Adaptive and Interactive Multi-Level Spatio-Temporal Network for Traffic Forecasting

Yudong Zhang, *Graduate Student Member, IEEE*, Pengkun Wang, *Member, IEEE*, Binwu Wang, Xu Wang, Zhe Zhao, Zhengyang Zhou, *Member, IEEE*, Lei Bai, *Member, IEEE*, and Yang Wang, *Senior Member, IEEE*

***Abstract*— Traffic forecasting is a challenging research topic due to the complex spatial and temporal dependencies among different roads. Though great efforts have been made on traffic forecasting, existing works still have the following shortcomings: i) Most methods only directly perform on the original road network topology which cannot accommodate the diverse traffic patterns and multi-granularity traffic forecasting requirements driven by the natural multi-level urban structure and layout, ii) The existing studies based on the spatio-temporal multi-granularity perspective ignore the interactions between the fine-grained information and coarse-grained information, resulting in the spatio-temporal correlation under multi-granularity inaccurately modeled. To solve the problems, we propose an Adaptive and Interactive Multi-level Spatio-Temporal network (AIMST) for traffic forecasting. Specifically, we first devise a learnable adaptive hierarchical clustering method to automatically generate more coarse-grained graphs from the initial road networks and the traffic data. Then, the spatio-temporal graph convolutional networks are executed on the constructed hierarchical traffic graph of each level correspondingly to capture the spatio-temporal patterns. Furthermore, a multi-level bidirectional interaction module is designed to emphasize the multi-grained interaction patterns among different levels. Extensive experiments on two real-world traffic datasets demonstrate that our framework is superior to several state-of-the-art baselines.**

***Index Terms*— Spatio-temporal data mining, traffic forecasting, multi-level traffic network, urban computing.**

Manuscript received 20 July 2023; revised 26 December 2023 and 16 April 2024; accepted 19 April 2024. This work was supported in part by the National Natural Science Foundation of China under Grant 12227901 and Grant 62072427, in part by the Project of Stable Support for Youth Team in Basic Research Field, in part by Chinese Academy of Sciences (CAS) under Grant YSBR-005, in part by the Academic Leaders Cultivation Program, and in part by the University of Science and Technology of China (USTC). The Associate Editor for this article was S. C. Wong. *(Corresponding authors: Yang Wang; Lei Bai.)*

Yudong Zhang, Binwu Wang, and Zhe Zhao are with the School of Data Science, University of Science and Technology of China, Hefei 230026, China (e-mail: zyd2020@mail.ustc.edu.cn; wbw1995@mail.ustc.edu.cn; zz4543@mail.ustc.edu.cn).

Pengkun Wang, Xu Wang, and Zhengyang Zhou are with the School of Software Engineering, University of Science and Technology of China, Hefei 230026, China, and also with the Suzhou Institute for Advanced Research, University of Science and Technology of China, Suzhou 215123, China (e-mail: pengkun@ustc.edu.cn; wx309@ustc.edu.cn; zzy0929@ustc.edu.cn).

Lei Bai is with Shanghai AI Laboratory, Shanghai 200030, China (e-mail: baisanshi@gmail.com).

Yang Wang is with the Key Laboratory of Precision and Intelligent Chemistry and the School of Data Science, University of Science and Technology of China, Hefei 230026, China, and also with the Suzhou Institute for Advanced Research, University of Science and Technology of China, Suzhou 215123, China (e-mail: angyan@ustc.edu.cn).

Digital Object Identifier 10.1109/TITS.2024.3392975

I. Introduction

With accelerated urbanization, a large number of people are rapidly gathering in cities. Traffic congestion, frequent traffic accidents, long commute times, and other transportation problems have severely reduced the operational efficiency of cities and reduced the travel experience of commuters. To address these challenges, many cities are committed to developing intelligent transportation systems (ITS) to provide efficient traffic management, accurate transportation resource allocation, and high-quality transportation services [1], [2], [3]. Modern urban systems are equipped with an array of sensors spread across traffic networks and key areas to monitor dynamic traffic conditions such as flow and speed [4]. Traffic forecasting is one of the most fundamental tasks in ITS [5], [6], which aims to predict future traffic conditions (e.g., traffic flow or speed) in the road network based on historical observations (e.g., collected by sensor records). Accurate traffic forecasting can not only help transportation administrators make better traffic control policies for reducing congestion [7], [8], [9] and facilitate commuters estimating travel time to avoid delays but also provides insights for urban planning and improving public safety [10], [11], [12].

However, traffic forecasting is challenging and has undergone different stages of evolution. In the early stage, traditional models are mainly based on mathematical statistics to predict traffic states. Among them, the principle of the Historical Average (HA) model uses the historical average of data as the prediction result, which is simple to calculate but has low prediction accuracy [13]. Time series models such as Auto Regressive Moving Average (ARMA) [14] and its variants [15], [16] further utilize the temporal dependencies between current data and historical data for forecasting and perform modeling and analysis considering the periodicity and trend of data. However, they are based on the time series stability assumption and thus unable to capture traffic flow mutations. Traditional machine learning approaches such as Support Vector Machine (SVM) [17], Bayesian inference method [18], and K-Nearest Neighbors (KNN) [19] can model the non-linearity and extract some complex correlations in traffic data. However, their shallow architecture leads to limited capacity for capturing more complex patterns and makes them inferior in big data scenarios [20], [21].

In recent years, deep learning-based techniques have been widely employed and achieved state-of-the-art performance in

various traffic applications. The Recurrent Neural Networks (RNN) [7] or its variants [22], [23] are employed to extract the temporal correlations in traffic data. And the Convolutional Neural Networks (CNN) are used to capture the spatial correlations in grid-based traffic network [24], [25]. However, many traffic networks are graph-structured in nature, such as road network [26], [27] and subway network [28], to which CNNs are not applicable for spatial features representation. To overcome this problem, Graph Convolutional Networks (GCN) are introduced to traffic forecasting and have achieved great success [29], [30], [31], [32], [33], [34]. Specifically, STGCN [35] applies ChebNet graph convolution and 1D convolution to extract spatial dependencies and temporal correlations in traffic data. ASTGCN [29] uses attention-based spatial-temporal graph convolutions to model dynamic spatial-temporal features of traffic flows. Graph WaveNet [36] captures the spatial correlation with a diffusion convolution layer and learns the temporal correlation using generic TCN. AGCRN [34] introduces node adaptive parameter learning to automatically capture node-specific spatial and temporal correlations in time-series data without a predefined graph. STFGNN [37] utilizes hidden states from previous multiple time steps to handle long sequences and achieves state-of-the-art performance in traffic flow forecasting.

While the detailed network structures or graph definitions of these advanced GCN models are diverse for achieving more efficient and accurate spatial correlations modeling, they all capture the spatial correlations in one granularity (e.g., the original road network granularity), which does not have enough capacity to model the complicated structure, layout, functions, and proximity of modern cities. In real-world urban systems, a city is normally planned by the administration or gradually developed into multiple smaller zones that have distinct main functions, such as commercial, residences, education, factory, and so on. Each zone will also contain many different blocks or streets that have their own characters, e.g., the main street or a small lane. Such complexities from the city layout, structure, functions, and proximity finally result in a multi-level hierarchical system. For example, Figure 1 illustrates the number of pick-ups and drop-offs of urban taxicabs during the morning traffic rush hours and afternoon rush hours in Manhattan. We can observe that the pick-ups in the morning peak and the drop-offs in the evening peak are relatively concentrated in some specific regions (containing several road segments) that may relate to the residence. Similarly, the morning peak drop-offs and evening peak pick-ups are also relatively concentrated in the vicinity of work regions. From a more macro perspective, several residences or work regions further form a larger functional zone, corresponding to the urban functional layout, e.g., residential areas, CBD.

Such a multi-level hierarchical structure of cities will have an inneglectable influence on traffic patterns from two perspectives. First, the traffic patterns of a basic road segment could have correlations and interact with other road segments at different levels. For example, two streets will have more similar traffic patterns than a street and a lane. At the same time, two streets of different resident regions also have a high probability to have more similar traffic patterns (e.g., high morning pick-ups) than a street from residence and a street from working regions (e.g., high morning drop-offs). Thus, modeling the traffic data from multiple hierarchical levels could help us discover more accurate and comprehensive spatio-temporal correlations. Second, the traffic patterns of basic units of each level are influenced and correlated with the units in other levels. For example, the congestion of a certain road segment will quickly spread to the surrounding area and influence the traffic of a region. Thus, capturing the bidirectional interaction patterns between coarse and fine granularity urban levels is essential for multi-level traffic prediction models to learn multi-grain spatio-temporal features.

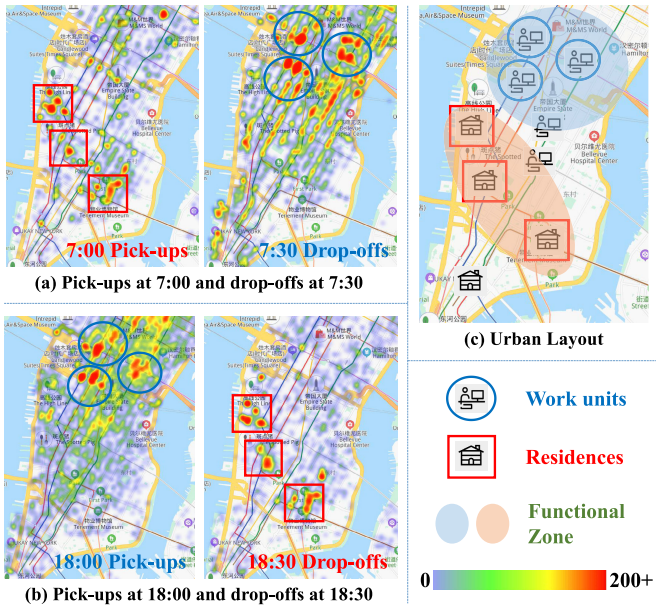


Fig. 1. Multi-level structure of natural urban layout.

Motivated by these observations, we propose a novel **A**daptive and **I**nteractive **M**ulti-level **S**patio-**T**emporal network (**AIMST**) for traffic forecasting. Specifically, to construct the hierarchical traffic graph satisfying the multi-level structure of the urban natural layout, we introduce an adaptive hierarchical clustering method, which can automatically generate the coarse-grained traffic graph according to the initially fine-grained graph structure. Our adaptive hierarchical clustering method can be trained with the forecasting together in an end-to-end manner, which is easier to use than the offline clustering step employed in [38] and can reveal more accurate hierarchies through learning. In addition, to explicitly consider the non-negligible spatio-temporal interaction between the fine-grained information and coarse-grained perspective, we propose a bidirectional multi-level interaction module based on the attention mechanism that mutually guides the feature interactions among different levels. The main **contributions** of this paper are summarized as follows:

- We propose a multi-level spatio-temporal network driven by the hierarchical urban layout for traffic forecasting, which can effectively simulate the complex urban structure and model the evolutionary relationship between traffic patterns and urban space.

TABLE I
KEY NOTATIONS AND CORRESPONDING DESCRIPTIONS

Notations	Descriptions
G_S, G_R, G_Z	Segment-, region- and zone-level graph
$\mathbf{X}^S$, $\mathbf{X}^R$, $\mathbf{X}^Z$	Feature matrices of segment-, region- and zone-level graph
$\mathbf{A}^S$, $\mathbf{A}^R$, $\mathbf{A}^Z$	Adjacency matrices of segment-, region- and zone-level graph
N^S, N^R, N^Z	Numbers of road segments, regions, zones
$\mathbf{S}_{SR}$, $\mathbf{S}_{RZ}$	Assignment matrices of segment-to-region and region-to-zone
$\mathbf{F}^S$, $\mathbf{F}^R$, $\mathbf{F}^Z$	Learned spatio-temporal features of segment-, region- and zone-level
$\mathbf{F}^S_{int}$, $\mathbf{F}^R_{int}$, $\mathbf{F}^Z_{int}$	Learned spatio-temporal features of segment-, region- and zone-level after interaction
P, Q	Time steps of historical sequence, prediction sequence
L_1, L_2	Regularization terms of assignment optimization
$\mathcal{L}_S$, $\mathcal{L}_R$, $\mathcal{L}_Z$	MAE of segment-, region- and zone-level predictions
Θ	All the learnable parameters in the model
λ_1, λ_2, λ_3	Regularization weights for balancing loss

- An adaptive spatio-temporal hierarchical clustering module is introduced to automatically achieve reasonable multi-level division of urban areas based on the original road networks and traffic data, which can be learned in an end-to-end manner.
- A multi-level bidirectional interaction module is devised to emphasize the bidirectional effects between coarse and fine information in spatiotemporal learning, which can couple the multi-grained spatio-temporal patterns to facilitate the downstream tasks.
- Extensive experiments are conducted on two real-world datasets and the results show that our model consistently outperforms all the baseline methods across various scenarios.

