



FDTI: Fine-Grained Deep Traffic Inference with Roadnet-Enriched Graph

Zhanyu Liu¹, Chumeng Liang¹, Guanjie Zheng¹(✉), and Hua Wei²

¹ Shanghai Jiao Tong University, Shanghai, China
{zhyliu00,gjzheng}@sjtu.edu.cn

² Arizona State University, Tempe, USA
hua.wei@asu.edu

Abstract. This paper proposes the fine-grained traffic prediction task (e.g. interval between data points is 1 min), which is essential to traffic-related downstream applications. Under this setting, traffic flow is highly influenced by traffic signals and the correlation between traffic nodes is dynamic. As a result, the traffic data is non-smooth between nodes, and hard to utilize previous methods which focus on smooth traffic data. To address this problem, we propose **F**ine-grained **D**eep **T**raffic **I**nferece, termed as **FDTI**. Specifically, we construct a fine-grained traffic graph based on traffic signals to model the inter-road relations. Then, a physically-interpretable dynamic mobility convolution module is proposed to capture vehicle moving dynamics controlled by the traffic signals. Furthermore, traffic flow conservation is introduced to accurately infer future volume. Extensive experiments demonstrate that our method achieves state-of-the-art performance and learned traffic dynamics with good properties. To the best of our knowledge, we are the first to conduct the city-level fine-grained traffic prediction.

Keywords: Spatio-Temporal Data · Traffic Forecasting

1 Introduction

Traffic prediction is an important part of an intelligent traffic system and benefits downstream tasks. Some downstream tasks are sensitive to the granularity of prediction results, such as traffic signal control, congestion discovery, and route planning. Taking traffic signal control as an example, predictions on the 1-minute level could timely evaluate the impact of the incoming traffic signal and improve traffic policy because the interval of traffic signal change is approximately 1 min [22]. Previous deep methods [8, 9, 38] focus on the coarse-grained traffic data. However, it remains unexplored that utilizes deep methods to solve traffic prediction tasks under the fine-grained setting.

Under the fine-grained setting, the traffic flow is determined by traffic signals [25]. When the signal turns green, the vehicles could flow into downstream roads. As a result, the correlations between these roads are strong under the traffic prediction context, which is shown in Fig. 1(a). However, previous research

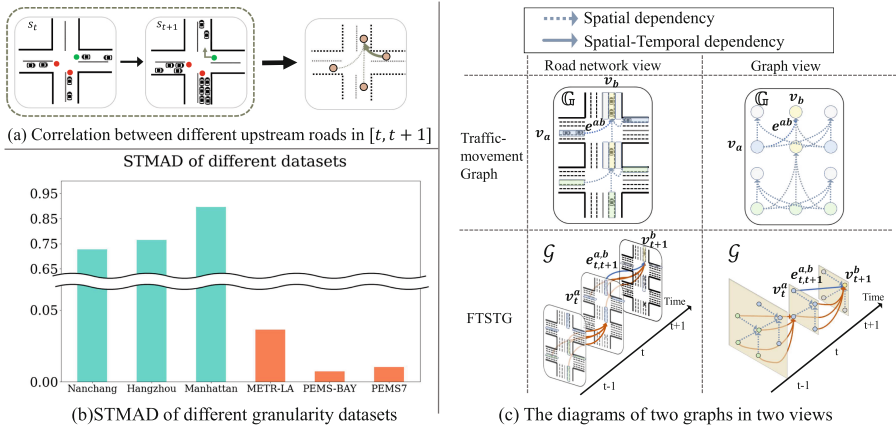


Fig. 1. (a) The traffic signal determines the traffic flow, thereby determining the correlation between roads. (b) Our fine-grained data is much more unsmooth than previous datasets. (Low STMAD indicates smoother and the wavy line indicates the omitted space). (c) Diagrams of Traffic-movement graph and FTSTG.

ignores the explicit highly dynamic correlation between nodes under the fine-grained settings and utilizes static graphs [28, 46] or data-driven graphs [2, 36, 37] to aggregate the knowledge of nodes.

Due to the highly dynamic correlations resulting from the traffic signals, spatial neighbors do not have similar traffic volumes. Therefore, as shown in Fig. 1(b), the fine-grained traffic data is non-smooth, which is evaluated by Spatial Temporal Mean Average Distance (STMAD) defined in this paper. Previous methods have satisfying results on coarse-grained smooth datasets. However, since smoothing is the essential nature of the GCN design [4], experiments show previous methods still make smooth predictions on the nonsmooth fine-grained data which leads to big errors.

