Applied Intelligence
https://doi.org/10.1007/s10489-021-02587-w

Spatial-temporal graph neural network for traffic forecasting: An overview and open research issues

Khac-Hoai Nam Bui[1] · **Jiho Cho**[2] · **Hongsuk Yi**[2]

Accepted: 2 June 2021

Abstract

Traffic forecasting plays an important role of modern Intelligent Transportation Systems (ITS). With the recent rapid advancement in deep learning, graph neural networks (GNNs) have become an emerging research issue for improving the traffic forecasting problem. Specifically, one of the main types of GNNs is the spatial-temporal GNN (ST-GNN), which has been applied to various time-series forecasting applications. This study aims to provide an overview of recent ST-GNN models for traffic forecasting. Particularly, we propose a new taxonomy of ST-GNN by dividing existing models into four approaches such as graph convolutional recurrent neural network, fully graph convolutional network, graph multi-attention network, and self-learning graph structure. Sequentially, we present experimental results based on the reconstruction of representative models using selected benchmark datasets to evaluate the main contributions of the key components in each type of ST-GNN. Finally, we discuss several open research issues for further investigations.

Keywords Traffic forecasing · Deep learning · Graph neural network · Spatial-temporal graph neural network

1 Introduction

The successful development of deep learning (DL), which is based on artificial neural networks, has revolutionized many machine learning (ML) tasks, and traffic forecasting is no exception. Specifically, end-to-end DL paradigms, such as Deep Belief Networks (DBNs) [37], recurrent neural networks (RNNs) [26, 32], Convolutional Neural Network (CNN) [5, 13], and Autoencoders (AEs) [9], have been successfully developed for improving the performance of the forecasting problem. Particularly, these paradigms can extract latent representations and exploit much more features of traffic data than traditional ML methods. For instance, CNN model are able to applied for extracting traffic data by modeling traffic road networks as grid structure [17]. However, they are not optimal methods for traffic forecasting since road networks are graph-based structure [2].

Recently, the graph neural network (GNN) has been introduced as a new DL paradigm for learning non-Euclidean data by applying graph analysis methods [7]. According to the authors of [39], GNNs can be categorized into four groups, namely recurrent GNNs (RecGNNs), convolutional GNNs (ConvGNNs), graph autoencoders (GAEs), and spatial-temporal GNNs (ST-GNNs). In this study, we focused on ST-GNN models for traffic prediction by learning hidden patterns of spatial-temporal graphs, which have made significant progress in recent years. Specifically, ST-GNN is based on the concept of simultaneously modeling spatial and temporal dependencies to deal with a dynamic graph problem. Therefore, this study aims to provide an overview of ST-GNN models in terms of mathematical methods and main

This work was partly supported by Institute for Information & communications Technology Promotion(IITP) grant funded by the Korean Ministry of Science and ICT (MSIT) (No. 2018-0-00494, Development of deep learning-based urban traffic congestion prediction and signal control solution system) and Korea Institute of Science and Technology Information(KISTI) grant funded by the Korean Ministry of Science and ICT (MSIT) K-20-L02-C09-S01). Corresponding Author: Hongsuk Yi.

✉ Hongsuk Yi
hsyi@kisti.re.kr

Khac-Hoai Nam Bui
hoainam.bk2012@gmail.com

Jiho Cho
jhcho@kisti.re.kr

1 Viettel Cyberspace Center, Hanoi, Vietnam

2 Korea Institute of Science and Technology Information, Seoul, South Korea

Published online: 21 June 2021

components for the traffic forecasting problem. Figure 1 provides a general summary of different methods that have been used to develop ST-GNN models for the spatial-temporal forecasting. Furthermore, Table 1 shows the detailed descriptions of general notations that we used in the paper.

1.1 Problem definition

Let $X_i = \{x_i^1, ..., x_i^T\}$ denotes time-series data where x_i^T represents the observed values of a location/sensor i at timestamp T. A traffic dataset at an certain area is a set of N time series $X = \{X_1, ..., X_N\} \in \mathbb{R}^{N \times T}$. Consequentially, traffic forecasting problem is to predict the next time-steps based on S time-steps of historical data, which can be formulated as follows:

$$\theta^* = \arg\min_{\theta} \sum_{t=1}^{T} \mathcal{L}_\theta(X^{(t-S):t}, X^{(t+1):(t+T)}) \tag{1}$$

where $X^{(t-S):t}$ and $X^{(t+1):(t+T)}$ denote the input and output values, respectively. In a neural network, the optimal parameter θ^* represents the weights of the model for solving the optimization problem of the non-linear loss function $\mathcal{L}_\theta$. Recently, spatial-temporal forecasting using GNNs has become a promising method for learning the non-Euclidean structures of road traffic networks by considering the spatial-temporal data as a dynamic graph problem, which includes a static graph structure and dynamic input signals [39]. Let $G(V, E, A)$ denote a graph structure representing spatial-temporal data, the traffic forecasting problem using ST-GNN-based models can be formulated as follows:

$$[X^{(t-S):t}, G] \xrightarrow{f} X^{(t+1):(t+T)} \tag{2}$$

where f denotes the learning function.

1.2 Contributions and organization

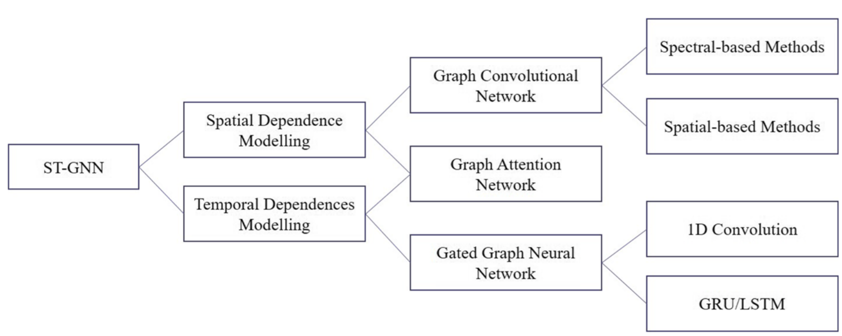
This paper aims to provide an overview of recent ST-GNN models for the traffic forecasting problem. Specifically, [28] and [39] are well-known survey papers on GNNs. Zhang et al. [36] presented a specific review of graph convolutional networks (GCNs), which are the most prominent GNN variants. In terms of traffic forecasting, [33] provided a comprehensive survey, covering traditional ML and DL models. In comparison with the aforementioned articles, this study focuses on ST-GNN models for traffic forecasting using spatial-temporal correlation. To the best of our knowledge, this is the first study for exploiting the potential solutions of ST-GNN for spatial-temporal datasets. In particular, the main contributions of this study is threefold as follows:

- We provide an overview of the state-of-the-art ST-GNN models for traffic forecasting. Moreover, we define a new taxonomy of ST-GNN for further exploitation in this research field. Specifically, with each type of ST-GNN, we provide detailed descriptions in terms of mathematical methods and model architectures of representative models.
- We examine the comparative experiment with selected benchmark datasets to evaluate the performance of the key components of representative models of each approach. Specifically, we considered two datasets: METR-LA, which is a well-known public traffic dataset, and our self-collected dataset, UVDS, in order

Table 1 List of general notations used in the paper

Notation	Description
$\mathbb{R}^n$	n-dimensions of euclidean space
N	number of nodes in the given graph
A, A_i^τ	adjacency matrix, transpose of the adjacency matrix
I, P	identity matrix, transition matrix
X, H, Z	input features, hidden state, output
S, T	look back and look forward steps of the prediction
$*_\mathcal{G}x$, $*_\mathcal{T}x$	convolution and temporal gated operation of signal x
h_i^k	hidden sate of node i at step k
tanh, LeakyReLU, Softmax	hyperbolic tangent, LeakyReLU, and softmax functions
$\odot$, $\|$	element-wise multiplication and concatenation operations
e, α	normalized and attention coefficients

Fig. 1 Overview of ST-GNN-based methods



to identify the effectiveness of each key component for further investigations.

– We discuss several limitations of the current ST-GNN models and propose several open research issues for the future research direction in the spatial-temporal forecasting domain. Specifically, ST-GNN models are able to extend for other forecasting problems such as Solar-Energy, Electricity, Exchange-Rate, and Human Action Recognition.

The remainder of this paper is organized as follows. In Section 2, we briefly review several GNN variants that form the basis for developing ST-GNN models. A new taxonomy for state-of-the-art ST-GNN models is presented in Section 3. Section 4 presents the experimental results for the evaluation of the aforementioned models with selected benchmark datasets. Open research issues for further investigation are discussed in Section 5. Finally, Section 6 concludes the paper.

2 Graph neural network variants

The original concept of GNN approach has several limitations, such as the node state updating problem. Therefore, several GNN variants have been sequentially introduced to improve the performance of the original GNN in terms of training methods and propagation steps. According to the survey study in [39], three well-known GNN variants are the GCN, graph attention network (GAT), and gated graph neural network (GGNN), which have been applied to many ML tasks with significantly improved performance. In this section, we generalize each variant using mathematical formulations, which are regarded as the main components for the development of ST-GNN models.

2.1 Graph convolutional network

The GCN plays a central role in the development of complex GNN models, where the main idea is to represent nodes in a graph by aggregating features from their neighboring nodes. Technically, GCN include two types of graph convolution operations such as spectral-based and spatial-based methods [36].

Spectral-based Methods: Spectral techniques have been widely used in graph signal processing. Specifically, convolutions can be calculated by computing the eigendecomposition of the graph Laplacian, in which the normalized graph is calculated as follows:

$$L = I_N - D^{-1/2} A D^{-1/2} \tag{3}$$

where $D \in \mathbb{R}^{N\times N}$ and I_N represent the diagonal degree matrix and identity matrix, respectively. Subsequently, the study in [14], which is the most cited works in graph learning, has defined the convolution of the graph signal x with N nodes and a filter $\Theta \in \mathbb{R}^n$ as follows:

$$\Theta *_{\mathcal{G}} x = \sum_{k=0}^{K-1} \theta_k T_k(\tilde{L}) x \tag{4}$$

where K represents the order polynomial in the Laplacian. T denotes the Chebyshev polynomials [8] using $\tilde{L} = \frac{2}{\lambda_{max}} L - I_N$ where λ_{max} is the largest eigenvalue of L. Consequently, the output Z, by integrating multiple convolutions, is formulated as follows:

$$\begin{aligned} Z &= \tilde{D}^{-1/2} \tilde{A} \tilde{D}^{-1/2} X \Theta \\ &s.t.: \tilde{A} = A + I_N \\ &\quad \tilde{D}_{ii} = \sum \tilde{A}_{ij} \end{aligned} \tag{5}$$

where X is the matrix of node features and Θ denotes a matrix of kernel/filter parameters that can be shared for the whole graph.

Spatial-based Methods The spatial approach directly performs convolution in the graph domain by aggregating the information from neighboring nodes, which can handle the scalability problem for large graphs because the entire graph is processed simultaneously. In particular, the spatial GCN is based on classic CNN-based models, which were successfully developed for regular data structures (e.g., image grids) and irregular data structures (such as social graphs, 3D meshes, molecular graphs, and road traffic networks). Specifically, the authors in [1] propose a diffusion convolutional neural network (DCNN), which defines graph convolutions as a diffusion process. Consequently, the output of the model is the concatenation of compositional layers, in which the k^{th} layer of the diffusion graph convolution is formulated as follows:

$$\begin{aligned} H^k &= f(W^k \odot P^k X) \\ &s.t.: P = D^{-1} A \end{aligned} \tag{6}$$

where $f(.)$ represents the activation function, and $P \in \mathbb{R}^{N\times N}$ and W denote the probability transition and model parameter matrices, respectively. Sequentially, a modification of the final output was proposed in [16], which calculates a sum at each diffusion step as follows:

$$Z = \sum_{k=0}^{K-1} f(P^k X W^k) \tag{7}$$

Another well-known model is GraphSAGE [12], which is a general inductive framework that generates embeddings using sampling and aggregating features from neighboring nodes. This model has been applied to various applications with promising results.