The remainder of this paper is organized as follows. Section II introduces the preliminaries and formalizes the problem. Section III investigates the proposed model in detail, followed by our empirical studies in Section IV. Moreover, Section V discusses future research directions. Finally, Section VI briefly reviews the related work, and Section VII concludes this paper.

II. PRELIMINARIES

In this section, we formally define some basic concepts as well as the key problem studied in this paper. Table I presents the frequently used notations and corresponding descriptions throughout this paper.

Definition 1 (Traffic forecasting): Given the historical traffic data (e.g., traffic flow, traffic speed) collected from N correlated traffic sensors located on the road network, the task of traffic forecasting is to forecast the future traffic status of the road network. Following previous studies [39], [40], [41], we denote a road network as a weighted graph $G = (V, E, \mathbf{A})$, where V is the set of $|V| = N$ nodes, E is the set of edges connecting the nodes, and $\mathbf{A} \in \mathbb{R}^{N\times N}$ is a weighted adjacency matrix representing the proximities among nodes, e.g., measured by the distance of node pairs in the road network. The traffic data observed on G at time step t are denoted as a graph signal $\mathbf{X}_t \in \mathbb{R}^{T\times D}$, where D is the feature dimension. The traffic forecasting problem aims to learn a function f that is able to forecast Q future graph signals given P historical graph signals and the graph G:

$$(\mathbf{X}_1, \cdots, \mathbf{X}_t, \cdots, \mathbf{X}_P; G) \xrightarrow{f} (\mathbf{X}_{P+1}, \cdots, \mathbf{X}_{P+Q})$$

where $\mathbf{X}_{1:P} \in \mathbb{R}^{N\times D\times P}$ and $\mathbf{X}_{(P+1):(P+Q)} \in \mathbb{R}^{N\times D\times Q}$.

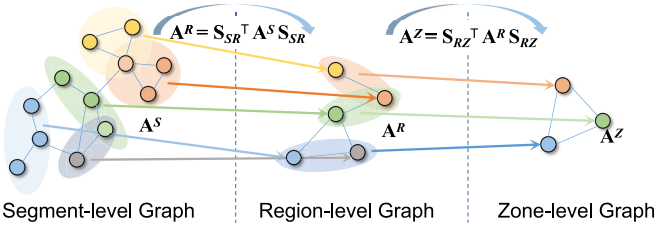


Fig. 2. Multi-level traffic graphs generating with adaptive spatio-temporal hierarchical clustering. The potential hierarchical relationships of traffic patterns can be automatically exploited through end-to-end learning.

Definition 2 (Multi-level traffic network): While the whole city could potentially have multiple-level hierarchies, we divide the traffic network into three levels of hierarchy motivated by the real hierarchical characters of urban layout, i.e., Segment-level, Region-level, and Zone-level in this paper. Based on Def. 1, the traffic network graph of each level can be defined as follows:

(i) Segment-level graph: the finest-grained traffic graph structure consisting of urban road segments, which is consistent with the graph $\mathbf{A} \in \mathbb{R}^{N\times N}$ used in most existing traffic forecasting works.

(ii) Region-level graph: a traffic graph structure composed of clusters gathered by several segment-level nodes.

(iii) Zone-level graph: the most macroscopic traffic graph structure composed of several clusters aggregated from nodes at the region level.

The multi-level graphs can be formally described as $G_S = (V^S, E^S, \mathbf{A}^S)$, $G_R = (V^R, E^R, \mathbf{A}^R)$ and $G_Z = (V^Z, E^Z, \mathbf{A}^Z)$, respectively, where $\mathbf{A}^S \in \mathbb{R}^{N^S\times N^S}$, $\mathbf{A}^R \in \mathbb{R}^{N^R\times N^R}$, and $\mathbf{A}^Z \in \mathbb{R}^{N^Z\times N^Z}$. And the graph signals of each level are $\mathbf{X}^S \in \mathbb{R}^{N^S\times D\times P}$, $\mathbf{X}^R \in \mathbb{R}^{N^R\times D\times P}$, and $\mathbf{X}^Z \in \mathbb{R}^{N^Z\times D\times P}$.

III. METHODOLOGY

In this section, we present the proposed Adaptive and Interactive Multi-level Spatio-Temporal network (AIMST) for traffic forecasting. The core idea is to establish a correspondence between traffic patterns and the hierarchy of urban layout by constructing a multi-level traffic prediction model. The overall architecture for the proposed AIMST is presented in Figure 3. We start with generating the hierarchical traffic graphs, then present how to extract temporal and spatial dependencies from historical traffic data and how to interact spatio-temporal features captured at each level. Finally, we discuss how to predict future traffic states and optimize the model.

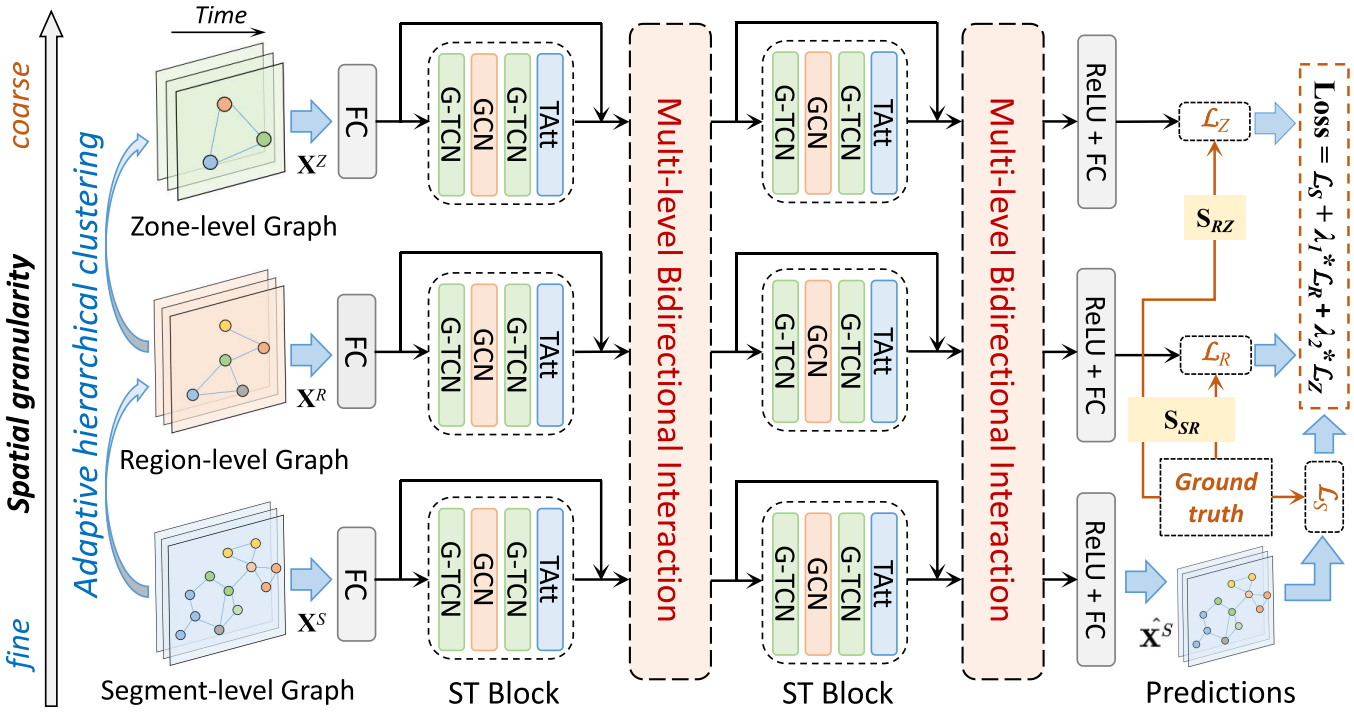


Fig. 3. Overview of AIMST: Adaptive and interactive multi-level spatio-temporal network for traffic forecasting.

A. Multi-Level Traffic Graph Generation

In this subsection, we first construct the segment-level traffic graph based on the original urban road network data, then learn the region-level and zone-level traffic graphs in an end-to-end manner by introducing the adaptive hierarchical clustering method which utilizes the learnable cross-granularity assignment matrix as shown in Figure 2. Besides, the adaptive clustering process is constrained by two regularization terms to facilitate the learning process.

1) Segment-Level Traffic Graph Construction: The segment-level traffic graph $G_S = (V^S, E^S, \mathbf{A}^S)$ is modeled on the original urban road network, which considers a traffic network with N^S sensors or road segments. Following STGCN [35], we compute the edge weight $\mathbf{A}^S_{i,j}$ between vertex v_i and v_j $(v_i, v_j \in V^S)$ via a threshold Gaussian kernel weighting function:

$$\mathbf{A}^S_{i,j} = \begin{cases} \exp\left(-\dfrac{d(i,j)^2}{\sigma^2}\right), & \text{if } d(i,j) \le \kappa \\ 0, & \text{otherwise} \end{cases} \tag{1}$$

where $d(i, j)$ is the Euclidean distance between vertex v_i and v_j, σ is the deviation of the distances, and κ is the threshold parameter to control the neighborhood size. The edges with zero weight can be considered disconnected.

As introduced in Def. 1, we denote the observed traffic data of the segment-level graph by $\mathbf{X}^S = \mathbf{X}^S_{1:P} \in \mathbb{R}^{N^S \times D \times P}$, which will be used for the prediction tasks.

2) Region- and Zone-Level Graph Construction: Different from HGCN [38] that generates coarse-level graphs with offline clustering, we attempt to learn the coarse-level graphs from data automatically, which makes the workflow easier and can reveal more accurate hierarchical structures. Specifically, the coarse-grained traffic graphs are adaptively generated from fine-grained road network representations by mining the spatial and semantic contextual information of road segments. Motivated by [42], we utilize a GNN layer to learn a cross-granularity assignment matrix for aggregating the nodes from the relatively fine-grained level and generating the coarser level graph.

Take the region-level graph generation as an example, the segment-to-region assignment matrix is learned based on the adjacency matrix $\mathbf{A}^S$ and input feature matrices $\mathbf{X}^S$ of the segment-level:

$$\mathbf{S}_{SR} = \text{Softmax}\left(\text{GCN}\left(\mathbf{A}^S, \mathbf{X}^S\mathbf{W}_1\right) \cdot \mathbf{W}_2\right) \tag{2}$$

where $\mathbf{S}_{SR} \in \mathbb{R}^{N^S \times N^R}$ is the segment-to-region assignment matrix, $Softmax$ is the softmax function and applied in a row-wise fashion, $\mathbf{W}_1 \in \mathbb{R}^P$, $\mathbf{W}_2 \in \mathbb{R}^{D \times N^R}$ are the learnable parameters. N^R corresponds to a pre-defined number of the region-level nodes and is a hyperparameter.

In fact, the functional region layout of a city is relatively stable and does not change in the short term (e.g., months or years). Whereas our work studies the traffic prediction problem of a city in the next few hours, during which the functional layout of the city does not change, so N^R is considered as a fixed hyperparameter in this paper. Meanwhile, N^R can be regarded as an a priori guide for the learning of adaptive clustering. When the value of hyperparameter N^R is reasonable, the obtained clusters are more reasonable and the division of regions is closer to the real urban layout, which brings superior performance to downstream prediction.

Notice that, the segment-to-region assignment matrix $\mathbf{S}_{SR}$ is a probabilistic assignment matrix essentially, and $\mathbf{S}_{SR}(i, j)$

represents the probability that segment-level node i is aggregated to region-level node j. Based on the segment-to-region assignment matrix, we can directly obtain the adjacency matrix of the region-level graph:

$$\mathbf{A}^R = \mathbf{S}_{SR}^\top \mathbf{A}^S \mathbf{S}_{SR} \in \mathbb{R}^{N^R \times N^R} \tag{3}$$

In addition, we can also pool the traffic data from the original segment level to the regional level based on the segment-to-region assignment matrix. Thus, the feature matrix of the region-level graph can be formulated as:

$$\mathbf{X}^R = \mathbf{S}_{SR}^\top \mathbf{X}^S \in \mathbb{R}^{N^R \times D \times P} \tag{4}$$

It is worth noting that, since the cross-granularity assignment matrix is a learnable probability distribution, the adaptive hierarchical clustering method we introduce is a soft clustering, whose benefits can be summarized into two aspects. On the one hand, some road segments in a city may be located between two functional regions, so its traffic state should be affected by both, and it is more consistent with the real-world layout of the city to cluster such segment-level nodes into the corresponding region-level nodes at the same time. On the other hand, when pooling the fine-grained graph using a cross-granularity assignment matrix, different aggregation scores are assigned to different nodes, which also well reflects the difference of importance coefficients of roads in cities, such as the influence of major and minor roads on regional traffic state is different.