To better model the dynamic correlations and tackle the non-smoothness of the data under the fine-grained setting, we propose a model called **F**ine-grained **D**eep **T**raffic **I**nferece (**FDTI**). First, to adapt the characteristic that traffic signal controls traffic flow in fine-grained traffic inference, we construct a **F**ine-grained **T**raffic **S**patial-**T**emporal **G**raph (**FTSTG**). Specifically, we build a road network feature enriched multi-layer traffic graph, in which each layer represents a time frame as shown in Fig. 1(c). Edges inside the graph represent traffic flow links between two nodes at adjacent time frames, which are controlled by traffic signals. Then, we propose a **D**ynamic **M**obility **C**onvolution **N**etwork to induce consistency with the traffic policy on FTSTG. People can make a metaphor between a traffic network and a water flow network, in which the traffic signal is similar to the tap controlling the flow. Based on the previous two modules, we further infer the traffic volume of each node following **F**low **C**onservative **T**raffic **S**tate **T**ransition. Our contribution can be summarized as follows.

- To the best of our knowledge, we are the first to complete the city-level fine-grained traffic prediction, which is important in intelligent traffic systems and will enable efficient and in-time traffic policy-making and other downstream tasks.
- We propose a model named Fine-grained Deep Traffic Inference (FDTI) to incorporate the dynamic spatial temporal dependency caused by traffic signals and then the future traffic is inferred in a flow-conservative perspective.
- Extensive experiments on traffic datasets have shown the superior performance of our proposed method. Graph smoothness analysis is conducted based on our proposed metric STMAD, which explains the mechanism of how other baselines fail under the fine granularity setting.

2 Related Work

Conventional Traffic Prediction. Traffic prediction research draws lots of attention [44], while conventional methods focus on statistical methods. Kalman filter based methods [30, 32] show good results for short-term traffic volume prediction. ML methods such as SVM [31] built on non-linear relationships achieve better performance. The spatial dependency is modeled by methods such as Bayesian Network [49], and probabilistic model [1]. However, they have not exploited the rich spatial information enough.

Deep Spatial-Temporal Traffic Prediction. The utilization of graph convolutional networks (GCNs) [21, 39] contributes significantly to the advancement of spatial-temporal traffic prediction. DCRNN [28], STGCN [46], GSTNet [10], STDN [45], STFGNN [27], LSGCN [19] combines modules such as diffusion, GCN and GRU to model the spatial and temporal relations. Recently many adaptive methods for spatial-temporal data have been proposed. Methodologies such as Graph Wavenet [43], AGCRN [2], GMAN [47], FC-GAGA [33], D2STGNN [37], HGCN [13], ST-WA [8], DSTAGNN [23] utilize techniques such as node embedding and attention to reconstruct the adaptive adjacent matrix and fuse the temporal long term relation. MDTP [12], MTGNN [42], and DMSTGCN [16] utilize multimodal data to help forecast the traffic. Z-GCNET [6] introduces time-aware persistent homology. STGODE [11], STG-NCDE [7], STDEN [20] use differential equations to model the traffic. However, most of those methods would utilize enormous parameters on learning graphs with node embeddings which ignores the influence of traffic signals between different nodes. [34, 35] researches the fine-grained volume inference. However, they focus on spatial-fine-grained grid-based data and utilize CNN-based methods, which can not be applied to our temporal-fine-grained graph-based data. A recent work [25] focuses on fine-grained graph-based traffic prediction, which incorporates a similar state transition function but uses a different setting of missing data.

3 Preliminaries

Definition 1 (Traffic-Movements Grap). We model the traffic system as a traffic-movement graph $\mathbb{G} = (\mathbb{V}, \mathbb{E})$ where \mathbb{V} is the set of N traffic-movements [48] and \mathbb{E} is the set of connections between traffic movements. Each traffic-movement v^i is a set of lanes with the same moving direction $d^i \in \{\text{Left}, \text{Straight}, \text{Right}\}$. Each directed edge e^{ij} denotes the link from traffic movement v^i to traffic movement v^j . Figure 1(c) shows a sample traffic-movement graph \mathbb{G} generated from the real traffic system.

Definition 2 (Traffic State). The traffic state \mathbf{x}_t^i of a traffic-movement v^i at timestamp t includes various measurements such as speed and volume. Thus, the traffic state of the whole system is represented as $\mathbf{X}_t = \{\mathbf{x}_t^1, \mathbf{x}_t^2, \dots, \mathbf{x}_t^N\}$. In this paper, we mainly focus on traffic volume, defined as the number of vehicles on the traffic-movement v^i at timestamp t . The time granularity of the traffic volume is 1 min.