2.2 Graph attention network

Attention mechanisms have been widely applied in the natural language processing domain for learning sequence-based tasks (e.g., machine translation). Specifically, the main idea of the attention method is to extract relevant features. Recent studies have applied attention techniques for the propagation process in a GNN. Accordingly, the GAT was introduced by Velickovic et al. [25], adopting attention to learn the relative weights between two neighboring nodes. In general, the graph convolution operation is formulated as follows:

$$h_i^0 = x_i$$
$$h_i^k = \sigma\left(\sum_j \alpha_{ij}^k W^k h_j^{k-1}\right)$$
$$z_i = h_i^K \tag{8}$$

where σ denotes the non-linear transformation (e.g., rectified linear unit). The final output of node i, Z_i, is the embedding process after K layers of the aggregation from its neighbor nodes based on the normalized coefficients α_{ij} for measuring the importance of neighbor nodes j to node i ($j \in N(i)$). Specifically, the attention coefficients (e.g., e_{ij}) are normalized to enable a comparison across all neighborhoods, which are formulated as follows:

$$e_{ij} = a(W_k h_i^{k-1}, W_k h_j^{k-1}) \tag{9}$$

$$\alpha_{ij} = \frac{exp(LeakyReLU(e_{ij}))}{\sum_m exp(LeakyReLU(e_{ik}))}$$
$$s.t. : \forall m \in N(i), m \neq j \tag{10}$$

where the vector a represents learnable parameters. Sequentially, multi-head attention [24] has been employed to improve the stability of the learning process using K independent attention mechanisms. Specifically, the final computation of the learning process can be represented as follows:

$$z_i = \sigma\left(\frac{1}{K}\sum_{k=0}^{K-1}\sum_j \alpha_{ij}^k W^k h_j\right) \tag{11}$$

Technically, the main difference between GATs and GCNs is that a GAT assigns larger weights to the more important neighboring nodes, which is suitable for dynamic input features of traffic datasets.

2.3 Gated graph neural network

Several studies have focused on gate mechanisms (such as gated recurrent unit (GRU) and long short-term memory (LSTM)) for the propagation process, which can improve the restrictions in the original GNN approach, particularly for long-term propagation across the graph structure. Specifically, the GGNN was introduced in [15] where a GRU was applied for the propagation step. Therefore, a general node hidden state can be updated as follows:

$$h_i^0 = x_i$$
$$h_i^k = GRU\left(h_i^{k-1}, \sum_{j \in N(i)} W h_j^{k-1}\right) \tag{12}$$

Particularly, the updating rate of the temporal state using GRU-like update functions can be formulated as follows:

$$h_i^t = u^t \odot \hat{h}_i^t + (1 - z^t) \odot h_i^{t-1}$$
$$s.t. : z_i^k = \sigma(W^z a_i^t + U^z h_i^{t-1})$$
$$r_i^t = \sigma(W^r a_i^t + U^r h_i^{t-1})$$
$$\hat{h}_i^t = tanh(W a_i^t + U(r_i^t \odot h_i^{t-1})) \tag{13}$$

where r^t and z^t represent reset and update gates at time t, respectively. h_i^t and $\hat{h}_i^{t-1}$ denote the temporal and previous state of node i, which are calculated based on the aggregation a from neighbor nodes in graph structure. Consequentially, a_i^t is formulated as follows:

$$a_i^t = A_i^\tau [h_1^{t-1} ... h_N^{t-1}]^\tau + b \tag{14}$$

where the sub-matrix A_i of the adjacency matrix A denotes the connection of node i with its neighbor nodes.

3 Taxonomy of spatial-temporal graph neural network

Conceptually, the spatial-temporal graph structure is dynamic in that the node/edge features change with time. For instance, a road traffic network is a typical application, where road networks can be modeled by a graph structure. Specifically, each node represents a sensor location that monitors the traffic characteristics (e.g., average speeds), which change over time. Figure 2 illustrates a general pipeline for traffic forecasting using spatial-temporal correlation.

In this section, we present recent state-of-the-art ST-GNN models and provide a new taxonomy for further exploitation in this research field.

3.1 Graph convolutional recurrent neural network

As mentioned above, the DCNN is a well-known spatial-based GCN. For the spatial-temporal forecasting problem, [16] presented a diffusion convolutional recurrent neural network (DCRNN) model by incorporating a diffusion GCN into the GRU network to capture spatial and temporal dependencies. Furthermore, the model adopts an encoder-decoder architecture to deal with the multi-step prediction,

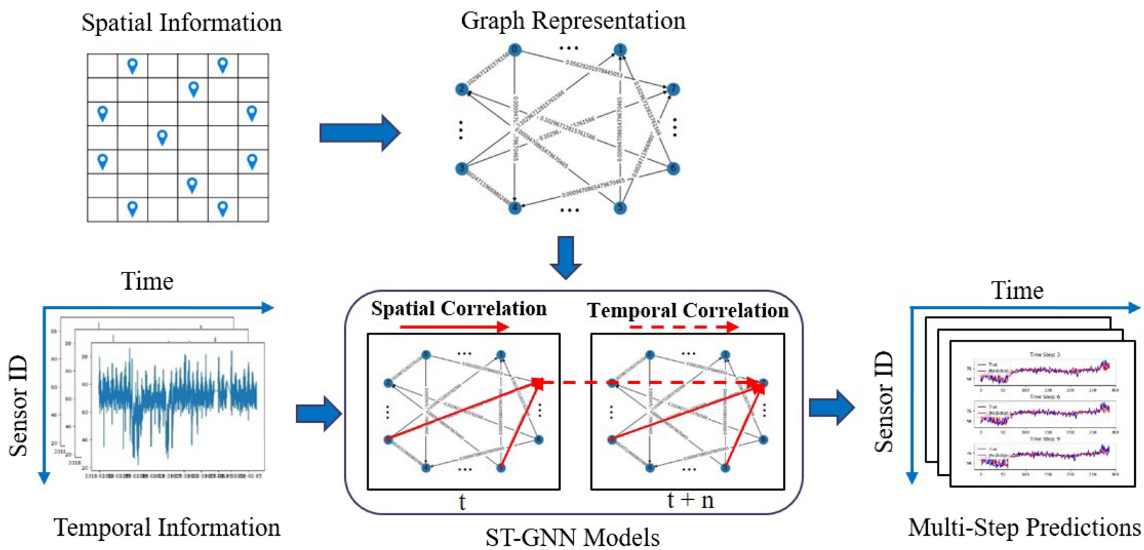


Fig. 2 A general pipeline of ST-GNN models for traffic prediction

which encodes the input sequence into a fixed-length vector and predicts the output of the next T time-steps. Specifically, the graph signal x can be represented by the diffusion convolution operation $*\mathcal{G}$ as follows:

$$\Theta *_{\mathcal{G}} x = \sum_{k=0}^{K} (\theta_{k,1}(D_O^{-1}A)^k + \theta_{k,2}(D_I^{-1}A^{\tau})^k)x \tag{15}$$

where $D_O^{-1}A$ and $D_I^{-1}A^{\tau}$ denote the transition and reverse matrices of the diffusion process, respectively. Then, the output of the spatial information is fed into a GRU network in which the diffusion convolution operation replaces the matrix multiplications in the GRU as follows:

$$\begin{aligned} r^t &= \sigma(W_{r*\mathcal{G}}[x^t, h^{t-1}] + b_r) \\ z^t &= \sigma(W_{z*\mathcal{G}}[x^t, h^{t-1}] + b_z) \\ \hat{h}^t &= tanh(W_{*\mathcal{G}}[x^t, (r^t \odot h^{t-1})] + b_r) \\ h^t &= u^t \odot \hat{h}^t + (1 - u^t) \odot h^{t-1} \end{aligned} \tag{16}$$