Similarly, we can further construct the zone-level graph on the basis of the region-level graph. The specific process is as follows:

$$\begin{aligned} \mathbf{S}_{RZ} &= \text{Softmax}\left(\text{GCN}\left(\mathbf{A}^R, \mathbf{X}^R \mathbf{W}_3\right) \cdot \mathbf{W}_4\right) \\ \mathbf{A}^Z &= \mathbf{S}_{RZ}^\top \mathbf{A}^R \mathbf{S}_{RZ} \\ \mathbf{X}^Z &= \mathbf{S}_{RZ}^\top \mathbf{X}^R \end{aligned} \tag{5}$$

where $\mathbf{S}_{RZ} \in \mathbb{R}^{N^R \times N^Z}$, $\mathbf{A}^Z \in \mathbb{R}^{N^Z \times N^Z}$ and $\mathbf{X}^Z \in \mathbb{R}^{N^Z \times D \times P}$. $\mathbf{W}_3 \in \mathbb{R}^P$, $\mathbf{W}_4 \in \mathbb{R}^{D \times N^Z}$ are the learnable parameters. N^Z corresponds to a pre-defined number of the zone-level nodes and is also a hyperparameter. The detailed studies of hyperparameters N^R and N^Z are presented in Section IV-F.1.

3) Assignment Regularization: Although the construction of region-level and zone-level graphs can be optimized together with the forecasting loss directly, it is difficult to train the two assignment matrices because this is a non-convex optimization problem and tends to fall into local optimum [42]. Therefore, we employ two regularization terms to alleviate the optimization issue.

Specifically, the first regularization term is defined as:

$$L_1 = \left\| \mathbf{A}^S, \mathbf{S}_{SR} \mathbf{S}_{SR}^\top \right\|_F + \left\| \mathbf{A}^R, \mathbf{S}_{RZ} \mathbf{S}_{RZ}^\top \right\|_F \tag{6}$$

where $\|\cdot\|_F$ denotes the Frobenius norm. By viewing the elements in adjacency matrix $\mathbf{A}^S$ and $\mathbf{A}^R$ represent the connection strength between two nodes, and elements in $\mathbf{S}_{SR}\mathbf{S}_{SR}^\top$ and $\mathbf{S}_{RZ}\mathbf{S}_{RZ}^\top$ represent the probabilities that two nodes are divided to the same cluster, this regularization term encourages that node pairs with larger connection strength are mapped to the same cluster.

Besides, the second regularization term is defined to clearly define the affiliation of each cluster. In real-world road networks, most road segments are contained in only one of their corresponding functional regions, and road segments located at the junction of two or more functional regions are in the minority. Therefore, we make each row in the assignment matrix $\mathbf{S}_{SR}$ and $\mathbf{S}_{RZ}$ approach to a one-hot vector by regularizing the entropy as follows,

$$L_2 = \frac{1}{N^S} \sum_{i=1}^{N^S} H\left(\mathbf{S}_{SR_i}\right) + \frac{1}{N^R} \sum_{j=1}^{N^R} H\left(\mathbf{S}_{RZ_j}\right) \tag{7}$$

where $H(\cdot)$ is the entropy function that can reduce the uncertainty of the mapping distribution, $\mathbf{S}_{SR_i}$ and $\mathbf{S}_{RZ_j}$ are the i-th and j-th row of the cross-granularity assignment matrix $\mathbf{S}_{SR}$ and $\mathbf{S}_{RZ}$, respectively. One special situation is that the i-th node is only mapped to one cluster in the upper level, and the entropy of the corresponding row is 0 at this time.

B. Spatio-Temporal Dependency Learning

After obtaining the traffic graphs at different levels, the complex and hierarchical spatial and temporal dependencies can be captured by appropriate learning networks. As presented in Figure 3, the extraction process of spatio-temporal characteristics of each level of traffic graphs consists of two spatio-temporal blocks (i.e., ST Block). Following STGCN [35] and HGCN [38], each ST Block is composed of a Gated Temporal Convolution Network (G-TCN) layer for modeling the local temporal correlations, a Graph Convolutional Network (GCN) layer for learning the spatial correlations, another G-TCN layer for extracting both temporal and spatial dependencies simultaneously, and a Temporal Attention Mechanism (TAtt) module for capturing the dynamic global temporal patterns of the traffic data. Following STGCN [35], the role of the second G-TCN in ST block is to recover the compressed spatio-temporal patterns in the graph convolution into a sequence form with the original feature dimension, which realizes the effective fusion of temporal and spatial information coherently. The details are presented in Figure 4 and described in the following.

1) Gated Temporal Convolution Layer: Although RNN-based models, such as LSTM and GRU, are widely applied in time-series analysis, recurrent networks still suffer from some intrinsic drawbacks like time-consuming iterations, unstable gradients, and delayed responses to dynamic changes. To enhance the performance of extracting long-term temporal dependencies, a dilated temporal convolutional network [43] that enables an exponentially large receptive field [44] along the time axis is adopted here.

$$\text{TCN}(\mathbf{X}) = \text{Conv}_{t_s}^{dil}(\mathbf{X}) \tag{8}$$

where $\text{Conv}_{t_s}^{dil}$ represents the temporal Convolution operator along the time dimension with a kernel size of t_s, dil is the exponential dilation rate in temporal convolution.

To better control the information flow and only keep the useful local temporal patterns for forecasting, the gating mechanism is further introduced owing to its superior performance

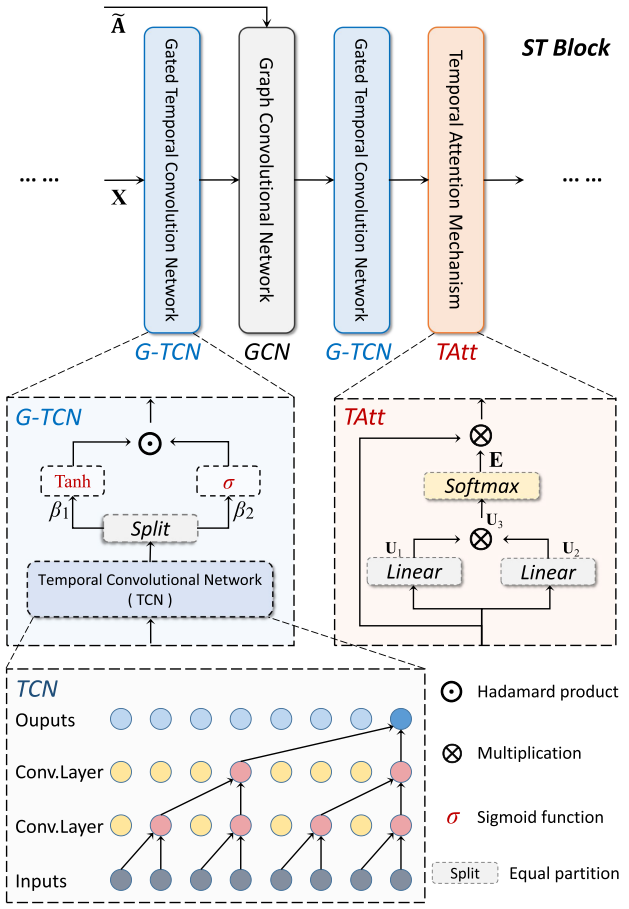


Fig. 4. Spatio-temporal block.

shown in recurrent networks and other graph convolutional networks [36]. The Gated Temporal Convolution Network (G-TCN) is defined as:

$$(\beta_1, \beta_2) = \text{split}(\text{TCN}(\mathbf{X}))$$
$$\text{G-TCN}(\mathbf{X}) = \text{Tanh}(\beta_1) \odot \sigma(\beta_2) \tag{9}$$

where split represents the operator of the equal partition along the channel dimension, $\sigma(\cdot)$ is the Sigmoid activation function, and $\odot$ corresponds to the Hadamard product.

2) Graph Convolutional Layer: In addition to paying attention to the temporal correlation of traffic states, capturing spatial correlation is also an important step in mining traffic patterns, where Graph Convolution Network (GCN) [45] is widely used. The same as existing GCN-based traffic forecasting works, the graph convolution layer is formulated as:

$$\text{GCN}(\mathbf{X}) = \sigma\left(\tilde{\mathbf{A}}\mathbf{X}\mathbf{W}\right) \tag{10}$$

where $\tilde{\mathbf{A}} \in \mathbb{R}^{N\times N}$ denotes the normalized adjacency matrix [34], [36], $\mathbf{X}$ is the input of the GCN layer, and $\mathbf{W}$ is the learnable parameters.

3) Temporal Attention Mechanism: After the temporal and spatial correlations modeling, we draw on the structure in HGCN [38], and further add an attention mechanism module proposed by ASTGCN [29] at the end of the ST Block to capture the dynamic global temporal patterns.

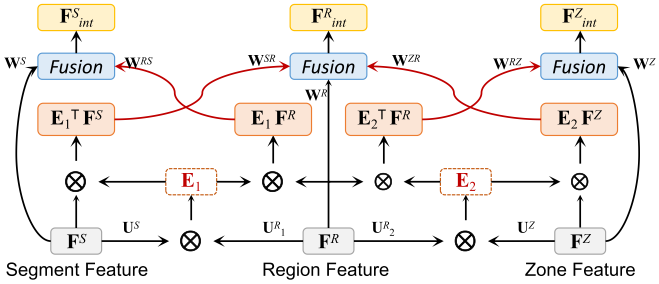


Fig. 5. Multi-level bidirectional interaction module, which can quantify the bidirectional effects between coarse and fine information and couple the multi-grained spatio-temporal patterns to facilitate the downstream tasks.

$$\mathbf{E} = \mathbf{V}_e \cdot \sigma\left(\left(\mathbf{X}^\top \mathbf{U}_1\right)\mathbf{U}_2\left(\mathbf{U}_3\mathbf{X}\right) + \mathbf{b}_e\right)$$
$$\mathbf{E}'_{i,j} = \frac{\exp\left(\mathbf{E}_{i,j}\right)}{\sum_{j=1}^{T_2}\exp\left(\mathbf{E}_{i,j}\right)}$$
$$\text{TAtt}(\mathbf{X}) = \mathbf{X}\mathbf{E}' \tag{11}$$

where $\mathbf{V}_e$, $\mathbf{b}_e \in \mathbb{R}^{P'\times P'}$ (P' denotes the time length of the feature $\mathbf{X}$), and $\mathbf{U}_1 \in \mathbb{R}^N$, $\mathbf{U}_2 \in \mathbb{R}^{D'\times N}$, $\mathbf{U}_3 \in \mathbb{R}^{D'}$ are all learnable parameters. The value of an element $\mathbf{E}_{i,j}$ in $\mathbf{E}$ semantically indicates the strength of dependencies between time i and j. At last, $\mathbf{E}$ is normalized by the softmax function and used to re-weight different time steps of $\mathbf{X}$.

C. Multi-Level Bidirectional Interaction Module

The traffic states of each level are not isolated but are intently related. For a particular road segment, its traffic state is always affected by the overall traffic pattern of the region where it is located. At the same time, the traffic state of a road segment could constantly propagate to the surrounding segments, which constitutes the factor of regional traffic pattern transformation. Motivated by this discovery, we propose to emphasize the bidirectional effects between coarse and fine information in spatiotemporal learning.

Therefore, as shown in Figure 5, we interact the output spatio-temporal features of each level (i.e., $\mathbf{F}^S$ for road segment level, $\mathbf{F}^R$ for region level, and $\mathbf{F}^Z$ for zone level) with the traffic graphs of adjacent levels after each ST Block execution to extract spatio-temporal information, and then use the features $\mathbf{F}^S_{int} \in \mathbb{R}^{N^S\times D'\times P'}$, $\mathbf{F}^R_{int} \in \mathbb{R}^{N^R\times D'\times P'}$, $\mathbf{F}^Z_{int} \in \mathbb{R}^{N^Z\times D'\times P'}$ obtained after the interaction for the following ST Blocks or prediction.

Specifically, we first compute two attention matrices $\mathbf{E}_1$ and $\mathbf{E}_2$, which are used to exploit potential interaction patterns between the features of adjacent levels, respectively. Formally:

$$\mathbf{E}_1 = \sigma\left(\left(\mathbf{F}^S\mathbf{U}^S\right)\left(\mathbf{F}^R\mathbf{U}^R_1\right)^\top\right)$$
$$\mathbf{E}_2 = \sigma\left(\left(\mathbf{F}^R\mathbf{U}^R_2\right)\left(\mathbf{F}^Z\mathbf{U}^Z\right)^\top\right) \tag{12}$$

where $\mathbf{E}_1 \in \mathbb{R}^{N^S\times N^R}$, $\mathbf{E}_2 \in \mathbb{R}^{N^R\times N^Z}$, and $\mathbf{U}^S, \mathbf{U}^R_1, \mathbf{U}^R_2, \mathbf{U}^Z \in \mathbb{R}^{P'}$ are all learnable parameters.

