A}\boldsymbol{X}\boldsymbol{W}_0\right), \\
\boldsymbol{u}_t &= \sigma\left(\boldsymbol{W}_u\left[\boldsymbol{f}\left(\boldsymbol{A}, \mathcal{X}_t\right), \boldsymbol{H}_{t-1}\right] + \boldsymbol{b}_u\right), \\
\boldsymbol{r}_t &= \sigma\left(\boldsymbol{W}_r\left[\boldsymbol{f}\left(\boldsymbol{A}, \mathcal{X}_t\right), \boldsymbol{H}_{t-1}\right] + \boldsymbol{b}_r\right), \\
\boldsymbol{c}_t &= \tanh\left(\boldsymbol{W}_c\left[\boldsymbol{f}\left(\boldsymbol{A}, \mathcal{X}_t\right), \left(\boldsymbol{r}_t * \boldsymbol{H}_{t-1}\right)\right] + \boldsymbol{b}_c\right), \\
\boldsymbol{H}_t &= \boldsymbol{u}_t * \boldsymbol{H}_{t-1} + (1 - \boldsymbol{u}_t) * \boldsymbol{c}_t,
\end{aligned} \tag{14}$$

where $f(A, \mathcal{X}_t)$ denotes the output of spatial GCN at time step t. Then $f(A, \mathcal{X}_t)$ is put forward into GRU to obtain the hidden state at t.

2) Coupled Neural Architecture: In coupled neural architectures, spatial learning networks are usually integrated into the architecture of temporal learning networks as embedded components. In STGNN, this type of neural architecture occurs almost exclusively in combinations of GNN-based spatial learning networks and RNN-based temporal learning networks. One example of a coupled neural architecture in STGNNs is the DCRNN [28], which integrates GCN into the architecture of GRU, as illustrated in Fig. 8. In this model, the original linear units in LSTM are replaced with a graph convolution operator, which can be written as:

$$\begin{aligned}
\boldsymbol{r}_t &= \sigma\left(\boldsymbol{\Theta}_r \star \mathcal{G}\left[\mathcal{X}_t, \boldsymbol{H}_{t-1}\right] + \boldsymbol{b}_r\right), \\
\boldsymbol{u}_t &= \sigma\left(\boldsymbol{\Theta}_u \star \mathcal{G}\left[\mathcal{X}_t, \boldsymbol{H}_{t-1}\right] + \boldsymbol{b}_u\right), \\
\boldsymbol{C}_t &= \tanh\left(\boldsymbol{\Theta}_C \star \mathcal{G}\left[\mathcal{X}_t, \left(\boldsymbol{r}_t \odot \boldsymbol{H}_{t-1}\right)\right] + \boldsymbol{b}_c\right), \\
\boldsymbol{H}_t &= \boldsymbol{u}_t \odot \boldsymbol{H}_{t-1} + (1 - \boldsymbol{u}_t) \odot \boldsymbol{C}_t,
\end{aligned} \tag{15}$$

where $\boldsymbol{\Theta}_r \star \mathcal{G}$ denotes the graph convolution operator with parameter $\boldsymbol{\Theta}_r$. Compared with the equation 9 of the original GRU, we can find that except for the internal graph convolution operator, the external calculation methods of the recurrent network are not much different. Similar to some neural translation models [154], DCRNN can also employ sequence-to-sequence structure to improve predictions.

VI. STGNN Variants

In Section V, we have introduced the basic neural architectures of STGNNs, thereby augmenting the comprehension of the spatio-temporal learning paradigm within this research domain. However, in recent years, there have been numerous innovative methods devised to enhance the learning of spatio-temporal dependencies. In this section, we elaborate on some advanced STGNN variants that can better capture spatio-temporal dependencies for predictive learning in urban computing.

A. Spatial Learning Methods

1) Multi-Graph Convolution: In urban systems, there are often multiple types of spatial relations that exist simultaneously. For instance, in transportation systems, adjacent regions and regions with similar POIs may exhibit similar traffic patterns. Hence, jointly considering multiple spatial relations is necessary for spatio-temporal learning in STGNN. In recent years, a series of STGNN variants that integrate multi-graph convolutions have been proposed to address this challenge [19], [21], [22], [25], [34], [155], [156], [157]. Among them, STMGCN [25] is a typical model for urban ride-haling demand prediction, as shown in Fig. 9. This model first constructs multi-graph based on neighborhood, function similarity, and connectivity to characterize multiple spatial correlations. For each graph, contextual gated RNN and ChebNet are respectively adopted to capture temporal and spatial dependencies. Finally, the final prediction is obtained by fusing the parallelized multi-graph spatio-temporal hidden information.

2) Adaptive Graph Learning: Despite its capability to capture multiple spatial correlations to some extent, multi-graph modeling still suffers from two limitations. Firstly, the graph construction process may be insufficient and fail to account for other implicit correlations. Secondly, the rationality of graph construction may be questioned, particularly in the absence of sufficient domain knowledge to support it. To overcome these challenges, adaptive graph learning methods have been developed gradually. According to existing literature, adaptive graph learning methods in STGNN can be broadly categorized into two main categories: random initialization-based and feature initialization-based approaches.

Random initialization-based methods perform adaptive graph structure learning via randomly initialized learnable matrices [35], [57], [59], [91], [158], [159], [160], [161]. Two prominent models in this category are Graph WaveNet [57] and MTGNN [159], which have been widely applied or improved upon in subsequent works. In Graph WaveNet, the adaptive graph is produced as follows:

$$\tilde{\boldsymbol{A}}_{adp} = \mathrm{SoftMax}\left(\mathrm{ReLU}\left(\boldsymbol{E}_1\boldsymbol{E}_2^T\right)\right), \tag{16}$$

where $\boldsymbol{E}_1 \in \mathbb{R}^{N\times C}$ and $\boldsymbol{E}_2 \in \mathbb{R}^{N\times C}$ are source node embedding and target node embedding, respectively. They are two learnable matrices with the random initialization, where N denotes the number of nodes in the graph and C denotes the dimension of the embedding.

In contrast, the generation process of the adaptive graph in MTGNN is defined as:

$$\boldsymbol{M}_1 = \tanh\left(\alpha\boldsymbol{E}_1\boldsymbol{\Theta}_1\right), \quad \boldsymbol{M}_2 = \tanh\left(\alpha\boldsymbol{E}_2\boldsymbol{\Theta}_2\right), \tag{17}$$

$$\tilde{\boldsymbol{A}}_{adp} = \mathrm{ReLU}\left(\tanh\left(\alpha\left(\boldsymbol{M}_1\boldsymbol{M}_2^T - \boldsymbol{M}_2\boldsymbol{M}_1^T\right)\right)\right), \tag{18}$$

where $\boldsymbol{E}_1 \in \mathbb{R}^{N\times C}$ and $\boldsymbol{E}_2 \in \mathbb{R}^{N\times C}$ represent two randomly initialized node embeddings; $\boldsymbol{\theta}_1$ and $\boldsymbol{\theta}_2$ are learnable parameters within the model; α is a hyperparameter for controlling the saturation rate of the activation function. Numerous subsequent random initialization-based adaptive graph learning methods are proposed based on the two aforementioned methods. For example, CCRNN [35] introduced a layer-wise adaptive graph learning mechanism to adjust the graph structures layer by layer. DMSTGCN [59] presented an adaptive graph learning approach with tensor decomposition.

Feature initialization-based approaches aim to construct adaptive graph structure learning based on the given inputs or the hidden states [58], [66], [94], [162], [163], [164]. These models usually adopt learnable matrices or attention mechanism to incorporate with the given features for generating the adaptive graph structures. For example, DGCRN [58] proposed a recurrent adaptive graph learning mechanism based on the hidden states to construct the graph structures for each time step. GTS [164] presented a novel probabilistic graph structure learning method based on input features.

3) Muti-Scale Spatial Learning: Due to the wide existence of spatial heterogeneity in urban systems, entities can be divided into communities with different functions. Entities in the same community may have inter-community correlations, while entities in different communities could also have cross-community correlations. In light of these facts, some recent methods have investigated multi-scale spatial learning based on community partitioning. These methods often leverage domain knowledge to guide the community partitioning process.

In this research line, some studies obtain partitioned communities by artificial division [45], [69] or clustering algorithms [21], [68], [165] while others obtain them by neural networks [89], [166], [167]. For example, ST-SHN [89] and ST-HSL [90] learn the hyperedges, *i.e.* communities, of the hypergraph to capture global spatial dependencies for crime prediction. Besides, GAGNN [166] is a group-aware STGNN model for air quality prediction among hundreds of Chinese cities. This model first proposed a differentiable grouping network for learning the assignment matrix, which automatically computes the mapping relationships between cities and city groups. Another notable research line involves THINK [168] and DMGCRN [169], which utilize hyperbolic graph neural networks on the Poincare ball to capture multi-scale spatial dependencies more directly. The hyperbolic space is particularly suitable for modeling hierarchies, including local and global dependencies of spatio-temporal data, which makes it a promising approach for improving STGNN models.

4) Heterogeneous Spatial Learning: As mentioned in the introduction section, heterogeneity is an essential property of spatio-temporal data in smart cities wherein it displays varying patterns across various temporal or spatial ranges. Different from the above multi-scale spatial learning methods, some works focused on the fine-grained node-to-node heterogeneous relationships in the spatio-temporal data. To distinguish between the influence of static undirected edges (e.g., distance-based edges) and the dynamic directed edges (e.g., vehicle's mobility-caused edges) in the spatio-temporal graph, HMGCN [134] performs heterogeneous aggregation on the spatial dimension. Similarly, MasterGNN [85] constructs a heterogeneous graph structure based on multiple relations between air quality and weather monitoring stations, while HTGNN aggregates heterogeneous information from spatial-based intra-edges, temporal-based inter-edges, and spatio-temporal-based across-time-edges. Another line of heterogeneous spatial learning utilizes transportation, time, and geographical information to capture intricate spatio-temporal message passing. For instance, HeGA [170] and MOHER [171] design multiple transportation mode-based heterogeneous graphs to receive information from multi-sources at the same time, e.g., bike, bus, vehicle, etc.

B. Temporal Learning Methods

1) Multi-Scale Temporal Learning: Given the prevalence of short- and long-range correlations in spatio-temporal data, capturing multi-scale temporal correlations has emerged as a crucial direction for improving temporal learning. So far, there are two mainstream design directions for multi-scale temporal learning in STGNNs. The first direction utilizes TCNs with receptive fields of varying scales [88], [159]. A typical example is MTGNN [159] which employs multiple TCNs with various kernel sizes for learning temporal dependencies in different scales. The second direction involves integrating other temporal learning networks [18], [21], [52]. For example, DMVST-VGNN [21]

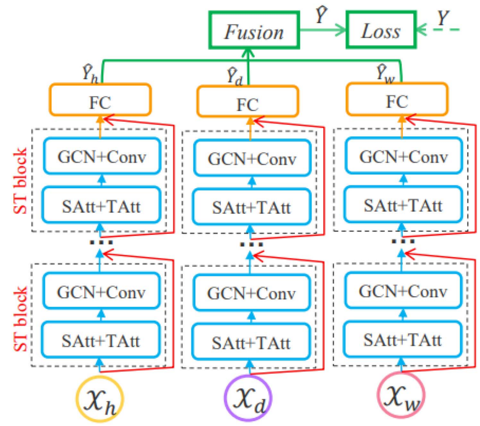


Fig. 10. Overview of ASTGCN [172].

jointly utilizes TCNs and Transformers for long-short range temporal learning.