Where x^t and h^t represent the input and output at time t, respectively. In particular, the main drawback of the DCRNN is computational and memory bottlenecks, which make it difficult to scale up with large networks. In this regard, the study of [23] presented a graph partitioning-based method for implementing the DCRNN in large-scale networks by dividing the graph G into m subgraphs ($G = G_1, ..., G_m$) to be trained independently.

3.2 Fully graph convolutional network

To address the computational problem caused by using recurrent units, Yu et al. [34] proposed the STGCN, a complete convolutional structure model for spatial-temporal forecasting. Specifically, STGCN model consists of several spatial-temporal blocks, where each block includes one graph convolution alternating in two convolutional sequence learning (1D convolution) layers. In particular, a graph convolution operator for extracting spatial features is employed following the concept of the spectral-based method using the Chebyshev polynomial (4). Furthermore, for capturing the dynamic temporal information, instead of using RNN-based models, the model uses 1D causal convolution with gated linear units (GLU). Specifically, the temporal gated convolution operator ($*_{\mathcal{T}}$) can be defined as follows:

$$\Gamma *_{\mathcal{T}} Y = P \odot \sigma(Q) \tag{17}$$

where P and Q are equal split values from input Y, which include sequences of each node in the graph and are mapped by the convolution kernel Γ. Therefore, assuming z^b denotes the input of a block b in the STGCN model, the output z^{b+1} can be generalized as follows:

$$z^{b+1} = \Gamma_1^b *_{\mathcal{T}} ReLU(\Theta^b *_{\mathcal{G}} (\Gamma_0^b *_{\mathcal{T}} z^b)) \tag{18}$$

where Γ_1^b and Γ_0^b denote the temporal kernels of the two corresponding convolutional sequence learning layers. Intuitively, compared with other RNN-based models, the STGCN can improve the training process with fewer parameters. The attention-based STGCN (ASTGCN) model [11] adopts the basic architecture of the STGCN to capture the dynamics of correlations of spatial-temporal dimensions by leveraging two attention layers. Specifically, (4) can be reformulated for the ASTGCN as follows:

$$\Theta *_{\mathcal{G}} x = \sum_{k=0}^{K-1} \theta_k(T_k(\tilde{L}) \odot S)x \tag{19}$$

where S denotes the spatial attention matrix, which is integrated into the Chebyshev polynomial $T_k\tilde{L}$ using the Hadamard product $\odot$.

Graph WaveNet (GraphWN) is a novel GNN architecture, which is a variant of the fully GCN for spatial-temporal graph modeling [30]. Specifically, the core idea of this approach is to handle long-range temporal sequences, based on the concept of WaveNet [19] by using stacked dilated casual convolutions. Mathematically, given a 1D sequence x and filter Γ, the dilated causal convolution operation at step t can be represented as follows:

$$\Gamma(t) *_{\mathcal{T}} x = \sum_{s=0}^{K-1} \Gamma(s)x(t - d \times s) \tag{20}$$

where d is the dilation factor. Moreover, an adaptive matrix based on the node embedding method has been adopted to capture hidden spatial dependency to improve the performance. In particular, considering the diffusion process for a directed graph, (16) can be generalized as follows:

$$Z = \sum_{k=0}^{K-1} P_f^k X W^{k1} + P_b^k X W^{k2} \tag{21}$$

where P_f and P_b denote the forward and backward transition matrices, respectively. Consequently, an adaptive matrix is introduced with end-to-end learning through stochastic gradient descent using the node embedding method, which is formulated as follows:

$$\tilde{A} = Softmax(ReLU(E_1 E_2^{\tau})) \tag{22}$$

where E_1 and E_2 are the learning parameters of the source and target node embeddings, respectively. Therefore, (21) can be updated as follows:

$$Z = \sum_{k=0}^{K-1} P_f^k X W^{k1} + P_b^k X W^{k2} + \tilde{A}^k X W^{k3} \tag{23}$$

Technically, GraphWN is currently one of the most advanced ST-GNN models for forecasting problems and has achieved promising results on various time-series datasets. Consequently, recent studies have adopted the concept of GraphWN for further investigation in this research field [10] [29].

3.3 Graph multi-attention network

Recently, attention-based models have been widely applied in various research fields, and the ST-GNN is no exception. The motivation is to improve long-term prediction, which is a remaining challenge using convolution modules. Zheng et. al. [38] presented the graph multi-attention network (GMAN) model, which includes multi-attention blocks in an encoder-decoder architecture for modeling spatial-temporal correlations. Accordingly, following the (10), the attention coefficients ($e_{i,j}$) between two nodes (i and j) at time t of the spatial attention ($e_{i,j}^{sp}$) are calculated as follows:

$$e_{i,j}^{sp} = \frac{(h_{i,t}^{k-1} \parallel E_{i,t}, h_{j,t}^{k-1} \parallel E_{j,t})}{\sqrt{2D}} \tag{24}$$

where $(\cdot,\cdot)$ represents the dot-product operator [24], $\parallel$ denotes the concatenation operation, $E_{i,t}$ is the spatial-temporal embedding of node i with time t, which is considered in both spatial and temporal embedding domains, and 2D represents the dimension. Similarity, the temporal attention between two time-steps t and t' at node i ($e_{t,t'}^{te}$) is defined as follows:

$$e_{t,t'}^{te} = \frac{(h_{i,t}^{k-1} \parallel E_{i,t}, h_{i,t'}^{k-1} \parallel E_{i,t'})}{\sqrt{2D}} \tag{25}$$

The spatial-temporal graph attention network (ST-GRAT) is a recent ST-GNN model that uses graph attention (GAT) to capture the dynamics of spatial-temporal information [21]. Specifically, compared with other attention-based ST-GNN models, ST-GRAT adds a new self-attention module for determining specific relevant nodes. In particular, existing attention-based methods assume that the total weight must be one, which is not effective in all cases. In this regard, the (8) can be revised as follows:

$$h_i^k = (1 - \sum_j \alpha_{ij}^k) W_v^k h_i^{k-1} + \sum_j \alpha_{ij}^k W^k h_j^{k-1} \tag{26}$$

where W_v indicates the linear transformation matrix of the sentinel value vector, which is calculated based on the concept of the sentinel mixture model [18]. Specifically, recent studies have focused on GAT as a promising approach for capturing diverse spatial correlations in this research field [20, 35].