Then, we quantify the spatio-temporal interaction information of adjacent levels using the attention matrices, and finally fuse them with several weight matrices to obtain the new features of multi-level pattern awareness. Formally:

$$\begin{aligned}\mathbf{F}_{int}^{S} &= \mathbf{W}^{S} \odot \mathbf{F}^{S} + \mathbf{W}^{RS} \odot \left(\mathbf{E}_1 \mathbf{F}^{R}\right) \\ \mathbf{F}_{int}^{R} &= \mathbf{W}^{R} \odot \mathbf{F}^{R} + \mathbf{W}^{SR} \odot \left(\mathbf{E}_1^{\top} \mathbf{F}^{S}\right) \\ &\quad + \mathbf{W}^{ZR} \odot \left(\mathbf{E}_2 \mathbf{F}^{Z}\right) \\ \mathbf{F}_{int}^{Z} &= \mathbf{W}^{Z} \odot \mathbf{F}^{Z} + \mathbf{W}^{RZ} \odot \left(\mathbf{E}_2^{\top} \mathbf{F}^{R}\right) \end{aligned} \tag{13}$$

where the weight matrices are all learnable, which reflects the influence degrees of spatio-temporal characteristics of each level on the corresponding adjacent levels. Similar feature interactions are carried out after the execution of each ST Block.

Algorithm 1 Learning Process of AIMST Framework

Require:
Observed segment-level traffic data $\mathbf{X}^S \in \mathbb{R}^{N^S \times D \times P}$; predefined segment-level graph $G_S = (V^S, E^S, \mathbf{A}^S)$; maximum epoch number E; regularization weights $\lambda_1, \lambda_2, \lambda_3$; learning rate η.

Ensure:
Trained parameters in Θ.

Initialize all parameters in Θ.
for $e = 1$ to E **do**
  Generate region- and zone-level graph G_R, G_Z, and the corresponding feature matrices $\mathbf{X}^R, \mathbf{X}^Z$;
  repeat
    Extract spatio-temporal dependencies $\mathbf{F}^S, \mathbf{F}^R, \mathbf{F}^Z$ by ST Blocks;
    Interact the learned spatio-temporal features of each adjacent level to obtain $\mathbf{F}_{int}^S, \mathbf{F}_{int}^R, \mathbf{F}_{int}^Z$;
  until all of ST Blocks have been executed.
  Make predictions $\hat{\mathbf{X}}$ for each level with stacked fully connected layers and Relu activation function;
  Calculate the MAE loss for each level prediction $\mathcal{L}_S, \mathcal{L}_R, \mathcal{L}_Z$ according to Eq. (14);
  Combine the loss terms together to get $\mathcal{L}$.
  for $\theta \in \Theta$ **do**
    $\theta = \theta - \eta \cdot \partial \mathcal{L} / \partial \theta$.
  end for
end for
return all parameters in Θ.

D. Model Optimization

Finally, a fully connected layer is appended to make sure the output of each level has the same dimension and shape as the forecasting target. The final fully connected layer uses ReLU as the activation function.

We choose to use mean absolute error (MAE) as the training objective of the proposed model, which is defined by

$$\mathcal{L} = \frac{1}{Q \times N} \sum_{i=P+1}^{P+Q} \sum_{j=1}^{N} \left| \mathbf{X}_{i,j} - \hat{\mathbf{X}}_{i,j} \right| \tag{14}$$

Notice that we constrained the training process of each level of traffic graphs with the above function, because the training of coarse-grained perspective may not accurately capture the corresponding spatio-temporal features without direct constraints. The noise-containing coarse-grained features will deteriorate the segment-level prediction results when performing the interactions in Section III-C. Thus, the loss function of the multi-level prediction model can be calculated by

$$Loss = \mathcal{L}_S + \lambda_1 \mathcal{L}_R + \lambda_2 \mathcal{L}_Z + \lambda_3 \mathcal{L}_{reg} \tag{15}$$

where λ_1, λ_2 and λ_3 are the regularization weights for balancing loss. Specifically, for $\mathcal{L}_S$, $\mathbf{X}_{P:P+Q}^S$ is the ground truth, and $\mathbf{X}_{P:P+Q}^R$, $\mathbf{X}_{P:P+Q}^Z$ in $\mathcal{L}_R$, $\mathcal{L}_Z$ can be denoted by

$$\begin{aligned}\mathbf{X}_{P:P+Q}^{R} &= \mathbf{S}_{SR}^{\top} \mathbf{X}_{P:P+Q}^{S} \\ \mathbf{X}_{P:P+Q}^{Z} &= \mathbf{S}_{RZ}^{\top} \mathbf{X}_{P:P+Q}^{R} \end{aligned} \tag{16}$$

In addition, the $\mathcal{L}_{reg} = L_1 + L2$ consists of the two regularization terms in Section III-A.3, which is added to the final objective function to obtain more reasonable cluster assignments. The detailed learning process of AIMST is presented in Algorithm 1.

IV. Experiments

In this section, we conduct experiments on XiAn and JiNan datasets to evaluate our method for traffic forecasting from multiple perspectives, including performance comparisons, ablation studies, hyperparameter studies, and efficiency analysis. Meanwhile, the following research questions can be answered.

- **RQ1**: Can our AIMST provide superior performance compared to several state-of-the-art baselines?
- **RQ2**: How is the effectiveness of the different components under our framework?
- **RQ3**: How is AIMST's sensitivity with respect to different hyperparameter settings?
- **RQ4**: How efficient of our AIMST framework when competing with different baselines?

A. Datasets

The two traffic speed datasets used in our experiments are collected by Didi Chuxing GAIA Initiative in XiAn and JiNan cities in China .[1] Both datasets contain the average speed of road segments with the time interval of 10 minutes. Each dataset consists of 52,286 samples. The two datasets contain 792 and 561 road segments (nodes) in the city center area for XiAn and JiNan, respectively. Detailed information on the two datasets is given in Table II below.

We adopt Z-score normalization to process the data in both datasets, which standardizes the features by removing the mean and scaling to unit variance with $\frac{\mathbf{X} - \text{mean}(\mathbf{X})}{\text{std}(\mathbf{X})}$, where mean$(\cdot)$ and std$(\cdot)$ are the mean and the standard deviation of the historical time series, respectively.

[1]The data is available at: https://github.com/guokan987/HGCN.

TABLE II
DETAILED INFORMATION OF THE EVALUATED DATASETS

Properties	Datasets	
	XiAn	JiNan
# of Nodes	792	561
# of Time spans	52,286	52,286
Time interval	10 minutes	10 minutes
Data type	Speed	Speed

B. Experimental Settings

1) Implementation Details: All experiments are conducted on NVIDIA TESLA V100 GPUs with 16G. The proposed model is implemented by Python 3.6 and Pytorch 1.2.0. We train our model using Adam optimizer with a learning rate of 0.001. The batch size is 64 and the training epoch is 60. Each dataset is split into 60% for training, 20% for validation and 20% for testing with chronological order, which is used to early-stop our training algorithm for each model based on the best validation score. We keep the length of the input historical sequence consistent with that of the forecasting sequence, i.e., $P = Q$. Meanwhile, we select the optimal settings of $\{N^R = 160, N^Z = 40\}$ and $\{N^R = 80, N^Z = 20\}$ for XiAn and JiNan datasets, respectively. The order of GCN in ST Block is set to 3 and dil in temporal convolution is set to 2 according to the previous works [45]. λ_1, λ_2 and λ_3 are set to 0.25, 0.15 and 1e-4, respectively. We initialize model parameters with uniform distribution. In addition, in order to verify the impact of the components in our framework, we also design five variants, whose settings are the same as those of our model, except for the corresponding component.

2) Definition of Evaluation Metrics: Three kinds of evaluation metrics are adopted to evaluate the performance of each model, including Root Mean Squared Errors (RMSE), Mean Absolute Errors (MAE), and Mean Absolute Percentage Errors (MAPE), which are defined as:

$$\begin{aligned}
\text{MAE} &= \frac{1}{N \times Q} \sum_{i=1}^{N} \sum_{j=1}^{Q} \left| \hat{\mathbf{X}}_{i,P+j} - \mathbf{X}_{i,P+j} \right| \\
\text{RMSE} &= \sqrt{\frac{1}{N \times Q} \sum_{i=1}^{N} \sum_{j=1}^{Q} \left(\hat{\mathbf{X}}_{i,P+j} - \mathbf{X}_{i,P+j} \right)^2} \\
\text{MAPE} &= \frac{1}{N \times Q} \sum_{i=1}^{N} \sum_{j=1}^{Q} \frac{\left| \hat{\mathbf{X}}_{i,P+j} - \mathbf{X}_{i,P+j} \right|}{\mathbf{X}_{i,P+j}}
\end{aligned} \tag{17}$$

Smaller values indicate higher prediction performance for the three terms of metrics above.

C. Description of Baselines

To demonstrate the effectiveness of the proposed AIMST in traffic forecasting tasks, we compare it with state-of-the-art methods. The methods involved in the experiments are introduced as follows.

- Historical Average (**HA**). HA predicts future traffic states by averaging the historical traffic data.
- Autoregressive Integrated Moving Average (**ARIMA**) [46]. ARIMA combines moving average with autoregression to predict future values in time series.
- Long Short-Term Memory (**LSTM**) [47]. LSTM is a specially designed RNN for additionally involving long-term dependence in future predictions.
- Gated Recurrent Unit (**GRU**) [48]. GRU is a deep learning model based on gating mechanism, which is often used to predict time series data.
- Attention based Spatial-Temporal Graph Convolutional Networks (**ASTGCN**) [29]. ASTGCN designs a trend-aware self-attention module and a dynamic graph convolution module to capture the dynamics flexibly and offers more accurate long-term prediction.
- Graph WaveNet (**GWNET**) [36]. Graph WaveNet is a spatio-temporal network that captures long-range temporal sequences with a stacked dilated 1D convolution component and learns a self-adaptive adjacency matrix through node embedding to capture the spatial dependency.
- Adaptive Graph Convolutional Recurrent Network (**AGCRN**) [34]. AGCRN is a traffic prediction model that automatically captures node-specific spatial and temporal correlations in time-series data without a pre-defined graph.
- Spatial-Temporal Fusion Graph Neural Networks (**STFGNN**) [37]. STFGNN is a GNN-based framework for traffic flow forecasting, which could handle long sequences by learning more spatial-temporal dependencies with layers stacked.
- Hierarchical Graph Convolution Networks (**HGCN**) [38]. HGCN integrates multiple GCNs on different layers of traffic graphs for traffic forecasting considering both the road segment and region features of the traffic system.
- Long Short-term Graph Convolutional Networks (**LSGCN**) [49]. LSGCN is a traffic prediction model that integrates both GCN and a graph attention network.
- Spatial-Temporal Synchronous Graph Convolutional Network (**STSCN**) [33]. STSGCN is a spatial-temporal synchronous graph convolutional networks for spatial-temporal network data forecasting.
- Spatial-Temporal Graph Ordinary Differential Equation Networks (**STODE**) [50]. STODE is a novel tensor-based spatial-temporal forecasting model.
- Meta-Graph Convolutional Recurrent Network (**MegaCRN**) [51]. MegaCRN is a meta-graph convolutional recurrent network along with spatio-temporal graph structure learning mechanism for traffic forecasting.
- **STWave** [52]. STWave is a novel disentangle-fusion framework for traffic forecasting.