2) Multi-Granularity Temporal Learning: There are multiple types of temporal characteristics in spatio-temporal data. For instance, the traffic flow at a given time is not only related to the recent traffic flow but may also exhibit similarities to the traffic flow at the same time on the previous day or even the previous week. This reflects the closeness, periodicity, and trend, respectively. To consider the temporal characteristics at these three granularities, many previous works [31], [32], [33], [172], [173] adopted a three-branch architecture to learn features from different temporal granularities separately, and then fuses the learned hidden states for predictions. As shown in Fig. 10, ASTGCN [172] employs a typical three-branch architecture for multi-granularity temporal learning, where $\mathcal{X}_h$, $\mathcal{X}_d$ and $\mathcal{X}_w$represent the spatio-temporal data for the latest one hour, the data of the same hour from the previous day, and the data of the same hour from the previous week, respectively. After going through separate branches, they are finally fused by the learnable weight matrix.

3) Decomposition Temporal Learning: Individual temporal patterns usually contain a variety of hidden components, such as inherent components, diffusion components, and periodic components. To better capture these complex temporal dependencies, decomposition-based temporal learning methods have been proposed, which can automatically decompose and integrate different temporal components through special neural designs [158], [174], [175], [176], [177]. FC-GAGA [174] is a noteworthy example of decomposition methods that adopt the subtraction residual from N-BEATS [178] to decompose different components in traffic data and model spatial correlations of each component. In addition to FC-GAGA, other works also adopted decomposition-based ideas. For example, StemGNN [176] decomposed the temporal components by the subtraction residual from N-BEATS, but modeled spatial correlations in the spectral domain. D2STGNN [158] proposed a temporal residual decomposition method to incorporate with graph structure learning. STWave [175] directly utilized the discrete wavelet transform to disentangle the event and the trend from the spatio-temporal graph data.

C. Spatio-Temporal Fusion Methods

1) Spatio-Temporal Joint Modeling: In Section V-E, we have discussed basic spatio-temporal fusion architectures of STGNNs, which are either factorized or coupled by spatial learning networks and temporal learning networks. Although these architectures can effectively learn spatial and temporal dependencies separately, they lack the ability to model the joint spatial-temporal dependencies, making it challenging to capture complex spatio-temporal relations across different time steps.

In recent years, some literature focused on jointly modeling spatial-temporal dependencies based on 3D GCN [179], Spatio-Temporal Joint GCN (STJGCN) [180] and Spatio-Temporal Synchronous GCN (STSGCN) [53]. Among them, STSGCN has become a mainstream method for spatio-temporal dependencies jointly fusion. This type of neural architecture enables the modeling of spatio-temporal dependencies in a unified graph structure, which can replace the separated spatial learning networks and temporal learning networks. The crucial part of STSGNN is the construction of the spatio-temporal synchronous graph. The original spatio-temporal synchronous graph is simple, whose nodes with the same location are connected to each other across adjacent time steps. This graph construction approach not only characterizes spatial neighbors, but also temporal neighbors, establishing unified spatio-temporal relations. After graph construction, STSGNN employs a simple GCN model to capture the spatio-temporal dependencies.

There are some follow-up works [23], [51], [52], [54], [181], [182], [183], [184] based on STSGCN to further improve the spatio-temporal synchronous graph modeling in recent years. For example, STFGNN [52] proposes to construct the spatio-temporal synchronous graph using not only topology-based but also similarity-based graphs, thereby making the spatio-temporal synchronous graph more informative. On the other hand, S2TAT [181] proposes a spatio-temporal synchronous Transformer framework that employs attention mechanisms to enhance the learning capability.

2) Automated Spatio-Temporal Fusion: Given the complexity of STGNNs, designing optimal neural architectures can be a challenging task. Existing spatio-temporal fusion methods are usually designed empirically, and may not generalize well to different data scenarios due to the various spatio-temporal attributes present in different scenarios. Neural architecture search (NAS) methods offer opportunities for automated spatio-temporal fusion in STGNNs, and have shown promising results in discovering optimal architectures for various applications.

AutoSTG [185] presented the first attempt to involve DARTS [186] (*i.e.*, the most classical gradient-based NAS method) into STGNN. In AutoSTG, the whole neural network is divided into different stacked cells, and these cells are the basic units to perform NAS. Following AutoSTG, a series of studies [23], [54], [75], [187], [188], [189], [190], [191] to integrate NAS into STGNNs in recent years. For example, AutoSTS [23] integrates NAS into the spatio-temporal synchronous graph neural networks for searching the optimal architecture of different GCNs and TCNs. Likewise, Auto-DSTSGN [54] also integrates NAS into the spatio-temporal synchronous graph neural networks, but focuses on searching the optimal adjacency matrices of spatio-temporal synchronous graphs.

VII. Advanced Learning Frameworks

In recent years, more and more advanced learning frameworks have been developed to enhance the performance of STGNN in terms of deep representation and prediction accuracy. In this section, we review and discuss some typical advanced learning frameworks that are combined with STGNNs.

A. Adversarial Learning

As traditional loss functions, such as L1 and L2 norms, are commonly used to measure prediction errors, they may lack the ability to capture the distribution and correlation between the predictions and real data. This limitation could potentially lead to distorted prediction results. Hence, the adversarial loss can be introduced to incorporate with the traditional loss for addressing this problem to some extent, which has been widely applied in time series prediction. To incorporate the adversarial loss, Generative Adversarial Networks (GANs) have been proposed, with the neural predictors as the generators and the neural architecture of discriminators designed separately. We have witnessed a blossom of works [85], [192], [193], [194], [195], [196] to combine the adversarial loss with STGNNs for predictive learning tasks.

B. Meta Learning

Meta learning is an advanced learning paradigm focusing on the concept of "learning to learn". Incorporating meta learning techniques in STGNN models is important since they can capture high-dimensional heterogeneity and dynamic spatio-temporal dependencies from raw data, and teaching them how to learn can significantly improve their prediction performance. Typically, meta-learning-based STGNNs involve extracting additional spatio-temporal attributes through a meta-learner. ST-MetaNet [197] (see Fig. 11) is the pioneering study to introduce meta learning into STGNNs, which is composed of RNN, Meta-GAT, and Meta-RNN, and utilizes two types of meta-knowledge learners, namely Node Meta-Knowledge (NMK) and Edge Meta-Knowledge (EMK) learners, to effectively incorporate additional spatio-temporal information.

In light of the success of ST-MetaNet, some other STGNN models have been proposed that incorporate meta learning. For example, ST-MetaNet+ [198] fuse the dynamic spatio-temporal state and meta-knowledge for weight generation of GAT and GRU. AutoSTG [185] also adopts a meta learning method similar to ST-MetaNet while introducing neural architecture search, using meta-knowledge to generate weight parameters for graph convolution and temporal convolution. MegaCRN [199] introduced an attention-based memory network, which stores the typical features in seen samples for further pattern matching, thus improving the capability of graph structure learning. In

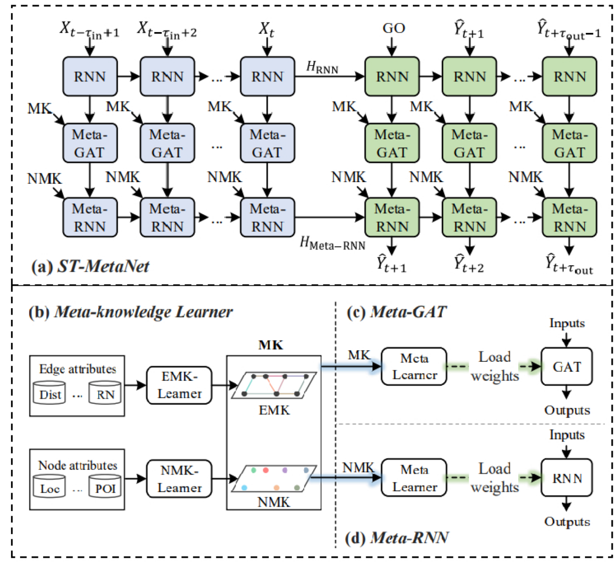


Fig. 11. Overview of ST-MetaNet [197].

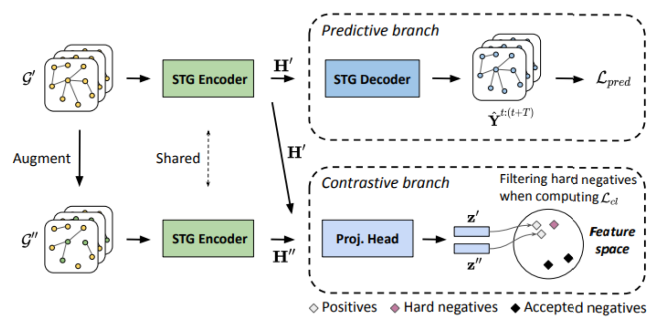


Fig. 12. Overview of STGCL [202].

addition, meta learning can also be used for spatio-temporal graph knowledge transfer in predictive learning scenarios [200], [201].

C. Self-Supervised Learning

Self-supervised learning is a type of method that transforms an unsupervised learning task into a supervised task by constructing its own labels. The goal of this learning paradigm is to learn better representations for downstream supervised tasks. By using self-supervised learning, a representation with strong generalization performance can be learned. Combining STGNN models with self-supervised learning can enhance the capability of spatio-temporal graph learning, which can improve the accuracy of downstream predictive learning tasks.

Contrastive learning is one of the most important self-supervised learning methods realized by constructing positive and negative samples, which has been introduced into STGNN models in recent years. One notable example is STGCL, which was introduced by Liu et al. [202] and was the first work to incorporate contrastive learning into STGNN architectures. As shown in Fig. 12, the first step of STGCL is the data augmentation to construct the positive and negative samples, where positive and negative samples are constructed using techniques such as edge masking, input masking, and temporal shifting. After obtaining the positive and negative samples, the same STG encoder is employed to learn the spatio-temporal graph representation for both the original data and the augmented data. Then, STGCL splits into two branches – a predictive branch and a contrastive branch. In the predictive branch, the STG decoder directly outputs the prediction results and traditional data point errors, such as mean absolute error (MAE), can be used as the loss function. In the contrastive branch, the two types of representation H' and H'' are put forward into the projection head to further obtain the latent representation z' and z''. For the two latent representation, the contrastive loss proposed in GraphCL [203] was adopted in this case

Based on STGCL, several other contrastive learning methods have been proposed to enhance the learning capabilities of STGNN in recent years. For example, SPGCL [204] proposed to learn the informative relations by maximizing the distinguishing margin between positive and negative neighbors for generating an optimal graph structure. ST-SSL [205] proposed an adaptive augmentation method over the spatio-temporal graph data at both attribute and structure levels. START [77] presented a spatio-temporal graph-based contrastive learning method for trajectory representation learning. This model proposed multiple negative trajectories construction methods such as trajectory trimming and road segments mask, to aid the STGNN model in achieving better performance in travel time prediction tasks.

D. Continuous Spatio-Temporal Modeling

Most existing STGNN-based approaches capture the spatial and temporal dependencies in a discrete way, leading to discontinuous latent state trajectories and higher prediction errors. To address this problem, some research has focused on continuous spatio-temporal modeling. Motivated by the success of Neural Ordinary Differential Equation (Neural-ODE) [206], a well-known approach for continuous system modeling, STGNNs combined with Neural-ODE can improve the capability of spatio-temporal graph representation learning in a continuous manner. STGODE [207] was the first attempt to introduce Neural-ODE into STGNNs, however, it only considers integrating Neural-ODE with GCN and neglects continuous modeling for temporal patterns. To achieve a joint continuous modeling for spatio-temporal dependencies, MTGODE [208] introduced the integration of Neural-ODE with graph convolution operators and temporal convolution operators to enable continuous spatio-temporal encoding. In addition, MixRNN+ [209] combined Neural-ODE and RNN for continuous recurrent hidden state modeling. STG-NCDE [210] developed a STGNN combined with the neural controlled differential equation (Neural-CDE) for better continuous modeling, compared with Neural-ODE-based methods.