3.4 Spatial-temporal graph structure learning

Most ST-GNN models are developed based on defined graph structures for the propagation process, which cannot be directly extended to multivariate problems. In this regard, recent studies have introduced graph learning layers to automatically learn hidden spatial dependencies among nodes by data-driven perspectives. Specifically, based on the basic concept of GraphWN, Wu et al. [29] proposed a new model, MTGNN, for learning a non-predefined graph structure. Accordingly, a graph learning layer was designed to learn graph adjacency, which is the input for the convolution modules. Sequentially, the adaptive graph adjacency can be formulated as follows:

$$\begin{aligned} \tilde{A} &= ReLU(tanh(\omega(E_1 E_2^{\tau} - E_2 E_1^{\tau}))) \\ s.t.:\ E_1 &= \tanh(\omega E_1^0 W_1) \\ E_2 &= \tanh(\omega E_2^0 W_2) \end{aligned} \tag{27}$$

Table 2 Benchmark datasets

Type	Dataset	Sample	Node	Time Interval	Source
Speed	METR-LA	34,272	207	5 min	Li et al. [16]
Flow	UVDS	25,632	104	5 min	Bui et al. [3]

Table 3 Baseline ST-GNN models

Approach	Model	Source(https://github.com)
Diffusion Convolution	DRCNN	/liyaguang/DCRNN
Fully GCN	STGCN	/VeritasYin/STGCN_IJCAI-18
Graph WaveNet	GraphWN	/nnzhan/Graph-WaveNet
Graph Attention	GMAN	/zhengchuanpan/GMAN
Adaptive Learning	MTGNN	/nnzhan/MTGNN

where E_1^0 and E_2^0 are randomly initialized node embedding with model parameters W. ω denotes a hyperparameter for controlling the saturation rate of the activation function. In contrast, Cao et al. [6] adopted self-attention mechanisms for the learning layer. In particular, the query Q, key K, and the output matrix W, which is regarded as the adjacency weight matrix for the input graph, can be sequentially formulated as follows:

$$\begin{aligned} Q &= RW^Q \\ K &= RW^K \\ \tilde{A} = W &= Softmax(\frac{QK^T}{\sqrt{d}}) \end{aligned} \tag{28}$$

4 Experimental analysis

In this section, we present the implemented results several typical ST-GNN models across different benchmark datasets to identify the advantages and disadvantage of each approach.

4.1 Experiment setup

Two benchmark datasets were considered for implementation: METRLA and UVDS. Table 2 summarizes the two datasets in more detail.

Specifically, METR-LA is a well-known traffic speed dataset that has collected from the highway system of Los Angeles County and preprocessed by Li et al. [16]. The data are from 207 sensors over 4 months. UVDS is a traffic dataset that we collected and preprocessed from vehicle detection systems at an urban area [3]. In particular, Fig. 3 illustrates the map locations of the two benchmark datasets, respectively.

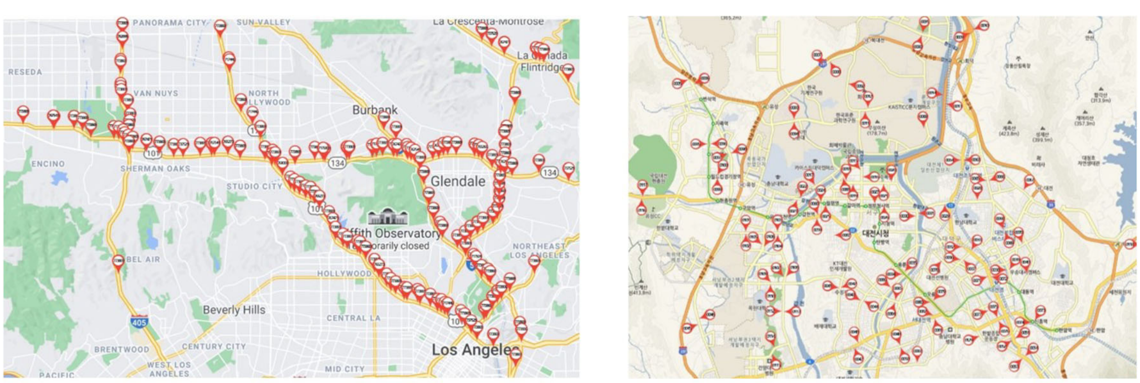
Specifically, the standard time interval in both two datasets are set to 5 min and normalized by using Z-score method. For each dataset, we use 70% for training, 10% for validation, and 20% for testing. The adjacency matrix of the road network graph is constructed based on the distance d between sensors as follows:

$$A_{i,j} = \begin{cases} exp(-\frac{d^2_{v_i,v_j}}{\sigma^2}), & \text{if } exp(-\frac{d^2_{v_i,v_j}}{\sigma^2}) \geq \beta \\ 0, & \text{otherwise.} \end{cases} \tag{29}$$

where β denotes the threshold for controlling the sparsity of the adjacency matrix, which is assigned to 0.1 for both datasets.

Regarding the ST-GNN models, DCRNN, STGCN, GraphWN, GMAN, and MTGNN are taken into account, in which each model represents a typical method of each approach in Section 3. Specifically, we re-constructed these models following the source codes which are provided by the authors (Table 3).