GWNET, AGCRN, STSGCN, and MegaCRN are the adaptive GCN-based methods. ASTGCN, LSGCN and STWave are the graph attention-based methods. These models are all representative baselines that are widely used for traffic forecasting. Both STFGNN and HGCN are recent advanced

TABLE III
AVERAGE PERFORMANCE COMPARISON OF DIFFERENT METHODS ON XIAN AND JINAN DATASETS

	Methods	30 min			1 hour			2 hour		
		MAE	RMSE	MAPE	MAE	RMSE	MAPE	MAE	RMSE	MAPE
XiAn	HA	6.02	8.16	21.79%	6.02	8.16	21.79%	6.02	8.16	21.79%
	ARIMA	3.70	6.05	12.96%	4.26	6.57	15.28%	5.04	7.24	18.26%
	LSTM	3.16	4.81	11.92%	3.70	5.52	14.22%	4.52	6.53	17.42%
	GRU	3.15	4.82	11.96%	3.69	5.52	14.25%	4.51	6.53	17.50%
	ASTGCN	3.06	4.67	11.54%	3.49	4.81	13.28%	3.85	5.47	14.32%
	GWNET	2.80	4.31	10.49%	3.21	4.64	11.73%	3.52	5.26	13.46%
	AGCRN	2.78	4.29	10.47%	3.11	4.61	11.69%	3.45	5.15	13.27%
	STFGNN	2.76	4.28	10.47%	2.99	4.56	11.55%	3.31	4.95	12.79%
	LSGCN	3.14	4.79	11.91%	3.66	5.31	13.69%	4.25	5.94	16.03%
	STSGCN	2.82	4.39	10.60%	3.21	4.68	11.80%	3.54	5.30	13.71%
	STODE	2.79	4.32	10.53%	3.12	4.65	11.77%	3.47	5.18	13.41%
	MegaCRN	2.78	4.30	10.50%	3.06	4.61	11.65%	3.40	5.09	13.19%
	STWave	2.76	4.29	10.48%	3.01	4.58	11.59%	3.37	5.02	13.01%
	HGCN	2.75	4.26	10.46%	2.96	4.54	11.44%	3.24	4.85	12.52%
	AIMST (ours)	**2.61**	**4.08**	**9.95%**	**2.80**	**4.33**	**10.80%**	**3.07**	**4.60**	**11.86%**
	Improvements	+5.09%	+4.16%	+4.88%	+5.41%	+4.63%	+5.59%	+5.25%	+5.15%	+5.27%
	Methods	30 min			1 hour			2 hour		
		MAE	RMSE	MAPE	MAE	RMSE	MAPE	MAE	RMSE	MAPE
JiNan	HA	5.69	7.60	20.02%	5.69	7.60	20.02%	5.69	7.60	20.02%
	ARIMA	3.96	6.14	14.12%	4.48	6.56	16.10%	5.09	7.01	18.24%
	LSTM	3.19	4.82	12.74%	3.64	5.43	14.80%	4.26	6.21	17.26%
	GRU	3.17	4.81	12.71%	3.63	5.40	14.76%	4.24	6.21	17.25%
	ASTGCN	3.15	4.68	12.49%	3.51	4.95	13.52%	3.74	5.29	14.43%
	GWNET	2.93	4.42	11.51%	3.17	4.74	12.49%	3.53	5.16	13.82%
	AGCRN	2.91	4.40	11.49%	3.15	4.70	12.48%	3.50	5.13	13.78%
	STFGNN	2.90	4.38	11.42%	3.14	4.69	12.40%	3.43	5.07	13.52%
	LSGCN	3.16	4.69	12.58%	3.56	5.16	14.07%	3.95	5.61	14.78%
	STSGCN	3.12	4.61	12.47%	3.54	5.03	13.98%	3.89	5.57	14.59%
	STODE	3.03	4.54	11.94%	3.48	4.89	12.97%	3.83	5.36	14.11%
	MegaCRN	3.01	4.42	11.60%	3.25	4.74	12.62%	3.63	5.17	13.78%
	STWave	2.90	4.39	11.44%	3.15	4.70	12.43%	3.46	5.09	13.57%
	HGCN	2.89	4.37	11.35%	3.11	4.68	12.31%	3.36	5.02	13.31%
	AIMST (ours)	**2.76**	**4.21**	**10.83%**	**2.96**	**4.46**	**11.72%**	**3.22**	**4.86**	**12.72%**
	Improvements	+4.50%	+3.66%	+4.58%	+4.82%	+4.70%	+4.79%	+4.17%	+3.19%	+4.43%

approaches, where HGCN is also a representative hierarchical model for traffic forecasting.

D. Performance Comparison (RQ1)

The average performances of our proposed approach and all alternative baselines on real-world datasets of both XiAn and JiNan are summarized in Table III. Based on these results, we have the following observations and corresponding analysis:

- Our proposed AIMST outperforms all alternatives in terms of MAE, RMSE, and MAPE for 30-minute, 1-hour, and 2-hour ahead predictions on both XiAn and JiNan datasets, which directly and powerfully verifies the validity of our proposed method in traffic prediction tasks.
- The performances of HA and ARIMA are relatively poor which are based on traditional mathematical statistics and cannot automatically capture complex nonlinear features in traffic data. For LSTM and GRU, they have similar performances and both can capture the temporal correlation of the temporal data to some extent, but cannot model the spatial dependency between road segments.
- Regarding the GCN-based approach, ASTGCN, GWNET and AGCRN can extract the spatio-temporal features of traffic states and perform better than the basic temporal network. The newer work STFGNN achieves better performance, which captures hidden spatial dependencies with a novel data-driven graph and its further fusion with a given spatial graph.
- Regarding the adaptive GCN-based approach, STSGCN and MegaCRN can adaptively learn the adjacency relationship between nodes in an end-to-end manner, which is used as input to the GCN for modeling spatial cor-

relation. However, this approach discards the original information of the road network, so that the factors with a great influence on the traffic pattern, such as the geographic structure of the road network, are not fully considered. Moreover, the optimization of the adjacency matrix is one of the difficulties of the network, which will bring challenges to the training of the model.

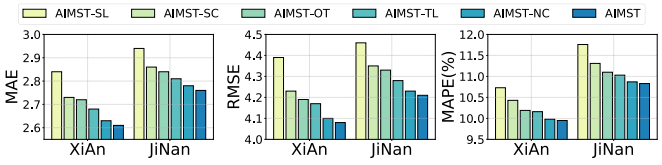
- Regarding the graph attention-based approach, the performance of STWave is considerable, due to incorporating a graph wavelet-based graph positional encoding into the full graph attention network to model dynamic spatial correlations. However, this method only calculates value-based spatial semantic correlations and lacks the structural information of the graph, which may result in over-fitting. Moreover, Due to the need to calculate the attention coefficient of each node pair, the GAT-based method has a large number of parameters and high complexity.
- The performance of HGCN is considerably improved, due to simultaneous consideration of segment- and region-level spatio-temporal characteristics. However, HGCN fails to divide reasonable multi-level spatial structure, in addition to ignoring the potential influence of the segment-level traffic state on regions.

Compared with other models, the improvement of our proposed AIMST can be attributed to the effective construction of multi-level traffic networks and the quantification of traffic pattern interactions in urban multi-grained spaces. We will further investigate our conjecture in the following parts.

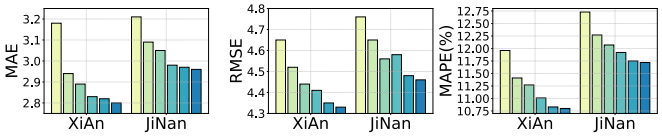
E. Ablation Study (RQ2)

To further illustrate the effectiveness of different components in AIMST, we design five variants to conduct ablation experiments and analyze experimental results on both XiAn and JiNan datasets, including: The RMSE, MAPE and MAE for 1-hour prediction results are listed below:

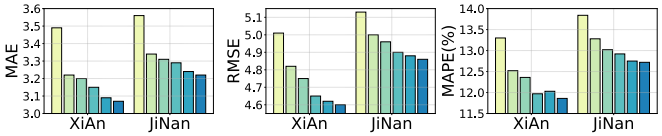
- **AIMST-SL** (**S**ingle **L**ayer): This variant removes the region- and zone-level, and only uses the road segment feature without multi-level structure for prediction.
- **AIMST-SC** (**S**pectral **C**lustering): This variant adopts the simple spectral clustering [38] to construct the multi-level traffic graphs, instead of the adaptive hierarchical clustering we introduced in Section III-A.
- **AIMST-OT** (**O**ne-way **T**ransfer): This variant does not possess the multi-level bidirectional interaction module proposed by us, and only concatenates coarse-grained features to the corresponding fine-grained features [38] to achieve one-way transfer among levels.
- **AIMST-TL** (**T**wo **L**ayers): This variant omits the zone-level, and models the traffic network as a two-level graph structure, i.e., segment- and region-level.
- **AIMST-NC** (**N**o **C**onstraint): This variant only applies the Eq. (14) to constrain the training process for the segment-level, and no additional constraints are considered to the other levels.



(a) Performance of different variants for 30-min forecasting



(b) Performance of different variants for 1-hour forecasting



(c) Performance of different variants for 2-hour forecasting

Fig. 6. Component ablation analysis of AIMST.

As demonstrated in Figure 6, we observe that removing the constraints from other levels causes the performance to worsen a little, which demonstrates constraints on the training process are beneficial at all levels. Constructing the traffic graph as a two-level structure greatly hurts the performance, which proves urban layout is not a simple two-level structure and only two-level networks cannot effectively model hierarchical traffic patterns driven by urban layout. Therefore, when adopting a multi-level structure, we can exploit hierarchical relationships better to make the constructed multi-level graph structure closer to the real natural layout of the city. Replacing the multi-level bidirectional interaction with the one-way transfer is much worse than the proposed AIMST, which demonstrates that the effects between different granularities are mutual in multi-granularity spatio-temporal learning, rather than a simple one-way auxiliary relationship. Using spectral clustering to construct multi-level traffic graphs degrades the performance quite severely, demonstrating that spectral clustering has a limited ability to delineate urban hierarchies. Our proposed adaptive hierarchical clustering method can automatically exploit the potential hierarchical relationships of traffic patterns through end-to-end learning, so as to adaptively generate more reasonable multi-level traffic graphs. Only considering the road segment feature without multi-level structure causes the worst performance, similar to the case of two layers, which demonstrates that the traffic network is a multi-level system closely related to urban layout.

In a nutshell, each component we proposed in our AIMST, such as the multi-level bidirectional interaction, adaptive hierarchical clustering and multi-level traffic network, contributes to the final forecasting performance and plays an indispensable role.

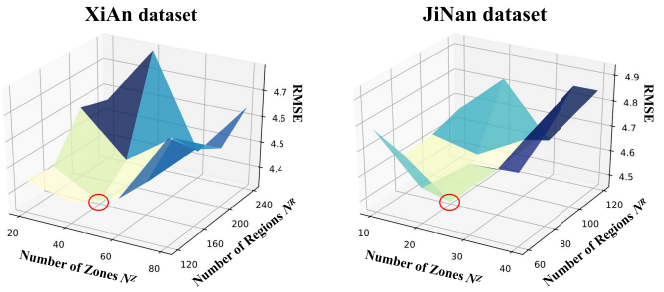


Fig. 7. Performance changes under different N^R and N^Z. The coordinates of the red circles "○" indicate the optimal settings.

F. Hyperparameter Study (RQ3)

To investigate the effect of different parameter settings, we perform experiments to evaluate the performance of our proposed AIMST framework with different configurations of key hyperparameters (e.g., numbers of region- and zone-level nodes, regularization weights in the loss function, etc). When varying a specific hyperparameter for effect investigation, we keep other parameters with default values.

1) Sensitivity to Numbers of Region- and Zone-Level Nodes: First, to investigate the effect of different numbers of N^R and N^Z on the performance of the proposed AIMST. We conduct experiments and observe the performance variations under different values of N^R and N^Z. Here, N^R is the number of region-level nodes, and N^Z is the number of zone-level nodes.

The performance variations of 1-hour ahead prediction on two datasets are characterized in Figure 7. According to the results, we obtain the optimal settings of $\{N^R = 160, N^Z = 40\}$ and $\{N^R = 80, N^Z = 20\}$ on XiAn and JiNan datasets, respectively. Further, the corresponding values of N^R and N^Z should be different for cities of different sizes and development levels. Generally speaking, cities with larger scales and higher development levels not only have more sound road networks and more developed transportation infrastructure, but also have more perfect urban functional zoning due to the large population and high level of economic development. As shown in Figure 7, urban transportation in XiAn may be more developed than that in JiNan. Therefore, these two parameters in the multi-level traffic prediction model also reflect the level of traffic service and the soundness of urban function of the city to some extent.

2) Sensitivity to Regularization Weights: Then, to explore the effect of regularization weights in the loss function, we fix the weight of main task as 1, and tune λ_1, λ_2 by grid searching. Specifically, we respectively vary them from $\{0.10, 0.15, 0.20, 0.25, 0.30\}$ and $\{0.05, 0.10, 0.15, 0.20\}$, and observe the variations in model performance.