Definition 3 (Roadnet-enriched Feature). Roadnet is an abbreviation for road network. The roadnet-enriched feature indicates the road-network-related feature that helps infer future traffic states. It contains the traffic signal (described as green signal time p_i) and the static information of traffic-movements (e.g., length l^i and direction d^i). Foramally, the system-wise roadnet-enriched feature is represented as $\mathbf{S}_t = \{\mathbf{s}_t^1, \mathbf{s}_t^2, \dots, \mathbf{s}_t^N\}$ where $\mathbf{s}_t^i = \{p_t^i, l^i, d^i\}$ is the features of traffic-movement v^i at time t .

3.1 Problem Definition

Problem 1 (One-step inference). Given a city-level traffic system $\mathbb{G} = (\mathbb{V}, \mathbb{E})$, the goal is to learn a model f to perform traffic inference of next time step \mathbf{X}_{t+1} based on traffic state observations $\mathbf{X}_{t-T+1:t}$ and roadnet-enriched feature $\mathbf{S}_{t-T+1:t}$ of previous T time steps. Formally, the problem is defined as

$$\hat{\mathbf{X}}_{t+1} = f(\mathbf{X}_{t-T+1:t}, \mathbf{S}_{t-T+1:t}). \quad (1)$$

Problem 2 (Q-step inference). Based on one-step state inference, Q-step state inference can be achieved by performing one-step inference Q times. Formally, this problem could be denoted as

$$\hat{\mathbf{X}}_{t'+1} = f(\hat{\mathbf{X}}_{t'-T+1:t'}, \mathbf{S}_{t'-T+1:t'}), t' = t + 1, \dots, t + Q \quad (2)$$

Here $\hat{\mathbf{X}}_{t'-T+1:t'}$ is the input of function f and it could include both predicted value and ground truth value.

4 Method

To solve the defined problem, we propose **F**ine-grained **D**eep **T**raffic **I**nference (**FDTI**) as illustrated in Fig. 2(a). Firstly, traffic states and roadnet-enriched

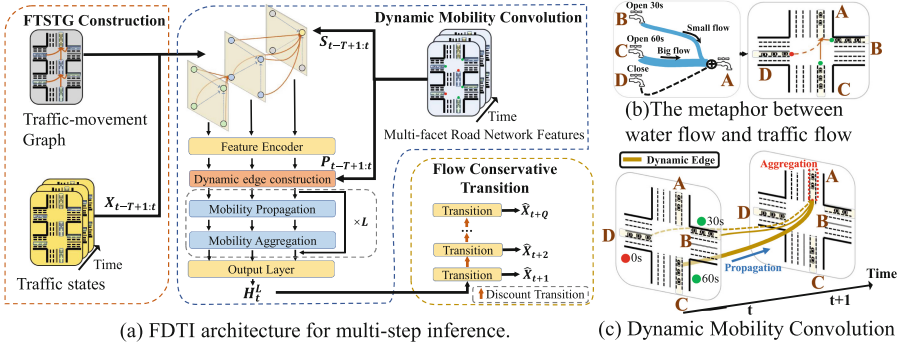


Fig. 2. Diagrams of Fine-grained Deep Traffic Inference (FDTI). (Color figure online)

features are organized to construct FTSTG, which represents the traffic node in a graph with multiple time layers. Then, a dynamic mobility convolution is conducted to model the traffic flow transition via dynamic edges. Lastly, the model predicted the traffic flow, and the future traffic volume of each node is inferred on considering the conservation of the traffic system.

4.1 Fine-Grained Traffic Spatial-Temporal Graph

Graph Construction. In this section, we introduce the construction of Fine-grained Traffic Spatial-Temporal Graph as shown in Fig. 1(c). For the sake of understanding, we make an analogy between a traffic flow network controlled by traffic signals and a water network controlled by taps (as shown in Fig. 2(b)). For the simple water network, the water volume of A (denoted as x_t^A at step t) can be inferred based on the flow-in volume ι_t^A and flow-out volume o_t^A as

$$x_{t+1}^A = x_t^A + \iota_t^A - o_t^A. \tag{3}$$

By considering the spatial dependency, ι_t^A and o_t^A can be calculated with the data B, C , and D , as shown below.

$$\iota_t^A = \sigma(x_t^B, \tau_t^B, x_t^C, \tau_t^C, x_t^D, \tau_t^D) \tag{4}$$

$$o_t^A = \phi(x_t^A, \tau_t^A) \tag{5}$$

where σ and ϕ calculate the flow-in volume and flow-out volume based on x and the turn-on time of the water tap τ .