In addition to epidemic prediction tasks, there are also a few works in other domains. For example, STDEN [27] proposed a unified framework that combines traffic potential energy field differential equations and neural networks for traffic flow prediction.

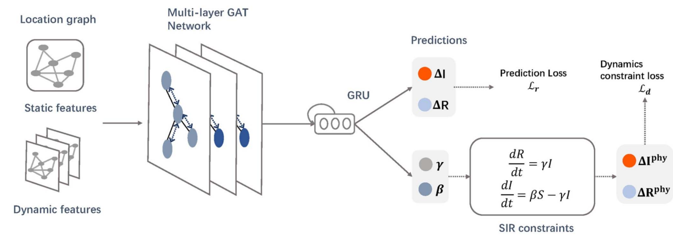


Fig. 13. Overview of STAN [130].

E. Physics-Informed Learning

In the last few years, a new paradigm called Physics-Informed Neural Networks (PINNs) [211] have emerged for exploring and computing real-world dynamics integrating physical differential equations and neural networks with powerful fitting capabilities. The main advantage of PINNs is their ability to enforce physical constraints on the predictions, thereby ensuring that the model's outputs are consistent with the laws of physics. Inspired by PINNs that are based on simple neural networks, physical-informed learning methods can be also combined with STGNNs, especially in epidemic prediction tasks [104], [130], [132], [133]. As shown in Fig. 13, STAN first integrates the constraints of SIR differential equations into the STGNN architecture. This model used GAT and GRU to capture the spatial and temporal dependencies respectively and performed a multi-task prediction. There are four components in the output of this model: transmission rate β, recovery rate γ, time-varying number of infections ΔI and recoveries ΔR. These components need to satisfy physical constraints based on the SIR equation.

F. Transfer Learning

Due to the scarcity of some spatio-temporal graph data, transfer learning techniques have become a cost-effective approach to extend the same basic STGNN model to different data scenarios. However, there are two main limitations in conducting transfer learning for STGNNs. The first one is the heterogeneity of spatial structures and the other one is the heterogeneity of temporal patterns in different circumstances. To be specific, in different scenarios, the spatial topology, relations, etc. are completely different as well as the temporal patterns such as periodicity and trend.

The existing literature on spatio-temporal graph transfer learning can be roughly divided into three categories: clustering-based [212], [213], [214], domain adaptation-based [215], [216] and meta-learning-based [200], [201]. For example, TL-DCRNN [212] proposed a graph partitioning method to divide the entire highway network into different sub-clusters and then used the DCRNN model to learn the spatio-temporal dependencies from source sub-clusters to target sub-clusters. DASTNet [215] combined the graph representation learning and multi-domains adversarial adaptation methods to obtain domain-invariant node embeddings, achieving the knowledge transfer among different scenarios with different spatial structures.

VIII. Challenges and Future Directions

We have investigated the applications, basic neural architectures, and recent advancements of STGNN for predictive learning in urban computing. Although STGNN models have achieved remarkable performance in recent years, there are still some challenging problems to be addressed, which point to potential future research directions. We summarize these challenges and suggest potentially feasible research directions as follows:

- *Lack of interpretability:* So far, the vast majority of STGNN-related work has focused on improving predictive performance through sophisticated model design. However, research on the interpretability of models has been relatively lacking, that is, we cannot clearly understand which spatio-temporal features take a leading role in improving predictive performance. In the most recent work, STNSCM [217] proposed to construct a causal graph to describe the bike flow prediction and analyze the causal relationship between the spatio-temporal features and prediction results. In addition to the spatial perspective, some deep time series models have incorporated statistical modeling techniques to enhance the understanding of predictive outcomes [218]. Hence, constructing interpretable STGNN models from both spatial and temporal perspectives is a potential direction.
- *Lack of calibration methods:* How to establish trust among frontline urban managers regarding the predictive capabilities of STGNNs is a practical problem that needs further exploration. Hence, the significance of uncertainty quantification that can reflect the trustworthiness of prediction results needs to be emphasized. In order to improve the trustworthiness of the deep models, appropriate model calibration methods are necessary, which have been widely used in image recognition [219] and graph representation learning [220] in recent years. At present, only works [221], [222] have studied the uncertainty of STGNN models, and there is a lack of research on calibration methods. Calibration for the STGNN models need to take into account the characteristics of spatial and temporal simultaneously, thus it is more challenging than previous related works.
- *Lack of physical constraints:* Most STGNN models capture the complex spatio-temporal dependencies through the integration of deep neural networks, while ignoring the consideration of physical constraints in different application domains, which makes the model less recognized in some professional fields. In recent years, although some STGNN models for epidemic prediction have combined professional differential equations as physical constraints [104], [130], [132], [133], such work is still lacking and needs to be improved in other application fields.
- *Lack of pre-training techniques:* Pre-training techniques have been greatly developed in the fields of time series and graph representation learning in recent years, but they are relatively lacking in STGNN-related work. In the most recent work, STEP [223] proposed a pre-training model combined with the Mask Auto-Encoder (MAE) [224] architecture to efficiently learn temporal patterns from very long-term history spatio-temporal graph data. In the future, pre-training techniques for long-range spatial and long-term temporal learning are necessary, which are of great value to the scalability and deployability of the STGNN models.

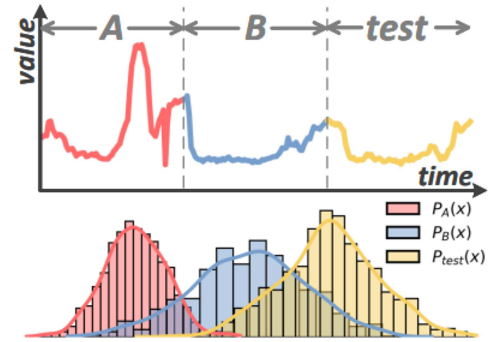


Fig. 14. Temporal Distribution Shift [228].

- *Lack of fine-grained NAS:* Due to the relatively complex components of STGNNs, automatically designing effective and reliable neural architectures is an urgent task. Although some existing works [54], [185], [189], [191] have proposed the integration of NAS and STGNNs, most of them are all limited to coarse-scale and lack of search for fine-grained architectures in GNN (eg, aggregation methods, activation functions, etc.). Therefore, inspired by some current state-of-the-art NAS methods for GNNs [225], [226], proposing efficient fine-grained NAS for STGNNs is a promising direction.
- *Hurdle of distribution shifts:* Spatio-temporal dataare often collected from various locations and time periods, resulting in significant differences in the distribution of the training, validation, and test sets. For instance, we visualize the temporal distribution on Beijing Air Quality dataset. As shown in Fig. 14, the training data (periods A and B) and test data derive from different distributions, namely $P_A(x) \neq P_B(x) \neq P_{test}(x)$. This can pose a challenge for STGNNs, as training a model on one dataset may not perform well on validation and test sets due to *distribution shifts*, which is similar to the distribution shift issue in domain adaptation (where the joint distribution of inputs and outputs differs between the training and test stages). Despite its importance, this problem has received less attention in the spatio-temporal research community. While several studies [227] investigated defeating distribution shifts in time series, they fail to encode the spatial correlations among locations.
- *Exploring new training strategies:* Previous studies have primarily focused on introducing novel STGNNs with sophisticated layers or modules to enhance human mobility analytics. However, another promising direction is to investigate new training strategies. For instance, in traffic prediction tasks, every location is treated equally, and the data belonging to these locations are jointly fed into

neural networks. Nevertheless, the complexity of modeling the spatio-temporal correlations of each location can vary significantly, necessitating a new training strategy such as curriculum learning. Curriculum learning trains a machine learning model on increasingly difficult data, starting from simpler data, and may be effective in addressing this issue.

- *Scalability issue:* One particularly challenging case for designing efficient STGNNs is when the number of locations in the sensor network is very large. For example, there are over ten thousand of loop detectors in PEMS systems. In this scenario, there is a need to develop STGNNs that can efficiently process and analyze the vast amounts of spatio-temporal data generated by the network while maintaining high prediction accuracy. Under this circumstance, more efficient AI solutions are appreciated, e.g., through model pruning/distillation, graph sampling techniques, or exploring the next-generation AI models with high efficiency. There are also a few studies probing into graph-free approaches [229] to reduce computational costs when scaling up to large-scale sensor networks. In addition, some advancements in time series prediction research have also challenged the necessity of employing overly complex temporal learning models [230], [231]. Therefore, reducing the complexity of both spatial and timing computations to improve the scalability of STGNNs is a promising direction.

IX. Conclusion

In this paper, we present a systematic survey of spatio-temporal graph neural networks (STGNNs) for predictive learning in urban computing. We start with a basic form and construction method of spatio-temporal graph data, and then summarize the predictive learning tasks involving STGNNs from different application domains in urban computing. Next, Moving on, we delve into the fundamental neural network architectures that underpin STGNNs, including the spatial learning network and temporal learning network, which consist of graph neural networks (GNNs), recurrent neural networks (RNNs), temporal convolutional networks (TCNs), self-attention networks (SANs), and explore the basic fusion techniques used to integrate these spatio-temporal neural architectures. To stay up-to-date with the latest developments in STGNNs, we review notable recent works, focusing on spatial learning methods, temporal learning methods, spatio-temporal fusion methods, and other advanced techniques that can be combined. Finally, we summarize the challenges of current research and suggest some potential directions.