For succinctness, we only report the MAE errors of 1-hour ahead prediction on the XiAn dataset in Table IV. The experimental results demonstrate that the best performance can be achieved with $\{\lambda_1 = 0.25, \lambda_2 = 0.15\}$, and the results on the JiNan dataset show a similar trend. Moreover, when the regularization weights are tiny, the performance of the model decreases obviously. This phenomenon further proves that proper constraints can help the spatio-temporal learning

TABLE IV
PERFORMANCE CHANGES UNDER DIFFERENT VALUES OF λ_1, λ_2

λ_2	λ_1				
	0.10	0.15	0.20	0.25	0.30
0.05	2.93	2.89	2.88	2.84	2.86
0.10	2.91	2.90	2.87	2.83	2.85
0.15	2.88	2.89	2.85	**2.80**	2.83
0.20	2.90	2.92	2.86	2.84	2.87

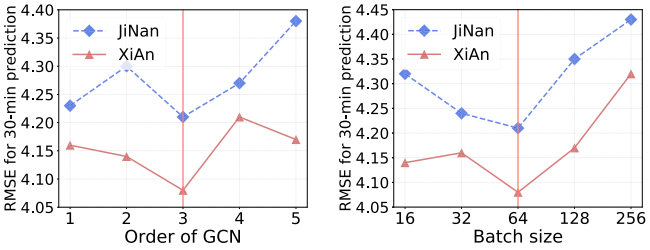


Fig. 8. Impacts of GCN's order and batch size in the proposed AIMST. (a) Study on the order of GCN in ST Blocks; (b) Study on the batch size.

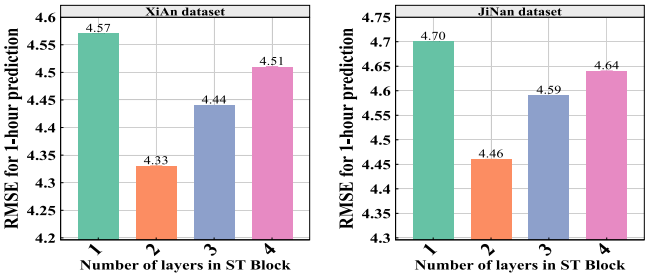


Fig. 9. Impacts of the number of the layers of ST Block in AIMST.

for each level during the training, and thus provide more accurate guidance for fine-grained target tasks.

3) Sensitivity to GCN's Order and Batch Size: After that, to test the impact of the GCN's order in the ST Blocks, we vary it from 1 to 5. The results are reported in Figure 8 (a). Overall, our model achieves the best performance when the order is set to 3.

Finally, we vary the batch size from the range of {16, 32, 64, 128, 256}. The results are reported in Figure 8 (b). We can observe that the model achieves the best result when the batch size is set to 64. Then the performance drops severely by further decreasing or increasing it. Thus, we set the batch size to 64 in our model to obtain optimal performance.

4) Sensitivity to the Layer of ST Block: To test the impact of the layer of ST Block, we vary it from 1 to 4. The results are reported in Figure 9. We can observe that the model achieves the best result when the layer of ST Block is set to 2. Then the performance drops severely by further decreasing or increasing it. Thus, we set the layer of ST Block to 2 in our model to obtain optimal performance.

G. Efficiency Analysis (RQ4)

We compare the computation time of the key baselines, our proposed AIMST, and the AIMST-TL variant. The results are presented in Table V.

TABLE V
COMPUTATION TIME COST INVESTIGATION

Methods	Computation Time (JiNan/XiAn)	
	Training (s/epoch)	Inference (s)
ASTGCN	215.60/308.09	14.28/28.57
GWNET	176.69/267.16	11.15/21.72
AGCRN	190.01/271.89	13.02/26.74
STFGNN	245.37/351.20	15.21/28.96
HGCN	80.16/124.17	5.56/7.63
AIMST-TL	93.54/137.35	6.16/11.71
AIMST	**126.63/174.49**	**8.37/15.22**

TABLE VI
COMPUTATION TIME COST INVESTIGATION ON LARGE-SCALE GRAPH

Methods	Computation Time on LargeST	
	Training (s/epoch)	Inference (s)
LSGCN	2016.08	257.43
STSGCN	2160.45	259.18
STODE	1866.37	249.28
MegaCRN	2305.71	282.91
STWave	2505.19	295.53
AIMST (ours)	**1056.42**	**107.19**

Taken together, the data in Table V indicates that AIMST achieves better efficiency than most baselines with respect to the model computational cost, which mainly benefits from applying the temporal convolution with fully parallel training. Meanwhile, compared with HGCN, our AIMST is slightly slower than HGCN, while much more accurate than HGCN, which illustrates that the cost of multi-level structure and automatic online clustering is deserved. However, we notice that HGCN is implemented with the offline clustering process, while ours is totally end-to-end, making these two solutions not directly comparable. Further analysis can be obtained that, the AIMST-TL variant is not much slower than HGCN. This is because the designed adaptive spatio-temporal hierarchical clustering only incurs a small computational cost with the regularization terms of assignment optimization. Therefore, the substantial experimental results effectively validate that AIMST is an efficient method with high performance.

We specially select a dataset termed LargeST [53] with a large number of road segments to further investigate the computational cost in the large-scale graph, which has 8600 nodes with 5-minute intervals and is the largest spatio-temporal dataset to our best knowledge. LSGCN, STSGCN, STODE, MegaCRN and STWave are chosen for the comparison. We compare the computation time of the above baselines with our proposed AIMST, and the results of 1-hour ahead forecasting are presented in Table VI.

The results in Table VI indicate that AIMST achieves better efficiency than most baselines with respect to the model computational cost for the large-scale graph, which mainly benefits from applying the layer-wise linear structure defined by stacking multiple localized graph convolutional layers with the first-order approximation of graph Laplacian [35], [54], and temporal convolution with fully parallel training. The graph attention-based models, such as LSGCN and STWave, have a quadratic calculation complexity about the sensor number N due to calculating the attention coefficient of each node pair, and N is very large in the used dataset, thus bringing unaffordable computation needs. Regarding the adaptive GCN-based models, STSGCN and MegaCRN need more parameters to optimize the adjacency, which brings a great burden on the computation cost. STODE achieves sub-optimal results because it predetermined the graph structure of the road network based on geographic and temporal distances during the data processing phase, allowing the computational overhead to be reduced. Therefore, the substantial experimental results effectively validate that AIMST remains an efficient and high-performance method even on large-scale road networks.

Our model is trained offline and does not need to be retrained for online prediction. As shown in Table V and Table VI, such prediction time is much lower than the duration of a time slot, demonstrating that our model can be applied in real-world applications.

H. Case Study

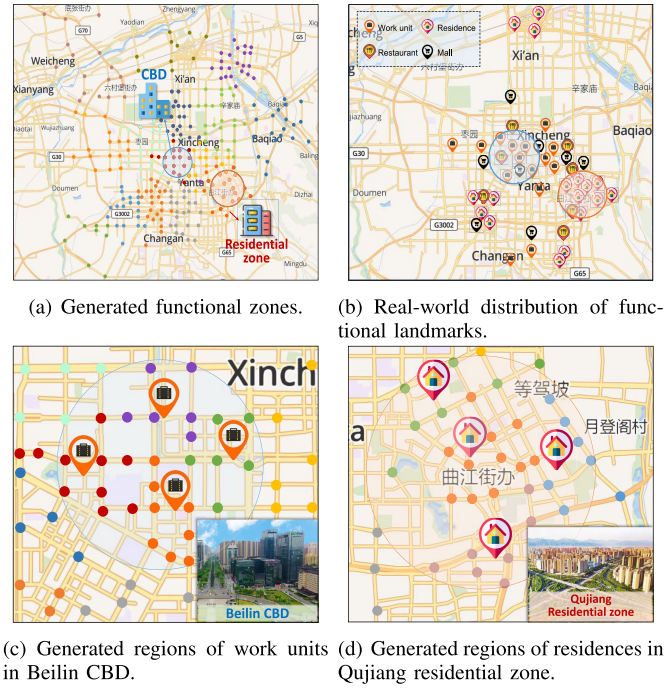
To further intuitively illustrate the effectiveness of our proposed adaptive spatio-temporal hierarchical clustering for generating multi-level urban structure, we visualize the generated regions and zones by our method and the real functional landmarks of XiAn city as depicted in Figure 10.

Figure 10 (a) and (b) clearly show that the zones generated by our proposal are exactly consistent with the real urban functional area, e.g., Beilin CBD and Qujiang residential zone. We further analyze the CBD and residential zone obtained in Figure 10 (a) to visualize the corresponding regions. According to Figure 10 (c) and (d), each functional zone consists of more fine-grained functional units, i.e., regions. These generated regions also accurately correspond to the real urban functional regions. Therefore, our proposal can effectively model a reasonable multi-level urban structure, and decouple the evolutionary relationship between traffic patterns and urban space.

V. DISCUSSION

A. Contributions to Intelligent Transportation System

Effective traffic management is essential to improve the service quality of Intelligent Transportation Systems (ITS), thereby advancing the progression of smart cities [4], [55], [56]. Traffic forecasting is an indispensable part of urban traffic management and a promising research topic in the field of ITS [4], [56], [57]. In this paper, we provide a brand-new solution to tackle traffic forecasting from a spatio-temporal multi-granularity perspective. Specifically, inspired by natural multi-level urban structure and layout, we first devise an adaptive hierarchical clustering method to automatically generate a multi-level traffic network. Besides, we discover the importance of bidirectional effects across granularities in spatio-temporal learning, and design a bidirectional interaction module to couple the multi-grained spatio-temporal patterns.



(a) Generated functional zones.

(b) Real-world distribution of functional landmarks.

(c) Generated regions of work units in Beilin CBD.

(d) Generated regions of residences in Qujiang residential zone.

Fig. 10. Visualization study on the result of generating multi-level urban structure. Each dot represents the road segment where located, and several dots of the same color are gathered into regions or zones, indicating the agglomeration effect of urban functions in real-world scenarios.

These findings and implementations carry important implications for traffic forecasting and make substantial contributions to existing spatio-temporal learning techniques. Consequently, this paper provides solid technical support and application assurances for the development of ITS.

B. Spatial Semantic Information Modeling in AIMST

In this paper, while focusing on the natural spatial structure of the city, we also involve the modeling of spatial semantic information, which is mainly achieved from two aspects.

On the one hand, we divide semantically similar nodes/regions by learning a reasonable clustering assignment from the feature matrices end-to-end, thus making the spatial semantic richer, and forming subspaces that both fit with the actual urban layout and possess functional semantics. And this paper is exactly inspired by the multi-level functional structure driven by the natural urban layout to study multi-granularity spatio-temporal learning. Although the spatial semantic we extract is more coarse-grained, this spatial representation we extract not only reflects the semantic information, but also models the native road network structure information well, which is crucial to the motivation of this paper.

On the other hand, we adaptively interact the global spatial semantics with the fine-grained node representations through a multi-level bidirectional interaction module, which enables the fusion and utilization of semantic information beneficial to the task of fine-grained node prediction, whereas directly interacting node-level semantics is too costly and difficult to assist downstream prediction. Therefore, through the multi-level interaction proposed in this paper, the fine-grained node representation contains both spatial structure information and spatial semantic information that facilitates the downstream task.

C. Future Work

Research on the multi-level spatio-temporal network is crucially promising in the field of data mining and urban computing. Regarding our proposed AIMST framework, we discuss below the possible future directions of the proposal.

i) Further explore the implications of hierarchy in multi-granularity spatio-temporal modeling. As studied in this paper, the hierarchical division of urban space has a significant impact on both spatial modeling and downstream urban computing tasks. Substantial experiments have shown that the three-level traffic network we designed is a superb simulation of the multi-level urban structure, thus better improving the accuracy of fine-grained prediction tasks. Also, we will investigate the impact of modeling higher-level urban space, such as four or more levels. Meanwhile, experimental results in Section IV-F.1 show that hierarchical division is sensitive to different datasets, which inspires us to explore the synergistic relationship between hierarchy and spatio-temporal data properties. Therefore, in multi-granularity spatio-temporal modeling, decoupling the implications of hierarchy will help us develop more effective multi-level spatio-temporal network architecture.

ii) Generalize the AIMST to broader spatio-temporal data scenarios. In the field of spatio-temporal data mining, the hierarchical property is ubiquitous in a variety of spatio-temporal data, such as COVID-19 pandemic data [58], air quality data [59] and urban crime data [60]. Specifically, for the pandemic prediction task, communities in highly contagious areas will arise more COVID-19 cases, making modeling the hierarchical structure of pandemic spread critical to early discovery of COVID-19 cases. Similarly, urban air quality possesses the corresponding aggregation effects at different geographical scales, and crime incidents are also concentrated in high-risk criminal areas. Our proposed AIMST framework enables effective hierarchical spatio-temporal modeling, thus providing a novel and generalizable scheme for various spatio-temporal data scenarios.

VI. Related Work

We briefly review related literature, including spatio-temporal forecasting and road network modeling.