For the hyperparameter setting, we re-use the configurations from original proposals. Particularly, more details of



a) METR-LA   b) UVDS

Fig. 3 Selected benchmark datasets

the reprehensive models for the evaluation are described as follows:

- DCRNN [16]: The best hyperparameters are chosen using the Tree-structured Parzen Estimator (TPE). Specifically, two recurrent layers are adopted for both encoder and decoder in which the unit is set to 64. The maximum step of random walks is set to 3 (K=3). The model is trained using Adam optimizer with the initial learning rate is set to $1e^{-}2$.
- STGCN [34]: The model includes two spatio-temporal convolutional blocks (ST-Conv blocks) in which each block contains one spatial graph convolution layer and two temporal gated convolution layers. The channels of the layers are set to 64, 16, and 64, respectively. Both graph convolution kernel size and temporal convolution kernel size are set to 3. The model is trained with the initial learning rate is set to $10^{-}3$ and using AdamW optimizer.
- GraphWN [30]: The model consists of eight spatial-temporal layers. The diffusion step is set to 2 (K = 2). The model is trained using Adam optimizer with the initial learning rate is set to $10^{-}3$.
- GMAN [38]: The number of attention blocks, attention heads , and the dimension of each attention head are set to 3, 8, and 8, respectively. The model is trained using Adam optimizer with the initial learning rate is set to $10^{-}3$.
- MTGNN [29]: The number of graph convolution and temporal convolution modules are both set to 3. The dilation exponential factor is set to 1. The number of neighbors for each node is set to 20. The model is trained using Adam optimizer with the initial learning rate is set to $10^{-}3$.

Furthermore, all the experiments are tested with 60 min as the historical time window (S = 12) to forecast traffic conditions in the next 60 min (T = 12). The experiments were implemented on a PC with a Core i7 16 GB CPU and 32 GB GPU memory, where the GPU was used for acceleration.

4.2 Experimental results

Three metrics were used for evaluation: mean absolute error (MAE), root mean squared error (MAPE), and mean absolute percentage error (MRSE), which are sequentially formulated as follows:

$$MAE = \frac{1}{n}\sum_{i=1}^{n} \mid y_i - \widehat{y_i} \mid$$

$$MAPE = \frac{1}{n}\sum_{i=1}^{n} \frac{\mid y_i - \widehat{y_i} \mid}{y_i}$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n} (\mid y_i - \widehat{y_i} \mid)^2} \qquad (30)$$

where y and $\widehat{y}$ are observation and prediction values, respectively. Tables 4 and 5 show the comparisons of baseline ST-GNN models on the two selected benchmark datasets in terms of accuracy and computational time, respectively.

Accordingly, there are five predicted steps for the future prediction such as 5 and 15 min (short-term prediction), 30 and 45 min (middle-term prediction), and 60 min (long-term prediction). Based on the experimental results, there are several main conclusions can be drawn:

- Currently, there is no model that can achieve the best results for all cases of multi-step prediction. For instance, ST-GCN, by using the fully GCN for extracting both spatial and temporal features, is able to significantly improve the computational time, however, the accuracy of this approach is generally lower than other approaches.
- For extracting spatial feature, spatial-based methods (DCRNN, GraphWN) are slightly better than spectral-based methods (STGCN) for extracting spatial dependencies, especially in the case of large road network (the number of nodes/sensors). Moreover, using RNN-based method is more effective than a 1D CNN in term of extracting temporal feature, especially in the case of short-term prediction. However, we are able to improve this approach by using dilated convolution with multiple filter sizes (GraphWN, MTGNN).
- Using GAT-based methods (GMAN) can perform better than GCN-based methods (STGCN, DCRNN, GraphWN) in extracting spatial correlations, especially for long-time prediction. The main reason is that using the attention mechanism is able to capture the complex dynamic of traffic flow, which changes significantly over time. However, the drawback of GMAN model is the high computational time for training, which require a large number of hyperparameters.
- The comparable results of adaptive learning-based approach (MTGNN, an improved model of GraphWN by adding a self-adaptive learning layer) with GAT-based approach (GMAN) indicate the importance of learning the dynamics of traffic dataset for improving the performance, which is a potential research issue of forecasting problem using GNN.

5 Open research issues

In recent years, ST-GNN models have provided the capability of traffic forecasting with promising results.

Table 4 Experiment results for multi-step predictions

Dataset	Step	Metrics	DCRNN	STGCN	GraphWN	GMAN	MTGNN
METR-LA		MAE	2.177	2.820	2.240	2.458	2.253
	1	RMSE	3.774	5.330	3.863	4.398	3.965
		MAPE	5.183%	6.543%	5.472%	6.210%	5.462%
	3	MAE	2.673	3.900	2.696	2.817	2.682
		RMSE	5.184	7.795	5.159	5.511	5.197
		MAPE	6.843%	9.374%	6.988%	7.538%	6.893%
	6	MAE	3.078	4.986	3.077	3.119	3.034
		RMSE	6.312	10.077	6.192	6.372	6.162
		MAPE	8.393%	12.079%	8.355%	8.807%	8.229%
	9	MAE	3.350	5.758	3.333	3.314	3.279
		RMSE	7.008	11.710	6.843	6.898	6.771
		MAPE	9.471%	13.997%	9.310%	9.626%	9.146%
	12	MAE	3.562	6.525	3.510	3.458	3.479
		RMSE	7.524	13.063	7.268	7.249	7.320
		MAPE	10.339%	15.802%	10.031%	10.215%	9.849%
UVDS		MAE	3.492	3.513	3.462	3.610	3.459
	1	RMSE	5.311	5.395	5.271	5.511	5.264
		MAPE	7.476%	7.136%	7.337%	7.791%	7.380%
	3	MAE	3.700	3.767	3.666	3.694	3.650
		RMSE	5.685	5.810	5.617	5.739	5.585
		MAPE	8.111%	7.714%	7.985%	8.037%	7.969%
	6	MAE	3.855	4.015	3.809	3.776	3.774
		RMSE	5.947	6.168	5.870	5.915	5.803
		MAPE	8.554%	8.249%	8.437%	8.264%	8.335%
	9	MAE	3.952	4.239	3.881	3.828	3.845
		RMSE	6.096	6.457	5.981	6.004	5.918
		MAPE	8.822%	8.753%	8.653%	8.405%	8.552%
	12	MAE	4.031	4.452	3.934	3.878	3.924
		RMSE	6.216	6.714	6.054	6.076	6.030
		MAPE	9.038%	9.274%	8.780%	8.530%	8.804%

Table 5 Computational time (s) of training and inference in each epoch

Dataset	Metrics	DCRNN	STGNN	GraphWN	GMAN	MTGNN
METR-LA	Training	122.5	12.2	59.3	663.9	53.6
	Inference	10.8	10.8	1.8	27.9	1.6
UVDS	Training	84.1	5.4	29.3	242	27.6
	Inference	6.9	4.7	0.9	9.9	0.8

However, there are still many issues that require further investigation. In this section, we discuss several open research issues in this field.