References

[1] Y. Zheng, L. Capra, O. Wolfson, and H. Yang, "Urban computing: Concepts, methodologies, and applications," *Trans. Intell. Syst. Technol.*, vol. 5, no. 3, pp. 1–55, 2014.
[2] X. Wang, L. Li, Y. Yuan, P. Ye, and F.-Y. Wang, "ACP-based social computing and parallel intelligence: Societies 5.0 and beyond," *CAAI Trans. Intell. Technol.*, vol. 1, no. 4, pp. 377–393, 2016.
[3] S. Wang, J. Cao, and P. S. Yu, "Deep learning for spatio-temporal data mining: A survey," *IEEE Trans. Knowl. Data Eng.*, vol. 34, no. 8, pp. 3681–3700, Aug. 2022.
[4] H. Drucker, C. J. Burges, L. Kaufman, A. Smola, and V. Vapnik, "Support vector regression machines," in *Proc. Adv. Neural Inf. Process. Syst.*, 1996, pp. 155–161.
[5] L. Breiman, "Random forests," *Mach. Learn.*, vol. 45, pp. 5–32, 2001.
[6] A. Natekin and A. Knoll, "Gradient boosting machines, a tutorial," *Front. Neurorobot.*, vol. 7, 2013, Art. no. 21.
[7] J. Gu et al., "Recent advances in convolutional neural networks," *Pattern Recognit.*, vol. 77, pp. 354–377, 2018.
[8] Y. Yu, X. Si, C. Hu, and J. Zhang, "A review of recurrent neural networks: LSTM cells and network architectures," *Neural Comput.*, vol. 31, no. 7, pp. 1235–1270, 2019.
[9] X. Shi, Z. Chen, H. Wang, D.-Y. Yeung, W.-K. Wong, and W.-C. Woo, "Convolutional LSTM network: A machine learning approach for precipitation nowcasting," in *Proc. Adv. Neural Inf. Process. Syst.*, 2015, pp. 802–810.
[10] Y. Wang, M. Long, J. Wang, Z. Gao, and P. S. Yu, "PredRNN: Recurrent neural networks for predictive learning using spatiotemporal lstms," in *Proc. Adv. Neural Inf. Process. Syst.*, 2017, pp. 879–888.
[11] T. N. Kipf and M. Welling, "Semi-supervised classification with graph convolutional networks," 2016*arXiv:1609.02907*.
[12] J. Ye, J. Zhao, K. Ye, and C. Xu, "How to build a graph-based deep learning architecture in traffic domain: A survey," *IEEE Trans. Intell. Transp. Syst.*, vol. 23, no. 5, pp. 3904–3924, May 2022.
[13] K.-H. N. Bui, J. Cho, and H. Yi, "Spatial-temporal graph neural network for traffic forecasting: An overview and open research issues," *Appl. Intell.*, vol. 52, no. 3, pp. 2763–2774, 2022.
[14] W. Jiang and J. Luo, "Graph neural network for traffic forecasting: A survey," *Expert Syst. Appl.*, vol. 207, 2022, Art. no. 117921.
[15] Z. A. Sahili and M. Awad, "Spatio-temporal graph neural networks: A survey," 2023, *arXiv:2301.10569*.
[16] Y. Li, D. Yu, Z. Liu, M. Zhang, X. Gong, and L. Zhao, "Graph neural network for spatiotemporal data: Methods and applications," 2023, *arXiv:2306.00012*.
[17] S. Guo, Y. Lin, H. Wan, X. Li, and G. Cong, "Learning dynamics and heterogeneity of spatial-temporal graph data for traffic forecasting," *IEEE Trans. Knowl. Data Eng.*, vol. 34, no. 11, pp. 5415–5428, Nov. 2022.
[18] X. Wang et al., "Traffic flow prediction via spatial temporal graph neural network," in *Proc.Web Conf.*, 2020, pp. 1082–1092.
[19] D. Chai, L. Wang, and Q. Yang, "Bike flow prediction with multi-graph convolutional networks," in *Proc. 26th ACM SIGSPATIAL Int. Conf. Adv. Geographic Inf. Syst.*, 2018, pp. 397–400.
[20] B. Yu, H. Yin, and Z. Zhu, "Spatio-temporal graph convolutional networks: A deep learning framework for traffic forecasting," in *Proc. 27th Int. Joint Conf. Artif. Intell.* pp. 3634–3640, 2018.
[21] G. Jin, Z. Xi, H. Sha, Y. Feng, and J. Huang, "Deep multi-view graph-based network for citywide ride-hailing demand prediction," *Neurocomputing*, vol. 510, pp. 79–94, 2022.
[22] L. Liu, J. Chen, H. Wu, J. Zhen, G. Li, and L. Lin, "Physical-virtual collaboration modeling for intra-and inter-station metro ridership prediction," *IEEE Trans. Intell. Transp. Syst.*, vol. 23, no. 4, pp. 3377–3391, Apr. 2022.
[23] F. Li, H. Yan, G. Jin, Y. Liu, Y. Li, and D. Jin, "Automated spatio-temporal synchronous modeling with multiple graphs for traffic prediction," in *Proc. 24th ACM Int. Conf. Inf. Knowl. Manage.*, 2022, pp. 1084–1093.
[24] H. Shi et al., "Predicting origin-destination flow via multi-perspective graph convolutional network," in *Proc. IEEE 29th Int. Conf. Data Eng. Workshops*, 2020, pp. 1818–1821.
[25] X. Geng et al., "Spatiotemporal multi-graph convolution network for ride-hailing demand forecasting," in *Proc. AAAI Conf. Artif. Intell.*, 2019, vol. 33, no. 01 pp. 3656–3663.
[26] L. Zhao et al., "T-GCN: A temporal graph convolutional network for traffic prediction," *IEEE Trans. Intell. Transp. Syst.*, vol. 21, no. 9, pp. 3848–3858, Sep. 2020.
[27] J. Ji, J. Wang, Z. Jiang, J. Jiang, and H. Zhang, "STDEN: Towards physics-guided neural networks for traffic flow prediction," in *Proc. AAAI Conf. Artif. Intell.*, 2022, vol. 36, no. 4, pp. 4048–4056.
[28] Y. Li, R. Yu, C. Shahabi, and Y. Liu, "Diffusion convolutional recurrent neural network: Data-driven traffic forecasting," in *Proc. Int. Conf. Learning Representations*, 2018.
[29] Z. Zhang, M. Li, X. Lin, Y. Wang, and F. He, "Multistep speed prediction on traffic networks: A deep learning approach considering spatio-temporal dependencies," *Transp. Res. Part C: Emerg. Technol.*, vol. 105, pp. 297–322, 2019.

[30] Y. Wang et al., "Gallat: A spatiotemporal graph attention network for passenger demand prediction," in *Proc. IEEE 29th Int. Conf. Data Eng. Workshops*, 2021, pp. 2129–2134.

[31] X. Zhang et al., "Traffic flow forecasting with spatial-temporal graph diffusion network," in *Proc. AAAI Conf. Artif. Intell.*, 2021, vol. 35, no. 17, pp. 15008–15015.

[32] J. Sun, J. Zhang, Q. Li, X. Yi, Y. Liang, and Y. Zheng, "Predicting citywide crowd flows in irregular regions using multi-view graph convolutional networks," *IEEE Trans. Knowl. Data Eng.*, vol. 34, no. 5, pp. 2348–2359, May 2022.

[33] X. Zhang, C. Huang, Y. Xu, and L. Xia, "Spatial-temporal convolutional graph attention networks for citywide traffic flow forecasting," in *Proc. 29th ACM Int. Conf. Inf. Knowl. Manage.*, 2020, pp. 1853–1862.

[34] G. Jin, Y. Cui, L. Zeng, H. Tang, Y. Feng, and J. Huang, "Urban ride-hailing demand prediction with multiple spatio-temporal information fusion network," *Transp. Res. Part C: Emerg. Technol.*, vol. 117, 2020, Art. no. 102665.

[35] J. Ye, L. Sun, B. Du, Y. Fu, and H. Xiong, "Coupled layer-wise graph convolution for transportation demand prediction," in *Proc. AAAI Conf. Artif. Intell.*, 2021, vol. 35, no. 5, pp. 4617–4625.

[36] Y. Wang, H. Yin, H. Chen, T. Wo, J. Xu, and K. Zheng, "Origin-destination matrix prediction via graph convolution: A new perspective of passenger demand modeling," in *Proc. 25th ACM SIGKDD Int. Conf. Knowl. Discov. Data Mining*, 2019, pp. 1227–1235.

[37] B. Huang, K. Ruan, W. Yu, J. Xiao, R. Xie, and J. Huang, "ODformer: Spatial-temporal transformers for long sequence origin-destination matrix forecasting against cross application scenario," *Expert Syst. Appl.*, vol. 222, 2023, Art. no. 119835.

[38] J. Hu, B. Yang, C. Guo, C. S. Jensen, and H. Xiong, "Stochastic origin-destination matrix forecasting using dual-stage graph convolutional, recurrent neural networks," in *Proc. IEEE 36th Int. Conf. Data Eng. Workshops*, 2020, pp. 1417–1428.

[39] Z. Dapeng and F. Xiao, "Dynamic auto-structuring graph neural network: A joint learning framework for origin-destination demand prediction," *IEEE Trans. Knowl. Data Eng.*, vol. 35, no. 4, pp. 3699–3711, Apr. 2023.

[40] L. Liu, Y. Zhu, G. Li, Z. Wu, L. Bai, and L. Lin, "Online metro origin-destination prediction via heterogeneous information aggregation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 45, no. 3, pp. 3574–3589, Mar. 2023.

[41] Y. Wang et al., "Passenger mobility prediction via representation learning for dynamic directed and weighted graphs," *ACM Trans. Intell. Syst. Technol.*, vol. 13, no. 1, pp. 1–25, 2021.

[42] B. Wang, Y. Lin, S. Guo, and H. Wan, "GSNet: Learning spatial-temporal correlations from geographical and semantic aspects for traffic accident risk forecasting," in *Proc. AAAI Conf. Artif. Intell.*, 2021, vol. 35, no. 5, pp. 4402–4409.

[43] L. Yu, B. Du, X. Hu, L. Sun, L. Han, and W. Lv, "Deep spatio-temporal graph convolutional network for traffic accident prediction," *Neurocomputing*, vol. 423, pp. 135–147, 2021.

[44] Z. Wang, R. Jiang, H. Xue, F. D. Salim, X. Song, and R. Shibasaki, "Event-aware multimodal mobility nowcasting," in *Proc. AAAI Conf. Artif. Intell.*, 2022, vol. 36, no. 4, pp. 4228–4236.

[45] Z. Zhou, Y. Wang, X. Xie, L. Chen, and H. Liu, "RiskOracle: A minute-level citywide traffic accident forecasting framework," in *Proc. AAAI Conf. Artif. Intell.*, 2020, vol. 34, no. 01, pp. 1258–1265.

[46] G. Jin, L. Liu, F. Li, and J. Huang, "Spatio-temporal graph neural point process for traffic congestion event prediction," in *Proc. AAAI Conf. Artif. Intell.*, 2023, vol. 37, no. 12, pp. 14268–14276.

[47] X. Fang, J. Huang, F. Wang, L. Zeng, H. Liang, and H. Wang, "ConSTGAT: Contextual spatial-temporal graph attention network for travel time estimation at Baidu maps," in *Proc. 26th ACM SIGKDD Int. Conf. Knowl. Discov. Data Mining*, 2020, pp. 2697–2705.

[48] J. Huang et al., "DuETA: Traffic congestion propagation pattern modeling via efficient graph learning for eta prediction at Baidu maps," in *Proc. 31st ACM Int. Conf. Inf. Knowl. Manage.*, 2022, pp. 3172–3181.

[49] A. Derrow-Pinion et al., "ETA prediction with graph neural networks in google maps," in *Proc. 30th ACM Int. Conf. Inf. Knowl. Manage.*, 2021, pp. 3767–3776.

[50] K. Fu, F. Meng, J. Ye, and Z. Wang, "CompactETA: A fast inference system for travel time prediction," in *Proc. 26th ACM SIGKDD Int. Conf. Knowl. Discov. Data Mining*, 2020, pp. 3337–3345.

[51] Z. Wu, D. Zheng, S. Pan, Q. Gan, G. Long, and G. Karypis, "TraverseNet: Unifying space and time in message passing for traffic forecasting," *IEEE Trans. Neural Netw. Learn. Syst.*, early access, Jul. 14, 2022, doi: 10.1109/TNNLS.2022.3186103.

[52] M. Li and Z. Zhu, "Spatial-temporal fusion graph neural networks for traffic flow forecasting," in *Proc. AAAI Conf. Artif. Intell.*, 2021, vol. 35, no. 5, pp. 4189–4196.

[53] C. Song, Y. Lin, S. Guo, and H. Wan, "Spatial-temporal synchronous graph convolutional networks: A new framework for spatial-temporal network data forecasting," in *Proc. AAAI Conf. Artif. Intell.*, 2020, vol. 34, no. 01, pp. 914–921.

[54] G. Jin, F. Li, J. Zhang, M. Wang, and J. Huang, "Automated dilated spatio-temporal synchronous graph modeling for traffic prediction," *IEEE Trans. Intell. Transp. Syst.*, vol. 24, no. 8, pp. 8820–8830, Aug. 2023.

[55] C. Zheng, X. Fan, C. Wang, and J. Qi, "GMAN: A graph multi-attention network for traffic prediction," in *Proc. AAAI Conf. Artif. Intell.*, 2020, vol. 34, no. 01, pp. 1234–1241.

[56] W. Chen, L. Chen, Y. Xie, W. Cao, Y. Gao, and X. Feng, "Multi-range attentive bicomponent graph convolutional network for traffic forecasting," in *Proc. AAAI Conf. Artif. Intell.*, 2020, vol. 34, no. 04, pp. 3529–3536.