A. Spatio-Temporal Forecasting

Spatial-temporal forecasting plays a vital role in many application areas (e.g., traffic forecasting), and has attracted much attention from enormous researchers in the past decades [61], [62]. It constitutes a typical supervised learning paradigm, drawing insights from historical data to predict future trends. As per urban computing theories [10], spatio-temporal forecasting, grounded in vast urban datasets, forms a critical cornerstone for intelligent decision-making, scheduling, and management in smart cities [4]. In the early stage, traditional models are mainly based on mathematical statistics to perform

forecasting, such as HA and ARIMA [15]. These linear models can leverage accumulated traffic data, but they typically perform poorly due to the nonlinear correlations in traffic data. To relax the linear dependency assumption, machine learning-based methods such as SVR [63] and KNN [64] are proposed. These methods can model more complex dependencies and deliver better results than the linear models, but their shallow architecture leads to limited capacity for capturing more complex patterns and makes them inferior in big data scenarios [20].

In recent years, deep learning based techniques have been widely employed and achieved state-of-the-art performance in various spatio-temporal data applications. The RNN-based models, such as LSTM [47] and GRU [48], could exploit patterns from sequential data and be leveraged in traffic forecasting [65]. However, spatial dependencies in the spatio-temporal data were omitted which encourages researchers to propose methods taking spatial dependencies into consideration. And the Convolutional Neural Networks (CNN) are widely used to capture the spatial correlations in grid-based traffic network [24]. However, much spatio-temporal data is graph-structured in nature, such as road network [26] and subway network [28], to which CNNs are not applicable for spatial features representation. To incorporate spatial dependencies more effectively, recent works introduce Graph Neural Networks (GNN) to learn the traffic networks [66], [67], [68], [69]. STGCN [35] and Graph WaveNet [36] employ graph convolution on spatial domain and 1-D convolution along time axis. ASTGCN [29] uses attention-based spatial-temporal graph convolutions to model dynamic spatial-temporal features of traffic flows. AGCRN [34] utilizes node adaptive parameter learning to automatically capture node-specific spatial and temporal correlations in time-series data without a predefined graph. STFGNN [37] utilizes hidden states from previous multiple time steps to handle long sequences in spatio-temporal forecasting. Regarding the graph attention-based method, LSGCN [49] proposes to integrate both GCN and a graph attention network into the spatial gated block for long and short-term traffic prediction. Lastjomer [70] designs a spatio-temporal joint attention in the Transformer architecture to capture all dynamic dependencies in the traffic data. STWave [52] provides a novel disentangle-fusion framework to mitigate the distribution shift issue. On this basis, STWave^{+} [56] is proposed as a novel disentangle-fusion framework for traffic forecasting, which incorporates a multi-scale graph wavelet positional encoding to capture hierarchical spatial dependencies. One fatal drawback with these GNN-based models is that they capture the spatial correlations in one granularity, which cannot adapt to multi-level systems (e.g., traffic networks).

To this end, there are some preliminary attempts at this point. STHMLP [71] proposes a spatio-temporal hierarchical MLP network for traffic forecasting, which focuses to some extent on road-regional correlation using a learnable matrix. HGCN [38] adopts the hierarchical GNN module to simulate the traffic patterns in two-layer spatial granularity. However, these methods still fail to model the multi-level structure of urban road network effectively, and ignore the multi-granularity information interaction in spatio-temporal learning, which are the problems we tackle in this paper.

B. Road Network Modeling

Road network is the basic component of the transportation systems, and research on road network modeling is indispensable for various spatio-temporal learning tasks, such as traffic forecasting, route planning [41], [72], [73], [74], arrival time estimation [40], [75]. With the rapid growth of deep learning techniques, several studies try to learn effective node representations from the road network, including graph convolution networks [39], graph attention network [41] and other types of networks [76]. As aforementioned, a city is normally a multi-level hierarchical system in the real world. Although these studies have improved the application performance with the enhanced data representations, they all focus on learning the representation of the original road network from a single granularity perspective, which does not have enough capacity to model the complicated structure, layout, functions, and proximity of modern cities. Recently, several preliminary studies have looked at this problem, such as HGCN [38] and HRHR [77]. Specifically, HGCN [38] generates coarse-grained traffic graphs using traditional spectral clustering methods offline, which is not an end-to-end learning model and cannot model multi-level road networks adaptively. HRHR [77] uses the graph attention network to learn the hierarchical road network representation, but it lacks consideration of the multi-granularity information effective interaction which is indispensable for spatio-temporal learning in the hierarchical road network scenarios, making it not applicable for multi-granularity spatio-temporal forecasting.

In contrast, our proposal combines effective hierarchical road network modeling and multi-granularity information interaction, together providing a more comprehensive scheme for extensive spatio-temporal learning tasks.

VII. Conclusion

In this paper, we propose a novel Adaptive and Interactive Multi-level Spatio-Temporal network (AIMST) for traffic forecasting. We first construct the hierarchical traffic graph satisfying the multi-level structure of the urban natural layout by introducing a learnable adaptive hierarchical clustering method. Next, we carefully devise a multi-level bidirectional interaction module to mutually guide multi-grained spatio-temporal feature learning at each level. Performance evaluations on two real-world datasets and the ablation study both powerfully demonstrate the effectiveness of our proposal. Therefore, our work provides a brand-new solution to tackle traffic forecasting from a spatio-temporal multi-granularity perspective, which consequently not only provides insights for urban planning, improves the efficiency of public transportation, but also guarantees public safety. Regarding future work, we intend to investigate more general multi-granularity spatio-temporal learning tasks such as meteorological prediction, human mobility, and crime analysis.

REFERENCES

[1] J. Zhang, F.-Y. Wang, K. Wang, W.-H. Lin, X. Xu, and C. Chen, "Data-driven intelligent transportation systems: A survey," *IEEE Trans. Intell. Transp. Syst.*, vol. 12, no. 4, pp. 1624–1639, Dec. 2011.

[2] H. Qu, Y. Gong, M. Chen, J. Zhang, Y. Zheng, and Y. Yin, "Forecasting fine-grained urban flows via spatio-temporal contrastive self-supervision," *IEEE Trans. Knowl. Data Eng.*, vol. 35, no. 8, pp. 8008–8023, Aug. 2023.

[3] X. Wang et al., "Latent Gaussian processes based graph learning for urban traffic prediction," *IEEE Trans. Veh. Technol.*, vol. 73, no. 1, pp. 282–294, Jan. 2024.

[4] G. Jin et al., "Spatio-temporal graph neural networks for predictive learning in urban computing: A survey," *IEEE Trans. Knowl. Data Eng.*, early access, Nov. 23, 2023, doi: 10.1109/TKDE.2023.3333824.

[5] M. R. Jabbarpour, H. Zarrabi, R. H. Khokhar, S. Shamshirband, and K.-K.-R. Choo, "Applications of computational intelligence in vehicle traffic congestion problem: A survey," *Soft Comput.*, vol. 22, no. 7, pp. 2299–2320, Apr. 2018.

[6] 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.

[7] Z. Lv, J. Xu, K. Zheng, H. Yin, P. Zhao, and X. Zhou, "LC-RNN: A deep learning model for traffic speed prediction," in *Proc. 27th Int. Joint Conf. Artif. Intell.*, Jul. 2018, pp. 3470–3476.

[8] C. Zheng, X. Fan, C. Wen, L. Chen, C. Wang, and J. Li, "DeepSTD: Mining spatio-temporal disturbances of multiple context factors for citywide traffic flow prediction," *IEEE Trans. Intell. Transp. Syst.*, vol. 21, no. 9, pp. 3744–3755, Sep. 2020.

[9] K. Wang, Z. Zhou, X. Wang, P. Wang, Q. Fang, and Y. Wang, "A2DJP: A two graph-based component fused learning framework for urban anomaly distribution and duration joint-prediction," *IEEE Trans. Knowl. Data Eng.*, vol. 35, no. 12, pp. 11984–11998, Dec. 2023.

[10] Z. Yu, C. Licia, W. Ouri, and Y. Hai, "Urban computing: Concepts, methodologies, and applications," *ACM Trans. Intell. Syst. Technol.*, vol. 5, no. 3, pp. 1–55, Sep. 2014.

[11] 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.

[12] Z. Zhou et al., "Maintaining the status quo: Capturing invariant relations for OOD spatiotemporal learning," in *Proc. 29th ACM SIGKDD Conf. Knowl. Discovery Data Mining*, Aug. 2023, pp. 3603–3614.

[13] B. L. Smith and M. J. Demetsky, "Traffic flow forecasting: Comparison of modeling approaches," *J. Transp. Eng.*, vol. 123, no. 4, pp. 261–266, Jul. 1997.

[14] M. S. Ahmed and A. R. Cook, "Analysis of freeway traffic time-series data by using Box-Jenkins techniques," *Transp. Res. Rec.*, no. 722, 1979.

[15] B. M. Williams, P. K. Durvasula, and D. E. Brown, "Urban freeway traffic flow prediction: Application of seasonal autoregressive integrated moving average and exponential smoothing models," *Transp. Res. Rec.*, vol. 1644, no. 1, pp. 132–141, 1998.

[16] S. Lee and D. B. Fambro, "Application of subset autoregressive integrated moving average model for short-term freeway traffic volume forecasting," *Transp. Res. Record, J. Transp. Res. Board*, vol. 1678, no. 1, pp. 179–188, Jan. 1999.

[17] W. Hu, L. Yan, K. Liu, and H. Wang, "A short-term traffic flow forecasting method based on the hybrid PSO-SVR," *Neural Process. Lett.*, vol. 43, no. 1, pp. 155–172, Feb. 2016.

[18] C. Tebaldi and M. West, "Bayesian inference on network traffic using link count data," *J. Amer. Stat. Assoc.*, vol. 93, no. 442, p. 557, Jun. 1998.

[19] G. A. Davis and N. L. Nihan, "Nonparametric regression and short-term freeway traffic forecasting," *J. Transp. Eng.*, vol. 117, no. 2, pp. 178–188, Mar. 1991.

[20] J. Liu, T. Li, P. Xie, S. Du, F. Teng, and X. Yang, "Urban big data fusion based on deep learning: An overview," *Inf. Fusion*, vol. 53, pp. 123–133, Jan. 2020.

[21] X. Wang et al., "An observed value consistent diffusion model for imputing missing values in multivariate time series," in *Proc. 29th ACM SIGKDD Conf. Knowl. Discovery Data Mining*, Aug. 2023, pp. 2409–2418.

[22] H. Yao et al., "Deep multi-view spatial–temporal network for taxi demand prediction," in *Proc. AAAI Conf. Artif. Intell.*, vol. 32, no. 1, 2018, pp. 1–8.

[23] 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, Aug. 2019.

[24] X. Ma, Z. Dai, Z. He, J. Ma, Y. Wang, and Y. Wang, "Learning traffic as images: A deep convolutional neural network for large-scale transportation network speed prediction," *Sensors*, vol. 17, no. 4, p. 818, 2017.

[25] B. Wang et al., "Towards dynamic spatial–temporal graph learning: A decoupled perspective," in *Proc. AAAI Conf. Artif. Intell.*, vol. 38, no. 8, 2024, pp. 9089–9097.

[26] L. Yan, H. Shen, J. Zhao, C. Xu, F. Luo, and C. Qiu, "CatCharger: Deploying wireless charging lanes in a metropolitan road network through categorization and clustering of vehicle traffic," in *Proc. IEEE Conf. Comput. Commun. (INFOCOM)*, May 2017, pp. 1–9.

[27] B. Wang et al., "Pattern expansion and consolidation on evolving graphs for continual traffic prediction," in *Proc. 29th ACM SIGKDD Conf. Knowl. Discovery Data Mining*, Aug. 2023, pp. 2223–2232.

[28] J. Wang, Y. Zhang, Y. Wei, Y. Hu, X. Piao, and B. Yin, "Metro passenger flow prediction via dynamic hypergraph convolution networks," *IEEE Trans. Intell. Transp. Syst.*, vol. 22, no. 12, pp. 7891–7903, Dec. 2021.

[29] 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.*, vol. 33, no. 1, 2019, pp. 922–929.

[30] 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. Discovery Data Mining*, 2019, pp. 1720–1730.

[31] 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.*, Nov. 2019, pp. 2293–2296.

[32] X. Geng et al., "Spatiotemporal multi-graph convolution network for ride-hailing demand forecasting," in *Proc. AAAI Conf. Artif. Intell.*, vol. 33, 2019, pp. 3656–3663.

[33] 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, pp. 914–921.

[34] L. Bai, L. Yao, C. Li, and X. Wang, "Adaptive graph convolutional recurrent network for traffic forecasting," in *Proc. NIPS*, 2020, pp. 17804–17815.

[35] 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., (IJCAI)*, 2018, pp. 3634–3640.

[36] 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.

[37] M. Li and Z. Zhu, "Spatial–temporal fusion graph neural networks for traffic flow forecasting," in *Proc. 35th AAAI Conf. Artif. Intell.*, 2021, pp. 4189–4196.