5.1 Dynamic spatial dependencies for multi-step forecasting

Multi-step forecasting plays an essential role in the development of intelligent transportation systems. For instance, estimating an accurate multi-step traffic flow can enable a dynamic traffic light control system [4]. Most current ST-GNN models assume that the spatial dependencies are fixed with connectivity and distance with different time-steps. However, in a real-world traffic network, spatial dependencies are dynamic with different time-steps, which are based on many other factors, such as accidents, weather conditions, and rush and non-rush hours. Therefore, an investigation on how to develop models for capturing the dynamical spatial dependencies to improve performance across multi-step predictions is required.

5.2 Multivariate forecasting with different traffic characteristic

The multivariate time-series forecasting problem assumes interdependence among variables and has become the main challenge for DL-based methods for traffic forecasting in terms of fully exploiting latent dependencies among variables [22]. Current ST-GNN models are not able to extract spatial dependency effectively because they depend on a predefined graph structure. Consequently, a general framework for multivariate forecasting using GNNs is an emergent issue in this research field.

5.3 Graph theory-based road network graph construction

As mentioned above, the aforementioned models require well-defined graph structures for information propagation. Therefore, building an appropriate adjacency matrix for a road network graph plays an important role. Therefore, using graph theory to represent road traffic networks with deeper semantic information can significantly improve prediction performance [31].

5.4 Real-time prediction task

Real-time prediction is a significant challenge in this research field. Specifically, DL-based models require a large amount of data, which is the main reason that they are not able to achieve real-time processing. Therefore, designing lightweight ST-GNN models by reducing the sizes of parameters will be a great achievement in this research field [27].

6 Conclusion

Recently, with the rapid development of GNNs, traffic forecasting using spatial-temporal correlation has become an emerging topic in this research field. Specifically, many ST-GNN models have been introduced to improve the performance of traffic forecasting for long-term prediction. In this paper, we first summarized several well-known GNN variants, such as the GCN, GAT, and GGNN, which serve as the basis for developing a general framework for traffic forecasting using spatial-temporal dependencies. State-of-the-art ST-GNN models were then classified for their further exploitation. Finally, open research issues were presented based on experimental results on selected benchmark datasets for future development in this research field.

References

1. Atwood J, Towsley D (2016) Diffusion-convolutional neural networks. In: Proceedings of the 29th annual conference on neural information processing systems (NeurIPS), pp 1993–2001
2. Bronstein MM, Bruna J, LeCun Y, Szlam A, Vandergheynst P (2017) Geometric deep learning: Going beyond euclidean data. IEEE Signal Process Mag 34(4):18–42. https://doi.org/10.1109/MSP.2017.2693418
3. Bui KHN, Yi H, Cho J (2021) Uvds: a new dataset for traffic forecasting with spatial-temporal correlation. In: Proceedings of the 13th asian conference on intelligent information and database system (ACIIDS). Springer, pp 66–77
4. Bui KN, Jung JE, Camacho D (2017) Game theoretic approach on real-time decision making for iot-based traffic light control. Concurr Comput Pract Exp 29(11):e4077. https://doi.org/10.1002/cpe.4077
5. Bui KN, Oh H, Yi H (2020) Traffic density classification using sound datasets: an empirical study on traffic flow at asymmetric roads. IEEE Access 8:125,671–125,679. https://doi.org/10.1109/ACCESS.2020.3007917
6. Cao D, Wang Y, Duan J, Zhang C, Zhu X, Huang C, Tong Y, Xu B, Bai J, Tong J, Zhang Q (2020) Spectral temporal graph neural network formultivariate time-series forecasting. In: Proceedings of the 33rd annual conference on neural information processing systems (NeurIPS)
7. Cui Z, Henrickson K, Ke R, Wang Y (2020) Traffic graph convolutional recurrent neural network: A deep learning framework for network-scale traffic learning and forecasting. IEEE Trans Intell Trans Syst 21(11):4883–4894. https://doi.org/10.1109/TITS.2019.2950416
8. Defferrard M, Bresson X, Vandergheynst P (2016) Convolutional neural networks on graphs with fast localized spectral filtering. In: Proceedings of the 29th annual conference on neural information processing systems (NeurIPS), pp 3837–3845