[57] Z. Wu, S. Pan, G. Long, J. Jiang, and C. Zhang, "Graph wavenet for deep spatial-temporal graph modeling," in *Proc. 28th Int. Joint Conf. Artif. Intell*. 2019, pp. 1907–1913.

[58] F. Li et al., "Dynamic graph convolutional recurrent network for traffic prediction: Benchmark and solution," *ACM Trans. Knowl. Discov. Data*, vol. 17, no. 1, pp. 1–21, 2021.

[59] L. Han, B. Du, L. Sun, Y. Fu, Y. Lv, and H. Xiong, "Dynamic and multi-faceted spatio-temporal deep learning for traffic speed forecasting," in *Proc. 27th ACM SIGKDD Int. Conf. Knowl. Discov Data Mining*, 2023, pp. 547–555.

[60] D. Liu, J. Wang, S. Shang, and P. Han, "MSDR: Multi-step dependency relation networks for spatial temporal forecasting," in *Proc. 28th ACM SIGKDD Int. Conf. Knowl. Discov. Data Mining*, 2022, pp. 1042–1050.

[61] G. Li et al., "A lightweight and accurate spatial-temporal transformer for traffic forecasting," *IEEE Trans. Knowl. Data Eng.*, vol. 35, no. 11, pp. 10967–10980, Nov. 2023.

[62] Q. Xie, T. Guo, Y. Chen, Y. Xiao, X. Wang, and B. Y. Zhao, "Deep graph convolutional networks for incident-driven traffic speed prediction," in *Proc. 29th ACM Int. Conf. Inf. Knowl. Manage.*, 2020, pp. 1665–1674.

[63] L. Bai, L. Yao, S. S. Kanhere, X. Wang, W. Liu, and Z. Yang, "Spatio-temporal graph convolutional and recurrent networks for citywide passenger demand prediction," in *Proc. 28th ACM Int. Conf. Inf. Knowl. Manage.*, 2019, pp. 2293–2296.

[64] A. Ali, Y. Zhu, and M. Zakarya, "Exploiting dynamic spatio-temporal graph convolutional neural networks for citywide traffic flows prediction," *Neural Netw.*, vol. 145, pp. 233–247, 2022.

[65] F. Zhou, L. Li, K. Zhang, and G. Trajcevski, "Urban flow prediction with spatial–temporal neural odes," *Transp. Res. Part C: Emerg. Technol.*, vol. 124, 2021, Art. no. 102912.

[66] H. Peng et al., "Dynamic graph convolutional network for long-term traffic flow prediction with reinforcement learning," *Inf. Sci.*, vol. 578, pp. 401–416, 2021.

[67] S. Wang, H. Miao, J. Li, and J. Cao, "Spatio-temporal knowledge transfer for urban crowd flow prediction via deep attentive adaptation networks," *IEEE Trans. Intell. Transp. Syst.*, vol. 23, no. 5, pp. 4695–4705, May 2022.

[68] S. Wang, M. Zhang, H. Miao, Z. Peng, and P. S. Yu, "Multivariate correlation-aware spatio-temporal graph convolutional networks for multi-scale traffic prediction," *ACM Trans. Intell. Syst. Technol.*, vol. 13, no. 3, pp. 1–22, 2022.

[69] Z. Zhou, Y. Wang, X. Xie, L. Chen, and C. Zhu, "Foresee urban sparse traffic accidents: A spatiotemporal multi-granularity perspective," *IEEE Trans. Knowl. Data Eng.*, vol. 34, no. 8, pp. 3786–3799, Aug. 2022.

[70] Z. Wang et al., "Spatio-temporal-categorical graph neural networks for fine-grained multi-incident co-prediction," in *Proc. 30th ACM Int. Conf. Inf. Knowl. Manage.*, 2021, pp. 2060–2069.

[71] X. Wu, C. Huang, C. Zhang, and N. V. Chawla, "Hierarchically structured transformer networks for fine-grained spatial event forecasting," in *Proc. Int. Conf. World Wide Web Conf.*, 2020, pp. 2320–2330.

[72] Q. Wang, C. Xu, W. Zhang, and J. Li, "GraphTTE: Travel time estimation based on attention-spatiotemporal graphs," *IEEE Signal Process. Lett.*, vol. 28, pp. 239–243, 2021.

[73] H. Wang et al., "Multi-task weakly supervised learning for origin destination travel time estimation," *Proc. IEEE Trans, Knowl. Data Eng.*, vol. 35, no. 11, pp. 11628–11641, Nov. 2023.

[74] G. Jin, H. Yan, F. Li, J. Huang, and Y. Li, "Spatio-temporal dual graph neural networks for travel time estimation," 2021, *arXiv:2105.13591*.

[75] G. Jin, H. Yan, F. Li, Y. Li, and J. Huang, "Hierarchical neural architecture search for travel time estimation," in *Proc. 29th ACM SIGSPATIAL Int. Conf. Adv. Geographic Inf. Syst.*, 2021, pp. 91–94.

[76] G. Jin, M. Wang, J. Zhang, H. Sha, and J. Huang, "STGNN-TTE: Travel time estimation via spatial–temporal graph neural network," *Future Gener. Comput. Syst.*, vol. 126, pp. 70–81, 2022.

[77] J. Jiang, D. Pan, H. Ren, X. Jiang, C. Li, and J. Wang, "Self-supervised trajectory representation learning with temporal regularities and travel semantics," in *Proc. IEEE 29th Int. Conf. Data Eng. Workshops*, 2023, pp. 843–855.

[78] Y. Peng, G. Zhang, X. Li, and L. Zheng, "STIRNet: A. spatial-temporal interaction-aware recursive network for human trajectory prediction," in *Proc. IEEE/CVF Int. Conf. Comput. Vis.*, 2021, pp. 2285–2293.

[79] L. Shi et al., "SGCN: Sparse graph convolution network for pedestrian trajectory prediction," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2021, pp. 8994–9003.

[80] A. Mohamed, K. Qian, M. Elhoseiny, and C. Claudel, "Social-STGCNN: A social spatio-temporal graph convolutional neural network for human trajectory prediction," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2020, pp. 14424–14432.

[81] S. Malla, C. Choi, and B. Dariush, "Social-STAGE: Spatio-temporal multi-modal future trajectory forecast," in *Proc. IEEE Int. Conf. Robot. Automat.*, 2021, pp. 13938–13944.

[82] Y. Liu, L. Yao, B. Li, X. Wang, and C. Sammut, "Social graph transformer networks for pedestrian trajectory prediction in complex social scenarios," in *Proc. 31st ACM Int. Conf. Inf. Knowl. Manage.*, 2022, pp. 1339–1349.

[83] J. Lian, W. Ren, L. Li, Y. Zhou, and B. Zhou, "PTP-STGCN: Pedestrian trajectory prediction based on a spatio-temporal graph convolutional neural network," *Appl. Intell.*, vol. 53, no. 3, pp. 2862–2878, 2022.

[84] Y. Lin et al., "Exploiting spatiotemporal patterns for accurate air quality forecasting using deep learning," in *Proc. 26th ACM SIGSPATIAL Int. Conf. Adv. Geographic Inf. Syst.*, 2018, pp. 359–368.

[85] J. Han, H. Liu, H. Zhu, H. Xiong, and D. Dou, "Joint air quality and weather prediction based on multi-adversarial spatiotemporal networks," in *Proc. AAAI Conf. Artif. Intell.*, 2021, vol. 35, no. 5, pp. 4081–4089.

[86] H. Lin, Z. Gao, Y. Xu, L. Wu, L. Li, and S. Z. Li, "Conditional local convolution for spatio-temporal meteorological forecasting," in *Proc. AAAI Conf. Artif. Intell.*, 2022, vol. 36, no. 7, pp. 7470–7478.

[87] T. Stańczyk and S. Mehrkanoon, "Deep graph convolutional networks for wind speed prediction," 2021, *arXiv:2101.10041*.

[88] N. Rathore, P. Rathore, A. Basak, S. H. Nistala, and V. Runkana, "Multi scale graph wavenet for wind speed forecasting," in *Proc. IEEE Int. Conf. Big Data*, 2021, pp. 4047–4053.

[89] L. Xia et al., "Spatial-temporal sequential hypergraph network for crime prediction with dynamic multiplex relation learning," in *Proc. Int. Joint Conf. Artif. Intell.*, 2021, pp. 1631–1637.

[90] Z. Li, C. Huang, L. Xia, Y. Xu, and J. Pei, "Spatial-temporal hypergraph self-supervised learning for crime prediction," in *Proc. IEEE 38th Int. Conf. Data Eng. Workshops*, 2022, pp. 2984–2996.

[91] M. Sun, P. Zhou, H. Tian, Y. Liao, and H. Xie, "Spatial-temporal attention network for crime prediction with adaptive graph learning," in *Proc. Int. Conf. Artif. Neural Netw.*, 2022, pp. 656–669.

[92] S. F. Tekin and S. S. Kozat, "Crime prediction with graph neural networks and multivariate normal distributions," *Signal, Image Video Process.*, vol. 17, no. 4, pp. 1053–1059, 2022.

[93] Y. Zhang and T. Cheng, "Graph deep learning model for network-based predictive hotspot mapping of sparse spatio-temporal events," *Computers, Environ. Urban Syst.*, vol. 79, 2020, Art. no. 101403.

[94] G. Jin, C. Liu, Z. Xi, H. Sha, Y. Liu, and J. Huang, "Adaptive dual-view wavenet for urban spatial–temporal event prediction," *Inf. Sci.*, vol. 588, pp. 315–330, 2022.

[95] X. Yang, F. Zhang, P. Sun, X. Li, Z. Du, and R. Liu, "A spatio-temporal graph-guided convolutional LSTM for tropical cyclones precipitation nowcasting," *Appl. Soft Comput.*, vol. 124, 2022, Art. no. 109003.

[96] J. Zhou, J. Xiang, and S. Huang, "Classification and prediction of typhoon levels by satellite cloud pictures through GC–LSTM deep learning model," *Sensors*, vol. 20, no. 18, 2020, Art. no. 5132.

[97] X. Zhang, W. Reichard-Flynn, M. Zhang, M. Hirn, and Y. Lin, "Spatiotemporal graph convolutional networks for earthquake source characterization," *J. Geophysical Res.: Solid Earth*, vol. 127, no. 11, 2022, Art. no. e2022JB024401.

[98] B. Feng and G. Fox, "GTrans: Spatiotemporal autoregressive transformer with graph embeddings for nowcasting extreme events," 2022, *arXiv:2201.06717*.

[99] L. Wang, A. Adiga, J. Chen, A. Sadilek, S. Venkatramanan, and M. Marathe, "CausalGNN: Causal-based graph neural networks for spatio-temporal epidemic forecasting," in *Proc. AAAI Conf. Artif. Intell.*, 2022, vol. 36, no. 11, pp. 12191–12199.

[100] F. Xie, Z. Zhang, L. Li, B. Zhou, and Y. Tan, "EpiGNN: Exploring spatial transmission with graph neural network for regional epidemic forecasting," in *Proc. Joint Eur. Conf. Mach. Learn. Knowl. Discov. Databases*, 2022, pp. 469–485.

[101] Z. Li, X. Luo, B. Wang, A. L. Bertozzi, and J. Xin, "A study on graph-structured recurrent neural networks and sparsification with application to epidemic forecasting," in *Proc. Optim. Complex Syst.: Theory, Models, Algorithms Appl.*, 2020, pp. 730–739.

[102] S. Yu, F. Xia, S. Li, M. Hou, and Q. Z. Sheng, "Spatio-temporal graph learning for epidemic prediction," *ACM Trans. Intell. Syst. Technol.*, vol. 14, no. 2, pp. 1–25, 2023.

[103] A. Kapoor et al., "Examining COVID-19 forecasting using spatio-temporal graph neural networks," 2020, *arXiv:2007.03113*.