[38] K. Guo, Y. Hu, Y. Sun, S. Qian, J. Gao, and B. Yin, "Hierarchical graph convolution networks for traffic forecasting," in *Proc. 35th AAAI Conf. Artif. Intell.*, 2021, pp. 151–159.

[39] T. S. Jepsen, C. S. Jensen, and T. D. Nielsen, "Graph convolutional networks for road networks," in *Proc. 27th ACM SIGSPATIAL Int. Conf. Adv. Geographic Inf. Syst.*, Nov. 2019, pp. 460–463.

[40] Y. Li, K. Fu, Z. Wang, C. Shahabi, J. Ye, and Y. Liu, "Multi-task representation learning for travel time estimation," in *Proc. 24th ACM SIGKDD Int. Conf. Knowl. Discov. Data Mining*, London, U.K., Jul. 2018, pp. 1695–1704.

[41] J. Wang, N. Wu, W. X. Zhao, F. Peng, and X. Lin, "Empowering A* search algorithms with neural networks for personalized route recommendation," in *Proc. 25th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, 2019, pp. 539–547.

[42] Z. Ying, J. You, C. Morris, X. Ren, W. Hamilton, and J. Leskovec, "Hierarchical graph representation learning with differentiable pooling," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 31, 2018, pp. 1–12.

[43] S. Bai, J. Z. Kolter, and V. Koltun, "An empirical evaluation of generic convolutional and recurrent networks for sequence modeling," 2018, *arXiv:1803.01271*.

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

[45] M. Welling and T. N. Kipf, "Semi-supervised classification with graph convolutional networks," in *Proc. J. Int. Conf. Learn. Represent.*, 2017, pp. 1–11.

[46] B. Pan, U. Demiryurek, and C. Shahabi, "Utilizing real-world transportation data for accurate traffic prediction," in *Proc. IEEE 12th Int. Conf. Data Mining*, Dec. 2012, pp. 595–604.

[47] Z. Cui, R. Ke, Z. Pu, and Y. Wang, "Stacked bidirectional and unidirectional LSTM recurrent neural network for forecasting network-wide traffic state with missing values," *Transp. Res. C, Emerg. Technol.*, vol. 118, Sep. 2020, Art. no. 102674.

[48] A. F. M. Agarap, "A neural network architecture combining gated recurrent unit (GRU) and support vector machine (SVM) for intrusion detection in network traffic data," in *Proc. 10th Int. Conf. Mach. Learn. Comput.*, Feb. 2018, pp. 26–30.

[49] R. Huang, C. Huang, Y. Liu, G. Dai, and W. Kong, "LSGCN: Long short-term traffic prediction with graph convolutional networks," in *Proc. 29th Int. Joint Conf. Artif. Intell.*, Jul. 2020, pp. 2355–2361.

[50] Z. Fang, Q. Long, G. Song, and K. Xie, "Spatial–temporal graph ODE networks for traffic flow forecasting," in *Proc. 27th ACM SIGKDD Conf. Knowl. Discov. Data Min.*, 2021, pp. 364–373.

[51] 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.

[52] Y. Fang et al., "When spatio-temporal meet wavelets: Disentangled traffic forecasting via efficient spectral graph attention networks," in *Proc. IEEE 39th Int. Conf. Data Eng. (ICDE)*, Apr. 2023, pp. 517–529.

[53] X. Liu et al., "LargeST: A benchmark dataset for large-scale traffic forecasting," 2023, *arXiv:2306.08259*.

[54] T. N. Kipf and M. Welling, "Semi-supervised classification with graph convolutional networks," 2016, *arXiv:1609.02907*.

[55] G. Dimitrakopoulos and P. Demestichas, "Intelligent transportation systems," *IEEE Veh. Technol. Mag.*, vol. 5, no. 1, pp. 77–84, Mar. 2010.

[56] Y. Fang, Y. Qin, H. Luo, F. Zhao, and K. Zheng, "STWave+: A multi-scale efficient spectral graph attention network with long-term trends for disentangled traffic flow forecasting," *IEEE Trans. Knowl. Data Eng.*, early access, Oct. 17, 2023, doi: 10.1109/TKDE.2023.3324501.

[57] A. Boukerche and J. Wang, "Machine learning-based traffic prediction models for intelligent transportation systems," *Comput. Netw.*, vol. 181, Nov. 2020, Art. no. 107530.

[58] Y. Ma, P. Gerard, Y. Tian, Z. Guo, and N. V. Chawla, "Hierarchical spatio-temporal graph neural networks for pandemic forecasting," in *Proc. 31st ACM Int. Conf. Inf. Knowl. Manage.*, Oct. 2022, pp. 1481–1490.

[59] J. Han, H. Liu, H. Xiong, and J. Yang, "Semi-supervised air quality forecasting via self-supervised hierarchical graph neural network," *IEEE Trans. Knowl. Data Eng.*, vol. 35, no. 5, pp. 5230–5243, May 2023.

[60] 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. (ICDE)*, May 2022, pp. 2984–2996.

[61] Y. Lv, Y. Duan, W. Kang, Z. Li, and F.-Y. Wang, "Traffic flow prediction with big data: A deep learning approach," *IEEE Trans. Intell. Transp. Syst.*, vol. 16, no. 2, pp. 865–873, Apr. 2014.

[62] 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.

[63] C.-H. Wu, J.-M. Ho, and D. T. Lee, "Travel-time prediction with support vector regression," *IEEE Trans. Intell. Transp. Syst.*, vol. 5, no. 4, pp. 276–281, Dec. 2004.

[64] H. van Lint and C. van Hinsbergen, "Short-term traffic and travel time prediction models," *Artif. Intell. Appl. Critical Transp.*, vol. 22, pp. 22–41, Nov. 2012.

[65] X. Ma, Z. Tao, Y. Wang, H. Yu, and Y. Wang, "Long short-term memory neural network for traffic speed prediction using remote microwave sensor data," *Transp. Res. C, Emerg. Technol.*, vol. 54, pp. 187–197, May 2015.

[66] B. Wang et al., "Knowledge expansion and consolidation for continual traffic prediction with expanding graphs," *IEEE Trans. Intell. Transp. Syst.*, vol. 24, no. 7, pp. 7190–7201, Jul. 2023.

[67] B. Wang et al., "Kill two birds with one stone: Rethinking data augmentation for deep long-tailed learning," in *Proc. ICLR*, 2024.

[68] Z. Zhao, P. Wang, H. Wen, Y. Zhang, Z. Zhou, and Y. Wang, "A twist for graph classification: Optimizing causal information flow in graph neural networks," in *Proc. AAAI Conf. Artif. Intell.*, vol. 38, no. 15, 2024, pp. 17042–17050.

[69] Y. Zhang, B. Wang, Z. Shan, Z. Zhou, and Y. Wang, "CMT-Net: A mutual transition aware framework for taxicab pick-ups and drop-offs co-prediction," in *Proc. 15th ACM Int. Conf. Web Search Data Mining*, Feb. 2022, pp. 1406–1414.

[70] 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.

[71] Y. Qin, H. Luo, F. Zhao, Y. Fang, X. Tao, and C. Wang, "Spatio-temporal hierarchical MLP network for traffic forecasting," *Inf. Sci.*, vol. 632, pp. 543–554, Jun. 2023.

[72] J. Dai, B. Yang, C. Guo, and Z. Ding, "Personalized route recommendation using big trajectory data," in *Proc. IEEE 31st Int. Conf. Data Eng.*, Seoul, South Korea, Apr. 2015, pp. 543–554.

[73] E. Kanoulas, Y. Du, T. Xia, and D. Zhang, "Finding fastest paths on a road network with speed patterns," in *Proc. 22nd Int. Conf. Data Eng. (ICDE)*, Atlanta, GA, USA, Apr. 2006, p. 10.

[74] L.-Y. Wei, Y. Zheng, and W.-C. Peng, "Constructing popular routes from uncertain trajectories," in *Proc. 18th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, Aug. 2012, pp. 195–203.

[75] X. Li, G. Cong, A. Sun, and Y. Cheng, "Learning travel time distributions with deep generative model," in *Proc. World Wide Web Conf.*, May 2019, pp. 1017–1027.

[76] M.-X. Wang, W.-C. Lee, T.-Y. Fu, and G. Yu, "Learning embeddings of intersections on road networks," in *Proc. 27th ACM SIGSPATIAL Int. Conf. Adv. Geographic Inf. Syst.*, Nov. 2019, pp. 309–318.

[77] N. Wu, X. W. Zhao, J. Wang, and D. Pan, "Learning effective road network representation with hierarchical graph neural networks," in *Proc. 26th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, Aug. 2020, pp. 6–14.

Yudong Zhang (Graduate Student Member, IEEE) received the bachelor's degree from the University of Electronic Science and Technology of China (UESTC) in 2020. He is currently pursuing the Ph.D. degree with the School of Data Science, University of Science and Technology of China (USTC). He has published more than ten research papers in top journals and conferences, such as IEEE TRANSACTIONS ON INTELLIGENT TRANSPORTATION SYSTEMS, IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE, IEEE TRANSACTIONS ON VEHICULAR TECHNOLOGY, SIGKDD, AAAI, and WSDM. His current research interests include spatio-temporal data mining and intelligent transportation systems.

Pengkun Wang (Member, IEEE) received the Ph.D. degree from the University of Science and Technology of China (USTC) in 2023. He is currently an Associate Researcher with Suzhou Institute for Advanced Research, USTC. He has published more than 20 papers on top conferences and journals, such as IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING, AAAI, ICLR, KDD, and WWW. His research interests include open environment data mining and generalization of deep models.

Binwu Wang is currently pursuing the Ph.D. degree with the School of Data Science, University of Science and Technology of China (USTC). He has published several papers on top conferences and journals, such as IEEE TRANSACTIONS ON INTELLIGENT TRANSPORTATION SYSTEMS, IEEE TRANSACTIONS ON MOBILE COMPUTING, KDD, DASFAA, and WSDM. His research interests include traffic data mining and continuous learning, especially their applications in urban computing.

Xu Wang received the bachelor's degree in automation from Northeastern University in 2017 and the Ph.D. degree from the University of Science and Technology of China (USTC) in 2023, under the supervision of Prof. Yang Wang. He is currently an Associate Researcher with USTC. His research interests include spatio-temporal data mining, time series analysis, and AI for science.

Zhe Zhao received the B.E. degree in computer science from Anhui University, Hefei, China, in 2021. He is currently pursuing the joint Ph.D. degree with the University of Science and Technology of China (USTC) and the City University of Hong Kong (CityU). His research interests include machine learning, data mining, and multi-objective optimization.

Zhengyang Zhou (Member, IEEE) received the Ph.D. degree from the University of Science and Technology of China (USTC) in 2023. He is currently an Associate Researcher with USTC. He has published more than 20 papers on top conferences and journals, such as IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING, IEEE TRANSACTIONS ON MOBILE COMPUTING, IEEE TRANSACTIONS ON VEHICULAR TECHNOLOGY, WWW, AAAI, and ICDE. His main research interests include spatio-temporal data mining and urban computing, and he is committed to improving the accuracy, reliability, and generalization of deep spatio-temporal learning models to empower the fields of traffic prediction, urban safety, and pollution control.

Lei Bai (Member, IEEE) received the Ph.D. degree in computer science from UNSW Sydney in 2021. He is currently a Research Scientist with Shanghai AI Laboratory. Prior to that, he was a Post-Doctoral Research Fellow with The University of Sydney, Australia. He has published a set of peer-reviewed papers on top AI conferences and journals, such as NeurIPS, CVPR, IJCAI, KDD, ICCV, Ubicomp, IEEE TRANSACTIONS ON INTELLIGENT TRANSPORTATION SYSTEMS, and IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE. His research interests include spatial–temporal intelligence, resource efficient machine learning, and their applications (e.g., earth system science and intelligent transportation systems). He was a recipient of the 2020 Google Ph.D. Fellowship, the 2020 UNSW Engineering Excellence Award, the 2021 Dean's Award for Outstanding Ph.D. Theses, and the 2022 World Artificial Intelligence Conference (WAIC) Yunfan Award.

Yang Wang (Senior Member, IEEE) received the Ph.D. degree from the University of Science and Technology of China (USTC) in 2007. He is currently an Associate Professor at the School of Computer Science and Technology, School of Software Engineering, and School of Data Science, USTC. Since then, he has continued working at USTC till now as a postdoc and an Associate Professor successively. Meanwhile, he also serves as the Vice Dean of the School of Software Engineering of USTC. His research interests mainly include wireless (sensor) networks, distributed systems, data mining, and machine learning, and he is also interested in all kinds of applications of AI and data mining technologies especially in urban computing and AI4Science.