A traffic system can be represented similarly. Traffic signals can naturally substitute the role of taps in the water network. Equations (3), (4) and (5) show that the states of spatial neighbors of A at timestamp t (x_t^B, x_t^C, x_t^D) are highly related to the state of A at timestamp $t + 1$ (x_{t+1}^A), which inspires us how to construct Fine-grained Traffic Spatial-Temporal Graph (FTSTG). Formally, FTSTG is denoted as $\mathcal{G} = \{\mathcal{V}, \mathcal{E}\}$, where each vertex $v_t^i \in \mathcal{V}$ denotes the node i at timestamp t . Here the size of \mathcal{V} is calculated as $|\mathcal{V}| = N \times T$ where N is

the number of traffic movements and T is the total number of timestamps. We model the spatial-temporal dependency by the edges.

$$\langle v_t^i, v_{t+1}^j \rangle = \begin{cases} 1 & \langle i, j \rangle \in \mathbb{E} \text{ or } i = j \\ 0 & \text{otherwise} \end{cases} \quad (6)$$

where \mathbb{E} is the edge set of the graph \mathbb{G} . (1) We add the edge between spatial neighbors of different time layers. (2) We add edges between the same node of the adjacent time layer. (3) There is no edge inside the same time layer, which is the key difference between FTSTG and STSGCN [38].

The roadnet-enriched features $\mathbf{S}_t = \{P_t, l, d\}$ along with the historical traffic states \mathbf{X}_t serve as the input of each node. There are two reasons why the features \mathbf{S}_t help forecast future traffic. Firstly, The green signal time P_t controls the traffic flow according to Eqs. (4) and (5) and thus significantly influence the future traffic volume. Secondly, the length l and the turning direction d influence the volume distribution of the traffic node since longer roads tend to have more traffic volume, and right-turning lanes tend to have less traffic volume.

4.2 Dynamic Mobility Convolution

To capture the spatial temporal dependencies, we propose Dynamic Mobility Convolution on FTSTG, which is shown in Fig. 2(c). This builds a model that approximates the function σ and ϕ in Eqs. (4) and (5).

Dynamic Edge Construction. To utilize the spatial temporal dependency, a traditional methodology is to apply graph convolution operation on the FTSTG. However, as shown in Fig. 2(b), the traffic flow between different nodes is highly related to the green signal time. Inspired by this fact, we add Dynamic Edge Construction on FTSTG as shown in Fig. 2(c). The dynamic edges are related to the green signal time of each vertex and could represent the traffic flow mobility. A higher weight of dynamic edges indicates higher mobility of traffic flow. Thus, by denoting the edge weight of $\langle v_t^i, v_{t'}^j \rangle$ as $w_{t,t'}^{i,j}$, and the green signal time of v_t^i as p_t^i , we build the edge of FTSTG as follow.

$$w_{t,t'}^{i,j} = \begin{cases} \frac{p_t^i}{t'-t} & \text{if } \langle v_t^i, v_{t'}^j \rangle \in \mathcal{E} \\ 0 & \text{otherwise} \end{cases} \quad (7)$$

Mobility Propagation and Mobility Aggregation. After the Dynamic Edge Construction, we conduct Mobility Propagation and Mobility Aggregation based on the idea of GraphSAGE [15], which is a representative inductive graph learning method. The key idea of Mobility Propagation and Mobility Aggregation is that the hidden states of FTSTG represent the traffic flow and the dynamic edge represents the traffic flow mobility. Then one propagation-aggregation operation layer is simulating the process that vehicles flow into

downstream nodes once, which is also an inductive operation. The output of the l -th layer can be derived as follows.

$$H_{t,i}^l = \text{Agg}(H_{t,i}^{l-1}, \text{Prop}(\{H_{t-1,j}^{l-1}|j\})) \quad (8)$$

Here $j \in \{k | \langle v_{t-1}^k, v_t^i \rangle \in \mathcal{E}\}$, which are the spatial neighbors of i and i itself at previous adjacent timestamp. For Mobility Propagation, we take the dynamic edge into the operation. Formally we can write.

$$\hat{H}_{t,i}^l = f(\{H_{t-1,j}^{l-1} \cdot w_{t-1,t}^{j,i}\}). \quad (9)$$

Then Mobility Aggregation is conducted to aggregate the result of Mobility Propagation and hidden states, which could be formulated as.