[104] V. La Gatta, V. Moscato, M. Postiglione, and G. Sperli, "An epidemiological neural network exploiting dynamic graph structured data applied to the COVID-19 outbreak," *Trans. Big Data*, vol. 7, no. 1, pp. 45–55, 2020.

[105] S. Deng, S. Wang, H. Rangwala, L. Wang, and Y. Ning, "Graph message passing with cross-location attentions for long-term ILI prediction," 2019, *arXiv:1912.10202*.

[106] Y. Lu, W. Wang, X. Hu, P. Xu, S. Zhou, and M. Cai, "Vehicle trajectory prediction in connected environments via heterogeneous context-aware graph convolutional networks," *IEEE Trans. Intell. Transp. Syst.*, vol. 24, no. 8, pp. 8452–8464, Aug. 2023.

[107] X. Mo, Z. Huang, Y. Xing, and C. Lv, "Multi-agent trajectory prediction with heterogeneous edge-enhanced graph attention network," *IEEE Trans. Intell. Transp. Syst.*, vol. 23, no. 7, pp. 9554–9567, Jul. 2022.

[108] X. Mo, Y. Xing, H. Liu, and C. Lv, "Map-adaptive multimodal trajectory prediction using hierarchical graph neural networks," *IEEE Robot. Automat. Lett.*, vol. 8, no. 6, pp. 3685–3692, Jun. 2023.

[109] Y. Liang et al., "Airformer: Predicting nationwide air quality in China with transformers," in *Proc. AAAI Conf. Artif. Intell.*, vol. 37, no. 12, 2023, pp. 14329–14337.

[110] H. Zhou, F. Zhang, Z. Du, and R. Liu, "Forecasting PM2. 5 using hybrid graph convolution-based model considering dynamic wind-field to offer the benefit of spatial interpretability," *Environ. Pollut.*, vol. 273, 2021, Art. no. 116473.

[111] S. Wang, Y. Li, J. Zhang, Q. Meng, L. Meng, and F. Gao, "PM2. 5-GNN: A domain knowledge enhanced graph neural network for pm2. 5 forecasting," in *Proc. 28th ACM SIGSPATIAL Int. Conf. Adv. Geographic Inf. Syst.*, 2020, pp. 163–166.

[112] Y. Liang, S. Ke, J. Zhang, X. Yi, and Y. Zheng, "Geoman: Multi-level attention networks for geo-sensory time series prediction," in *Proc. Int. Joint Conf. Artif. Intell.*, 2018, vol. 2018, pp. 3428–3434.

[113] S. Chen, J. A. Zwart, and X. Jia, "Physics-guided graph meta learning for predicting water temperature and streamflow in stream networks," in *Proc. 28th ACM SIGKDD Int. Conf. Knowl. Discov. Data Mining*, 2022, pp. 2752–2761.

[114] X. Jia et al., "Physics-guided recurrent graph model for predicting flow and temperature in river networks," in *Proc. SIAM Int. Conf. Data Mining*, 2021, pp. 612–620.

[115] H. Lira, L. Martí, and N. Sanchez-Pi, "Frost forecasting model using graph neural networks with spatio-temporal attention," in *Proc. AI: Model. Oceans Climate Change Workshop ICLR*, 2021.

[116] M. Khodayar and J. Wang, "Spatio-temporal graph deep neural network for short-term wind speed forecasting," *IEEE Trans. Sustain. Eng.*, vol. 10, no. 2, pp. 670–681, Apr. 2019.

[117] Y. Qian, L. Pan, P. Wu, and Z. Xia, "GeST: A grid embedding based spatio-temporal correlation model for crime prediction," in *Proc. IEEE 5th Int. Conf. Data Sci. Cyberspace*, 2020, pp. 1–7.

[118] X. Song, H. Wang, and B. Zhou, "DGCN-RS: A dilated graph convolutional networks jointly modelling relation and semantic for multi-event forecasting," in *Proc. 28th Int. Conf. Neural Inf. Process.*, 2021, pp. 666–676.

[119] G. Jin, H. Sha, Y. Feng, Q. Cheng, and J. Huang, "GSEN: An ensemble deep learning benchmark model for urban hotspots spatiotemporal prediction," *Neurocomputing*, vol. 455, pp. 353–367, 2021.

[120] B. Wang, X. Luo, F. Zhang, B. Yuan, A. L. Bertozzi, and P. J. Brantingham, "Graph-based deep modeling and real time forecasting of sparse spatio-temporal data," 2018, *arXiv:1804.00684*.

[121] C. Wang et al., "HAGEN: Homophily-aware graph convolutional recurrent network for crime forecasting," in *Proc. AAAI Conf. Artif. Intell.*, 2022, vol. 36, no. 4, pp. 4193–4200.

[122] J. Feng, H. Sha, Y. Ding, L. Yan, and Z. Yu, "Graph convolution based spatial-temporal attention LSTM model for flood forecasting," in *Proc. IEEE Int. Joint Conf. Neural Netw.*, 2022, pp. 1–8.

[123] F. Yuan, Y. Xu, Q. Li, and A. Mostafavi, "Spatio-temporal graph convolutional networks for road network inundation status prediction during urban flooding," *Comput., Environ. Urban Syst.*, vol. 97, 2022, Art. no. 101870.

[124] G. Jin, C. Zhu, X. Chen, H. Sha, X. Hu, and J. Huang, "UFSP-Net: A neural network with spatio-temporal information fusion for urban fire situation prediction," *IOP Conf. Ser.: Mater. Sci. Eng.*, vol. 853, no. 1, 2020, Art. no. 012050.

[125] H. Farahmand, Y. Xu, and A. Mostafavi, "A spatial-temporal graph deep learning model for urban flood nowcasting leveraging heterogeneous community features," *Scientific Reports*, vol. 13, no. 1, 2023, Art. no. 6768.

[126] A. Doğan and E. Demir, "Structural recurrent neural network models for earthquake prediction," *Neural Comput. Appl.*, vol. 34, no. 13, pp. 11049–11062, 2022.

[127] I. W. McBrearty and G. C. Beroza, "Earthquake location and magnitude estimation with graph neural networks," in *Proc. IEEE Int. Conf. Image Process.*, 2022, pp. 3858–3862.

[128] K. M. Ngoc and M. Lee, "Forecasting COVID-19 confirmed cases in South Korea using spatio-temporal graph neural networks," *Int. J. Contents*, vol. 17, no. 3, pp. 1–14, 2021.

[129] G. Panagopoulos, G. Nikolentzos, and M. Vazirgiannis, "Transfer graph neural networks for pandemic forecasting," in *Proc. AAAI Conf. Artif. Intell.*, 2021, vol. 35, no. 6, pp. 4838–4845.

[130] J. Gao et al., "STAN: Spatio-temporal attention network for pandemic prediction using real-world evidence," *J. Amer. Med. Informat. Assoc.*, vol. 28, no. 4, pp. 733–743, 2021.

[131] S. Deng, S. Wang, H. Rangwala, L. Wang, and Y. Ning, "Cola-GNN: Cross-location attention based graph neural networks for long-term ILI prediction," in *Proc. 24th ACM Int. Conf. Inf. Knowl. Manage.*, 2020, pp. 245–254.

[132] C. Sun, V. K. Kumarasamy, Y. Liang, D. Wu, and Y. Wang, "Using a layered ensemble of physics-guided graph attention networks to predict COVID-19 trends," *Appl. Artif. Intell.*, vol. 36, no. 1, 2022, Art. no. 2055989.

[133] Y. Zheng, Z. Li, J. Xin, and G. Zhou, "A spatial-temporal graph based hybrid infectious disease model with application to COVID-19," in *Proc. 10th Int. Conf. Pattern Recognit. Appl. Methods*, 2021, pp. 357–364.

[134] Z. Wang et al., "Forecasting ambulance demand with profiled human mobility via heterogeneous multi-graph neural networks," in *Proc. IEEE 37th Int. Conf. Data Eng. Workshops*, 2021, pp. 1751–1762.

[135] R. Jin, T. Xia, X. Liu, T. Murata, and K.-S. Kim, "Predicting emergency medical service demand with bipartite graph convolutional networks," *IEEE Access*, vol. 9, pp. 9903–9915, 2021.

[136] T. Munasinghe and B. Behlendorf, "Using graph neural networks to investigate the relationship between the socioeconomic factors and emergency medical service (EMS) median response time in New York City," in *Proc. IEEE Int. Conf. Big Data*, 2022, pp. 6781–6783.

[137] R. Yu, Y. Sun, D. He, J. Gao, Z. Liu, and M. Yu, "Spatio-temporal graph cross-correlation auto-encoding network for wind power prediction," *Int. J. Mach. Learn. Cybern.*, pp. 1–13, 2022.

[138] Z. Li et al., "A spatiotemporal directed graph convolution network for ultra-short-term wind power prediction," *IEEE Trans. Softw. Eng.*, vol. 14, no. 1, pp. 39–54, Jan. 2023.

[139] J. Simeunović, B. Schubnel, P. -J. Alet, and R. E. Carrillo, "Spatio-temporal graph neural networks for multi-site PV power forecasting," *IEEE Trans. Softw. Eng.*, vol. 13, no. 2, pp. 1210–1220, Apr. 2022.

[140] B. Hui, D. Yan, W.-S. Ku, and W. Wang, "Predicting economic growth by region embedding: A multigraph convolutional network approach," in *Proc. 29th ACM Int. Conf. Inf. Knowl. Manage.*, 2020, pp. 555–564.

[141] F. Xu, Y. Li, and S. Xu, "Attentional multi-graph convolutional network for regional economy prediction with open migration data," in *Proc. 26th ACM SIGKDD Int. Conf. Knowl. Discov. Data Mining*, 2020, pp. 2225–2233.

[142] M. Defferrard, X. Bresson, and P. Vandergheynst, "Convolutional neural networks on graphs with fast localized spectral filtering," in *Proc. Adv. Neural Inf. Process. Syst.*, 2016, pp. 3844–3852.

[143] W. Hamilton, Z. Ying, and J. Leskovec, "Inductive representation learning on large graphs," in *Proc. Adv. Neural Inf. Process. Syst.*, 2017, pp. 1025–1035.

[144] P. Veličković, G. Cucurull, A. Casanova, A. Romero, P. Lio, and Y. Bengio, "Graph attention networks," in *Proc. Int. Conf. Learn. Representations*, 2018.

[145] J. Zhou et al., "Graph neural networks: A review of methods and applications," *AI Open*, vol. 1, pp. 57–81, 2020.

[146] D. K. Hammond, P. Vandergheynst, and R. Gribonval, "Wavelets on graphs via spectral graph theory," *Appl. Comput. Harmon. Anal.*, vol. 30, no. 2, pp. 129–150, 2011.

[147] M. Lukoševičius and H. Jaeger, "Reservoir computing approaches to recurrent neural network training," *Comput. Sci. Rev.*, vol. 3, no. 3, pp. 127–149, 2009.

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

[149] R. Dey and F. M. Salem, "Gate-variants of gated recurrent unit (GRU) neural networks," in *Proc. IEEE 60th Int. Midwest Symp. Circuits Syst.*, 2017, pp. 1597–1600.

[150] Y. N. Dauphin, A. Fan, M. Auli, and D. Grangier, "Language modeling with gated convolutional networks," in *Proc. Int. Conf. Mach. Learn.*, 2017, pp. 933–941.

[151] A. v. d. Oord et al., "WaveNet: A generative model for raw audio," 2016, *arXiv:1609.03499*.

[152] F. Yu and V. Koltun, "Multi-scale context aggregation by dilated convolutions," 2015, *arXiv:1511.07122*.

[153] A. Vaswani et al., "Attention is all you need," in *Proc. Adv. Neural Inf. Process. Syst.*, 2017, pp. 6000–6010.