9. Essien A, Giannetti C (2020) A deep learning model for smart manufacturing using convolutional LSTM neural network autoencoders. IEEE Trans Ind Inform 16(9):6069–6078. https://doi.org/10.1109/TII.2020.2967556
10. Feng D, Wu Z, Zhang J, Wu Z (2020) Dynamic global-local spatial-temporal network for traffic speed prediction. IEEE Access 8:209,296–209,307. https://doi.org/10.1109/ACCESS.2020.3038380
11. Guo S, Lin Y, Feng N, Song C, Wan H (2019) Attention based spatial-temporal graph convolutional networks for traffic flow forecasting. In: Proceedings of the 33rd AAAI Conference on Artificial Intelligence (AAAI). AAAI Press, pp 922–929
12. Hamilton WL, Ying Z, Leskovec J (2017) Inductive representation learning on large graphs. In: Proceedings of the 30th annual conference on neural information processing systems (NeurIPS), pp 1024–1034
13. Khodabandelou G, Kheriji W, Hadj-selem F (2021) Link traffic speed forecasting using convolutional attention-based gated recurrent unit. Appl Intell 51(4):2331–2352. https://doi.org/10.1007/s10489-020-02020-8
14. Kipf TN, Welling M (2017) Semi-supervised classification with graph convolutional networks. In: Proceedings of the 5th international conference on learning representation (ICLR). OpenReview.net
15. Li Y, Tarlow D, Brockschmidt M, Zemel RS (2016) Gated graph sequence neural networks. In: Proceedings of the 4th international conference on learning representation (ICLR). OpenReview.net
16. Li Y, Yu R, Shahabi C, Liu Y (2018) Diffusion convolutional recurrent neural network: Data-driven traffic forecasting. In: Proceedings of the 6th international conference on learning representation (ICLR). OpenReview.net
17. Ma X, Dai Z, He Z, Ma J, Wang Y, Wang Y (2017) Learning traffic as images: A deep convolutional neural network for large-scale transportation network speed prediction. Sensors 17(4):818. https://doi.org/10.3390/s17040818
18. Merity S, Xiong C, Bradbury J, Socher R (2017) Pointer sentinel mixture models. In: Proceedings of the 5th international conference on learning representation (ICLR). OpenReview.net. https://openreview.net/forum?id=Byj72udxe
19. Oord AVD, Dieleman S, Zen H, Simonyan K, Vinyals O, Graves A, Kalchbrenner N, Senior AW, Kavukcuoglu K (2016) Wavenet: a generative model for raw audio. In: Proceedings of the 9th ISCA speech synthesis workshop(SSW), vol 125. ISCA
20. Pan Z, Zhang W, Liang Y, Zhang W, Yu Y, Zhang J, Zheng Y (2020) Spatio-temporal meta learning for urban traffic prediction. IEEE Trans Knowl Data Eng. https://doi.org/10.1109/TKDE.2020.2995855. (Early Access)
21. Park C, Lee C, Bahng H, Tae Y, Jin S, Kim K, Ko S, Choo J (2020) ST-GRAT: A novel spatio-temporal graph attention networks for accurately forecasting dynamically changing road speed. In: Proceedings of the 29th ACM international conference on information and knowledge management (CIKM). ACM, pp 1215–1224
22. Shih S, Sun F, Lee H (2019) Temporal pattern attention for multivariate time series forecasting. Mach Learn 108(8-9):1421–1441. https://doi.org/10.1007/s10994-019-05815-0
23. Tanwi M, Prasanna B, Eric R, Jane M (2020) Graph-partitioning-based diffusion convolutional recurrent neural network for large-scale traffic forecasting. Transp Res Rec 2674(9):473–488
24. Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez AN, Kaiser L, Polosukhin I (2017) Attention is all you need. In: Proceedings of the 30th annual conference on neural information processing systems (NeurIPS), pp 5998–6008
25. Velickovic P, Cucurull G, Casanova A, Romero A, Liò P, Bengio Y (2018) Graph attention networks. In: Proceedings of the 6th international conference on learning representation (ICLR). OpenReview.net
26. Wang J, Zhu W, Sun Y, Tian C (2020) An effective dynamic spatiotemporal framework with external features information for traffic prediction. Appl Intell 1–15. https://doi.org/10.1007/s10489-020-02043-1
27. Wu F, Fan A, Baevski A, Dauphin YN, Auli M (2019) Pay less attention with lightweight and dynamic convolutions. In: Proceedings of the 7th international conference on learning representations (ICLR). OpenReview.net. https://openreview.net/forum?id=SkVhlh09tX
28. Wu Z, Pan S, Chen F, Long G, Zhang C, Yu PS (2021) A comprehensive survey on graph neural networks. IEEE Trans Neural Netw Learn Syst 32(1):4–24. https://doi.org/10.1109/TNNLS.2020.2978386
29. Wu Z, Pan S, Long G, Jiang J, Chang X, Zhang C (2020) Connecting the dots: Multivariate time series forecasting with graph neural networks. In: Proceedings of the 26th international conference on knowledge discovery & data mining (KDD). ACM, pp 753–763
30. Wu Z, Pan S, Long G, Jiang J, Zhang C (2019) Graph wavenet for deep spatial-temporal graph modeling. In: Proceedings of the 28th international joint conference on artificial intelligence (IJCAI), pp 1907–1913. ijcai.org
31. Yao H, Tang X, Wei H, Zheng G, Li Z (2019) Revisiting spatial-temporal similarity: A deep learning framework for traffic prediction. In: Proceedings of the 33rd AAAI conference on artificial intelligence (AAAI). AAAI Press, pp 5668–5675
32. Yi H, Bui KHN (2020) An automated hyperparameter search-based deep learning model for highway traffic prediction. IEEE Trans Intell Transp Syst. https://doi.org/10.1109/TITS.2020.2987614. (Early Access)
33. Yin X, Wu G, Wei J, Shen Y, Qi H, Yin B (2021) Deep learning on traffic prediction: Methods, analysis and future directions. IEEE Trans Intell Transp Syst. https://doi.org/10.1109/TITS.2021.3054840. (Early Access)
34. Yu B, Yin H, Zhu Z (2018) Spatio-temporal graph convolutional networks: A deep learning framework for traffic forecasting. In: Proceedings of the 27th international joint conference on artificial intelligence (IJCAI), pp 3634–3640. ijcai.org
35. Zhang C, Yu JJQ, Li Y (2019) Spatial-temporal graph attention networks: A deep learning approach for traffic forecasting. IEEE Access 7:166,246–166,256. https://doi.org/10.1109/ACCESS.2019.2953888
36. Zhang S, Tong H, Xu J, Maciejewski R (2019) Graph convolutional networks: A comprehensive review. Comput Soc Netw 6(11):1–23. https://doi.org/10.1186/s40649-019-0069-y
37. Zhao L, Zhou Y, Lu H, Fujita H (2019) Parallel computing method of deep belief networks and its application to traffic flow prediction. Knowl Based Syst 163:972–987. https://doi.org/10.1016/j.knosys.2018.10.025
38. Zheng C, Fan X, Wang C, Qi J (2020) GMAN: A graph multi-attention network for traffic prediction. In: Proceedings of the 34th AAAI conference on artificial intelligence (AAAI). AAAI Press, pp 1234–1241
39. Zhou J, Cui G, Zhang Z, Yang C, Liu Z, Sun M (2020) Graph neural networks: A review of methods and applications. AI Open 1:57–81. https://doi.org/10.1016/j.aiopen.2021.01.001

Khac-Hoai Nam Bui received the M.S. degree in computer science and information engineering from Aletheia University, New Taipei, Taiwan, in 2014, and the Ph.D. degree in computer science and engineering from Chung-Ang University, Seoul, South Korea, in 2018. He has been a researcher at Viettel Cyberspace Center (VTCC), Hanoi, Vietnam, since March 2021. Before joining VTCC, he was a Researcher with Korea Institute of Science and Technology Information (KISTI), Daejeon, South Korea, in February 2019. His research interests are the Internet of Things and ambient intelligence by using AI methodologies such as data mining, machine learning, and logical reasoning. Recently, he focuses on applying deep learning models for smart applications, such as transportation and home energy management systems.

Jiho Cho is a Postdoctoral Researcher at the Korea Institute of Science and Technology Information (KISTI) since 2019. He received the Ph.D. degree in computer engineering at the Hannam University, Daejeon, Korea in 2019. His research interests are intelligent transportation systems based on transportation big data and artificial intelligence, machine learning, network security, and online privacy.

Hongsuk Yi received the Ph.D. degree from Sogang University, Seoul, South Korea, in 1997. He has been a Principal Researcher with the Korea Institute of Science and Technology Information (KISTI), South Korea, since 2000. His research interests are traffic congestion problems, smart cities, supercomputing, and heterogeneous computing by using deep learning techniques. Recently, his research topic focuses on developing traffic signal control solutions by using deep learning methodologies.