[154] K. Cho et al., "Learning phrase representations using RNN encoder-decoder for statistical machine translation," in *Proc. 2014 Conf. Empirical Methods Natural Lang. Process.*, 2014, Art. no. 1724.

[155] Q. Ni and M. Zhang, "STGMN: A gated multi-graph convolutional network framework for traffic flow prediction," *Appl. Intell.*, vol. 52, no. 13, pp. 15026–15039, 2022.

[156] Y. He, L. Li, X. Zhu, and K. L. Tsui, "Multi-graph convolutional-recurrent neural network (MGC-RNN) for short-term forecasting of transit passenger flow," *IEEE Trans. Intell. Transp. Syst.*, vol. 23, no. 10, pp. 18155–18174, Oct. 2022.

[157] Z. Xu, Y. Kang, Y. Cao, and Z. Li, "Spatiotemporal graph convolution multifusion network for urban vehicle emission prediction," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 32, no. 8, pp. 3342–3354, Aug. 2021.

[158] Z. Shao et al., "Decoupled dynamic spatial-temporal graph neural network for traffic forecasting," in *Proc. VLDB Endowment*, vol. 15, no. 11, 2022, pp. 2733–2746.

[159] Z. Wu, S. Pan, G. Long, J. Jiang, X. Chang, and C. Zhang, "Connecting the dots: Multivariate time series forecasting with graph neural networks," in *Proc. 26th ACM SIGKDD Int. Conf. Knowl. Discov. Data Mining*, 2020, pp. 753–763.

[160] B. Lu, X. Gan, H. Jin, L. Fu, and H. Zhang, "Spatiotemporal adaptive gated graph convolution network for urban traffic flow forecasting," in *Proc. 29th ACM Int. Conf. Inf. Knowl. Manage.*, 2020, pp. 1025–1034.

[161] W. Zhang et al., "ADAPGL: An adaptive graph learning algorithm for traffic prediction based on spatiotemporal neural networks," *Transp. Res. Part C: Emerg. Technol.*, vol. 139, 2022, Art. no. 103659.

[162] S. Lan, Y. Ma, W. Huang, W. Wang, H. Yang, and P. Li, "DstaGNN: Dynamic spatial-temporal aware graph neural network for traffic flow forecasting," in *Proc. Int. Conf. Mach. Learn.*, 2022, pp. 11906–11917.

[163] K. Guo, Y. Hu, Z. Qian, Y. Sun, J. Gao, and B. Yin, "Dynamic graph convolution network for traffic forecasting based on latent network of Laplace matrix estimation," *IEEE Trans. Intell. Transp. Syst.*, vol. 23, no. 2, pp. 1009–1018, Feb. 2022.

[164] C. Shang, J. Chen, and J. Bi, "Discrete graph structure learning for forecasting multiple time series," in *Proc. Int. Conf. Learn. Representations*, 2020.

[165] K. Guo, Y. Hu, Y. Sun, S. Qian, J. Gao, and B. Yin, "Hierarchical graph convolution network for traffic forecasting," in *Proc. AAAI Conf. Artif. Intell.*, 2021, vol. 35, no. 1, pp. 151–159.

[166] L. Chen et al., "Group-aware graph neural network for Nationwide City Air Quality Forecasting," 2021, *arXiv:2108.12238*.

[167] Y. Tang, J. He, and Z. Zhao, "HGARN: Hierarchical graph attention recurrent network for human mobility prediction," 2022, *arXiv:2210.07765*.
[168] S. Agarwal, R. Sawhney, M. Thakkar, P. Nakov, J. Han, and T. Derr, "THINK: Temporal hypergraph hyperbolic network," in *Proc. IEEE Int. Conf. Data Mining*, 2022, pp. 849–854.
[169] Y. Qin, Y. Fang, H. Luo, F. Zhao, and C. Wang, "DMGCRN: Dynamic multi-graph convolution recurrent network for traffic forecasting," 2021, *arXiv:2112.02264*.
[170] S. Liu, X. Chen, Z. Wu, L. Deng, H. Su, and K. Zheng, "HeGA: Heterogeneous graph aggregation network for trajectory prediction in high-density traffic," in *Proc. 31st ACM Int. Conf. Inf. Knowl. Manage.*, 2022, pp. 1319–1328.
[171] Q. Zhou et al., "Modeling heterogeneous relations across multiple modes for potential crowd flow prediction," in *Proc. AAAI Conf. Artif. Intell.*, 2021, vol. 35, no. 5, pp. 4723–4731.
[172] S. Guo, Y. Lin, N. Feng, C. Song, and H. Wan, "Attention based spatial-temporal graph convolutional networks for traffic flow forecasting," in *Proc. AAAI Conf. Artif. Intell.*, 2019, vol. 33, no. 01, pp. 922–929.
[173] H. Wang, R. Zhang, X. Cheng, and L. Yang, "Hierarchical traffic flow prediction based on spatial-temporal graph convolutional network," *IEEE Trans. Intell. Transp. Syst.*, vol. 23, no. 9, pp. 16137–16147, Sep. 2022.
[174] B. N. Oreshkin, A. Amini, L. Coyle, and M. Coates, "FC-GAGA: Fully connected gated graph architecture for spatio-temporal traffic forecasting," in *Proc. AAAI Conf. Artif. Intell.*, 2021, vol. 35, no. 10, pp. 9233–9241.
[175] Y. Fang et al., "Spatio-temporal meets wavelet: Disentangled traffic flow forecasting via efficient spectral graph attention network," 2021, *arXiv:2112.02740*.
[176] D. Cao et al., "Spectral temporal graph neural network for multivariate time-series forecasting," in *Proc. Adv. Neural Inf. Process. Syst.*, 2020, vol. 33, pp. 17766–17778.
[177] S. Zheng, Z. Gao, W. Cao, J. Bian, and T.-Y. Liu, "Hierst: A unified hierarchical spatial-temporal framework for COVID-19 trend forecasting," in *Proc. 30th ACM Int. Conf. Inf. Knowl. Manage.*, 2021, pp. 4383–4392.
[178] B. N. Oreshkin, D. Carpov, N. Chapados, and Y. Bengio, "N-BEATS: Neural basis expansion analysis for interpretable time series forecasting," in *Proc. Int. Conf. Learn. Representations*, 2019.
[179] T. Xia et al., "3DGCN: 3-dimensional dynamic graph convolutional network for citywide crowd flow prediction," *ACM Trans. Knowl. Discov. Data*, vol. 15, no. 6, pp. 1–21, 2021.
[180] C. Zheng et al., "Spatio-temporal joint graph convolutional networks for traffic forecasting," *IEEE Trans. Knowl. Data Eng.*, no. 1, 2023, pp. 1–14, doi: 10.1109/TKDE.2023.3284156.
[181] T. Wang et al., "Synchronous spatiotemporal graph transformer: A new framework for traffic data prediction," *IEEE Trans. Neural Netw. Learn. Syst.*, early access, 06 May, 2022, doi: 10.1109/TNNLS.2022.3169488.
[182] H. Li, D. Jin, X. Li, J. Huang, and J. Yoo, "Multi-task synchronous graph neural networks for traffic spatial-temporal prediction," in *Proc. 29th ACM SIGSPATIAL Int. Conf. Adv. Geographic Inf. Syst.*, 2021, pp. 137–140.
[183] Y. Fang et al., "CDGNet: A cross-time dynamic graph-based deep learning model for traffic forecasting," 2021, *arXiv:2112.02736*.
[184] Y. Fang, F. Zhao, Y. Qin, H. Luo, and C. Wang, "Learning all dynamics: Traffic forecasting via locality-aware spatio-temporal joint transformer," *IEEE Trans. Intell. Transp. Syst.*, vol. 23, no. 12, pp. 23433–23446, Dec. 2022.
[185] Z. Pan et al., "AutoSTG: Neural architecture search for predictions of spatio-temporal graph," in *Proc. World Wide Web Conf.*, 2021, pp. 1846–1855.
[186] H. Liu, K. Simonyan, and Y. Yang, "DARTS: Differentiable architecture search," in *Proc. Int. Conf. Learn. Representations*, 2018.
[187] G. Jin, H. Yan, F. Li, Y. Li, and J. Huang, "Dual graph convolution architecture search for travel time estimation," *Trans. Intell. Transp. Syst.*, vol. 14, no. 4, pp. 1–23, 2023.
[188] C. Wang, K. Zhang, H. Wang, and B. Chen, "Auto-STGCN: Autonomous spatial-temporal graph convolutional network search," *ACM Trans. Knowl. Discov. Data*, vol. 17, no. 5, pp. 1–21, 2022.
[189] X. Wu, D. Zhang, C. Guo, C. He, B. Yang, and C. S. Jensen, "AutoCTS: Automated correlated time series forecasting," *Proc. VLDB Endowment*, vol. 15, no. 4, pp. 971–983, 2021.
[190] G. Jin, H. Sha, Z. Xi, and J. Huang, "Urban hotspot forecasting via automated spatio-temporal information fusion," *Appl. Soft Comput.*, vol. 136, 2023, Art. no. 110087.
[191] S. Ke, Z. Pan, T. He, Y. Liang, J. Zhang, and Y. Zheng, "AutoSTG : An automatic framework to discover the optimal network for spatio-temporal graph prediction," *Artif. Intell.*, vol. 318, 2023, Art. no. 103899.
[192] A. Khaled, A. M. T. Elsir, and Y. Shen, "TFGAN: Traffic forecasting using generative adversarial network with multi-graph convolutional network," *Knowl.-Based Syst.*, vol. 249, 2022, Art. no. 108990.
[193] Z. Huang, W. Zhang, D. Wang, and Y. Yin, "A GAN framework-based dynamic multi-graph convolutional network for origin–destination-based ride-hailing demand prediction," *Inf. Sci.*, vol. 601, pp. 129–146, 2022.
[194] J. Jin et al., "A GAN-based short-term link traffic prediction approach for urban road networks under a parallel learning framework," *IEEE Trans. Intell. Transp. Syst.*, vol. 23, no. 9, pp. 16185–16196, Sep. 2022.
[195] C. Liu, Z. Xiao, D. Wang, M. Cheng, H. Chen, and J. Cai, "Foreseeing private car transfer between urban regions with multiple graph-based generative adversarial networks," *World Wide Web*, vol. 25, no. 6, pp. 2515–2534, 2022.
[196] H. Miao, J. Shen, J. Cao, J. Xia, and S. Wang, "MBA-STNet: Bayes-enhanced discriminative multi-task learning for flow prediction," *IEEE Trans. Knowl. Data Eng.*, vol. 35, no. 7, pp. 7164–7177, Jul. 2023.
[197] Z. Pan, Y. Liang, W. Wang, Y. Yu, Y. Zheng, and J. Zhang, "Urban traffic prediction from spatio-temporal data using deep meta learning," in *Proc. 25th ACM SIGKDD Int. Conf. Knowl. Discov. Data Mining*, 2019, pp. 1720–1730.
[198] Z. Pan et al., "Spatio-temporal meta learning for urban traffic prediction," *IEEE Trans. Knowl. Data Eng.*, vol. 34, no. 3, pp. 1462–1476, Mar. 2022.
[199] R. Jiang et al., "Spatio-temporal meta-graph learning for traffic forecasting," in *Proc. AAAI Conf. Artif. Intell.*, vol. 37, no. 7, 2023, pp. 8078–8086.
[200] B. Lu, X. Gan, W. Zhang, H. Yao, L. Fu, and X. Wang, "Spatio-temporal graph few-shot learning with cross-city knowledge transfer," in *Proc. 28th ACM SIGKDD Int. Conf. Knowl. Discov. Data Mining*, 2022, pp. 1162–1172.
[201] J. Mo and Z. Gong, "Cross-city multi-granular adaptive transfer learning for traffic flow prediction," *IEEE Trans. Knowl. Data Eng.*, vol. 35, no. 11, pp. 11246–11258, Nov. 2023.
[202] X. Liu, Y. Liang, C. Huang, Y. Zheng, B. Hooi, and R. Zimmermann, "When do contrastive learning signals help spatio-temporal graph forecasting," in *Proc. 30th ACM SIGSPATIAL Int. Conf. Adv. Geographic Inf. Syst.*, 2022, pp. 1–12.
[203] Y. You, T. Chen, Y. Sui, T. Chen, Z. Wang, and Y. Shen, "Graph contrastive learning with augmentations," in *Proc. Adv. Neural Inf. Process. Syst.*, 2020, vol. 33, pp. 5812–5823.
[204] R. Li, T. Zhong, X. Jiang, G. Trajcevski, J. Wu, and F. Zhou, "Mining spatio-temporal relations via self-paced graph contrastive learning," in *Proc. 28th ACM SIGKDD Int. Conf. Knowl. Discov. Data Mining*, 2022, pp. 936–944.
[205] J. Ji et al., "Spatio-temporal self-supervised learning for traffic flow prediction," in *Proc. AAAI Conf. Artif. Intell.*, vol. 37, no. 4, 2023, pp. 4356–4364.
[206] R. T. Chen, Y. Rubanova, J. Bettencourt, and D. K. Duvenaud, "Neural ordinary differential equations," in *Proc. Adv. Neural Inf. Process. Syst.*, 31, 2018, pp. 6572–6583.
[207] Z. Fang, Q. Long, G. Song, and K. Xie, "Spatial-temporal graph ODE networks for traffic flow forecasting," in *Proc. 27th ACM SIGKDD Int. Conf. Knowl. Discov. Data Mining*, 2021, pp. 364–373.
[208] M. Jin, Y. Zheng, Y.-F. Li, S. Chen, B. Yang, and S. Pan, "Multivariate time series forecasting with dynamic graph neural odes," *IEEE Trans. Knowl. Data Eng.*, vol. 35, no. 9, pp. 9168–9180, Sep. 2023.
[209] Y. Liang et al., "Mixed-order relation-aware recurrent neural networks for spatio-temporal forecasting," *IEEE Trans. Knowl. Data Eng.*, vol. 35, no. 9, pp. 9254–9268, Sep. 2023.
[210] J. Choi, H. Choi, J. Hwang, and N. Park, "Graph neural controlled differential equations for traffic forecasting," in *Proc. AAAI Conf. Artif. Intell.*, 2022, vol. 36, no. 6, pp. 6367–6374.
[211] G. E. Karniadakis, I. G. Kevrekidis, L. Lu, P. Perdikaris, S. Wang, and L. Yang, "Physics-informed machine learning," *Nature Rev. Phys.*, vol. 3, no. 6, pp. 422–440, 2021.
[212] T. Mallick, P. Balaprakash, E. Rask, and J. Macfarlane, "Transfer learning with graph neural networks for short-term highway traffic forecasting," in *Proc. 25th Int. Conf. Pattern Recognit.*, 2021, pp. 10367–10374.
[213] Y. Huang, X. Song, S. Zhang, and J. J. Q. Yu, "Transfer learning in traffic prediction with graph neural networks," in *Proc. IEEE Int. Intell. Transp. Syst. Conf.*, 2021, pp. 3732–3737.

[214] B. An, A. Vahedian, X. Zhou, W. N. Street, and Y. Li, "HintNet: Hierarchical knowledge transfer networks for traffic accident forecasting on heterogeneous spatio-temporal data," in *Proc. SIAM Int. Conf. Data Mining*, 2022, pp. 334–342.
[215] Y. Tang, A. Qu, A. H. Chow, W. H. Lam, S. Wong, and W. Ma, "Domain adversarial spatial-temporal network: A transferable framework for short-term traffic forecasting across cities," in *Proc. 31st ACM Int. Conf. Inf. Knowl. Manage.*, 2022, pp. 1905–1915.
[216] Y. Liang, G. Huang, and Z. Zhao, "Cross-mode knowledge adaptation for bike sharing demand prediction using domain-adversarial graph neural networks," *IEEE Trans. Intell. Transp. Syst.*, 2023, doi: 10.1109/TITS.2023.3322717.
[217] P. Deng, Y. Zhao, J. Liu, X. Jia, and M. Wang, "Spatio-temporal neural structural causal models for bike flow prediction," 2023, *arXiv:2301.07843*.
[218] Y. Wang, A. Smola, D. Maddix, J. Gasthaus, D. Foster, and T. Januschowski, "Deep factors for forecasting," in *Proc. Int. Conf. Mach. Learn.*, 2019, pp. 6607–6617.
[219] Y. Ovadia et al., "Can you trust your model's uncertainty? evaluating predictive uncertainty under dataset shift," in *Proc. Adv. Neural Inf. Process. Syst.*, 2019, pp. 14003–14014.
[220] X. Wang, H. Liu, C. Shi, and C. Yang, "Be confident! towards trustworthy graph neural networks via confidence calibration," in *Proc. Adv. Neural Inf. Process. Syst.*, 2021, vol. 34, pp. 23768–23779.
[221] D. Wu et al., "Quantifying uncertainty in deep spatiotemporal forecasting," in *Proc. 27th ACM SIGKDD Int. Conf. Knowl. Discov. Data Mining*, 2021, pp. 1841–1851.
[222] H. Wen, Y. Lin, Y. Xia, H. Wan, R. Zimmermann, and Y. Liang, "DIFFSTG: Probabilistic spatio-temporal graph forecasting with denoising diffusion models," 2023, *arXiv:2301.13629*.
[223] Z. Shao, Z. Zhang, F. Wang, and Y. Xu, "Pre-training enhanced spatial-temporal graph neural network for multivariate time series forecasting," in *Proc. 28th ACM SIGKDD Int. Conf. Knowl. Discov. Data Mining*, 2022, pp. 1567–1577.
[224] K. He, X. Chen, S. Xie, Y. Li, P. Dollár, and R. Girshick, "Masked autoencoders are scalable vision learners," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2022, pp. 16000–16009.
[225] W. Zhang, Z. Lin, Y. Shen, Y. Li, Z. Yang, and B. Cui, "Deep and flexible graph neural architecture search," in *Proc. Int. Conf. Mach. Learn.*, 2022, pp. 26362–26374.
[226] P. Xu et al., "Do not train it: A linear neural architecture search of graph neural networks," in *Proc. Int. Conf. Mach. Learn.*, 2023, pp. 38826–38847.
[227] Y. Du et al., "ADARNN: Adaptive learning and forecasting of time series," in *Proc. 30th ACM Int. Conf. Inf. Knowl. Manage.*, 2021, pp. 402–411.
[228] Y. Xia et al., "Deciphering spatio-temporal graph forecasting: A causal lens and treatment," 2023, *arXiv:2309.13378*.
[229] X. Liu et al., "Do we really need graph neural networks for traffic forecasting?," 2023, *arXiv:2301.12603*.
[230] A. Zeng, M. Chen, L. Zhang, and Q. Xu, "Are transformers effective for time series forecasting?," in *Proc. AAAI Conf. Artif. Intell.*, 2023, vol. 37, no. 9, pp. 11121–11128.
[231] S.-A. Chen, C.-L. Li, N. Yoder, S. O. Arik, and T. Pfister, "TSMixer: An all-MLP architecture for time series forecasting," 2023, *arXiv:2303.06053*.

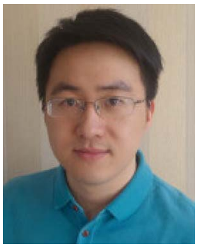
Guangyin Jin received the PhD degree from the College of Systems Engineering, National University of Defense Technology. He is currently an assistant research fellow with the National Innovative Institute of Defense Technology. He has authored or coauthored more than ten papers in JCR Q1-level international journals, such as *IEEE Transactions on Intelligent Transportation Systems*, TIST, TRC, INS, and top international conferences, such as AAAI, CIKM, and SIGSPATIL. His research interests include spatial-temporal data mining, graph neural networks, and urban computing.

Yuxuan Liang (Member, IEEE) received the PhD degree from the National University of Singapore. He is currently an assistant professor with the Intelligent Transportation Thrust, Hong Kong University of Science and Technology (Guangzhou). He is working on the research, development, and innovation of spatio-temporal data mining and AI, with a broad range of applications in smart cities. He has authored or coauthored more than 40 peer-reviewed papers in refereed journals and conferences, such as KDD, WWW, NeurIPS, ICLR, ECCV, AAAI, IJCAI, Ubicomp, and TKDE. Those papers have been cited more than 2,000 times (Google Scholar H-Index: 21). He was recognized as one out of ten most innovative and impactful PhD students focusing on data science in Singapore by Singapore Data Science Consortium (SDSC).

Yuchen Fang working toward the PhD with the University of Electronic Science and Technology of China. He has authored or coauthored several papers in top journals and conference proceedings, such as ICDE, SIGIR, AAAI, and *IEEE Transactions on Intelligent Transportation Systems*. His general research interests include spatio-temporal data mining, graph neural networks, and urban computing, with a special focus on traffic forecasting.

Zezhi Shao received the BE degree from Shandong University, Jinan, China, in 2019. He is currently working toward the PhD degree with the Institute of Computing Technology, Chinese Academy of Sciences, China. He has authored or coauthored several papers in top journals and conferences, such as *IEEE Transactions on Knowledge and Data Engineering*, SIGKDD, VLDB, and CIKM. His research interests include traffic forecasting, multivariate time series forecasting, graph neural networks, and spatial-temporal data mining.

Jincai Huang is currently a professor with the College of Systems Engineering of National University of Defense Technology, Changsha, Hunan, China. He has authored or coauthored more than 80 SCI / EI indexed papers. His main research interests include artificial general intelligence, deep reinforcement learning, and multi-agent systems. He also is the director of the Machine Learning Committee of the Chinese Association for Artificial Intelligence.



Junbo Zhang (Senior Member, IEEE) is currently a senior researcher of JD intelligent cities research. He is leading the Urban AI Product Department of JD iCity with JD Technology, and AI Lab of JD Intelligent Cities Research. Prior to that, he was with Micorsoft Research Asia from 2015 to 2018. He has authored or coauthored more than 60 research papers in spatio-temporal data mining and AI, urban computing, deep learning, and federated learning. He is an associate editor for *ACM Transactions on Intelligent Systems and Technology*. He was the recipient of a number of honors, including the Second Prize of the Natural Science Award of the Ministry of Education in 2021, 22nd and 23rd China Patent Excellence Award in 2021 and 2022, CCF Award for Science and Technology Award–Technological Progress Outstanding Award in 2021. He is an ACM/CCF Senior Member.

Yu Zheng (Fellow, IEEE) is currently the vice president of JD.COM and head JD Intelligent Cities Research. Before Joining JD.COM, he was a senior research manager with Microsoft Research. He is the editor-in-chief of *ACM Transactions on Intelligent Systems and Technology* and was the program co-chair of ICDE 2014 (Industrial Track), CIKM 2017 (Industrial Track) and IJCAI 2019 (industrial track). He is also a keynote speaker of AAAI 2019, KDD 2019 Plenary Keynote Panel and IJCAI 2019 Industrial Days. His monograph, entitled Urban Computing, has been used as the first text book in this field. In 2013, he was named one of the Top Innovators under 35 by MIT Technology Review (TR35) and featured by Time Magazine for his research on urban computing. In 2016, he was named an ACM Distinguished Scientist and elevated to an IEEE Fellow in 2020 for his contributions to spatio-temporal data mining and urban computing.