$$H_{t,i}^l = g(H_{t,i}^{l-1}, \hat{H}_{t,i}^l). \quad (10)$$

For f and g , multiple functions such as $MEAN(\cdot)$, $POOL(\cdot)$, $Concat(\cdot)$, $FC(\cdot)$ could be chosen. Furthermore, we add residual links [17] between adjacent blocks

A key observation is that one layer of propagation and aggregation feeds all the required input contained in Eqs. (3),(4), and (5) to state x_{t+1}^i . This means the number of layers of propagation and aggregation is equal to the number of historical horizons that are aggregated to $H_{t+1,i}^l$. Typically, the model only needs to consider several adjacent horizons and get good results, which keeps consistent with the fact that only traffic states of adjacent time stamps are useful in the fine-grained traffic inference scenario.

4.3 Flow Conservative Traffic State Transition

Traffic Flow Prediction. The Dynamic Mobility Convolution learns representations \mathbf{H}_t^L that capture the fine-grained spatial temporal dynamics. Based on that, we can predict the flow features, i.e., the out number $\hat{\mathbf{O}}_t$ and in number $\hat{\mathbf{I}}_t$ by using two fully connected layers.

$$\hat{\mathbf{O}}_t = FC(\mathbf{H}_t^L), \hat{\mathbf{I}}_t = FC(\mathbf{H}_t^L) \quad (11)$$

One-Step Traffic Inference. After the out number $\hat{\mathbf{O}}_t$ and in number $\hat{\mathbf{I}}_t$ is predicted, the future traffic could be inferred in a flow conservative perspective. Equation (12) shows the transition from current observation \mathbf{X}_t to the inference of next timestamp $\hat{\mathbf{X}}_{t+1}$ based on the out number $\hat{\mathbf{O}}_t$ and in number $\hat{\mathbf{I}}_t$.

$$\hat{\mathbf{X}}_{t+1} = \mathbf{X}_t + \hat{\mathbf{I}}_t - \hat{\mathbf{O}}_t \quad (12)$$

This shows a flow-conservative perspective for traffic inference. Intuitively, the volume of a node would stay conserved if there are no vehicles driving in or driving out. Hence, by considering each node as a closed traffic system, we only need to focus on the number of the drive-in and drive-out vehicles for future volume inference. This is a key difference between FDTI and other conventional

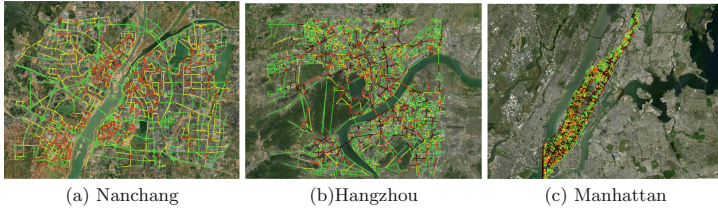


Fig. 3. The running screenshots of the traffic simulator.

approaches to traffic prediction. Conventional approaches focus on capturing the numerical pattern based on mechanisms such as convolution and ignore the conservative traffic state transition which is the intrinsic dynamics.

Multi-step Traffic Inference. For multi-step traffic volume inference, the future multi-faceted features $S_{t+1:t+P}$ is predefined since the traffic signal policy is set in advance. Thus, we can simply apply traffic state transition Eq. (12) multiple times. However, multi-step inference still suffers from error accumulation [3] when a vertex takes inaccurate information from the previous one. Thus, we propose a discounting mechanism to reduce the accumulated error. The discounted multi-step traffic volume inference could be formulated as Eq. (13) where λ denotes the discounting factor.

$$\hat{\mathbf{X}}_{t+Q} = \mathbf{X}_t + \sum_{q=0}^{Q-1} \lambda^q (\hat{\mathbf{I}}_{t+q} - \hat{\mathbf{O}}_{t+q}) \quad (13)$$

We choose MSE loss as the objective function for the one-step flow feature inference to train the model. Thus the loss function of FDTI for flow prediction can be formulated as

$$\mathcal{L} = \frac{1}{NT} \sum_{i=1}^N \sum_{t=1}^T (l_t^i - \hat{l}_t^i)^2 + \frac{1}{NT} \sum_{i=1}^N \sum_{t=1}^T (o_t^i - \hat{o}_t^i)^2. \quad (14)$$

5 Experiment

5.1 Experiment Settings

Datasets. We evaluate our model on three city-wide large-scale datasets of Nanchang, Manhattan, Hangzhou, and one small dataset Hangzhou-Small. Current city traffic data is sparse, coarse-grained, and lacks traffic signal information. Hence, utilizing real roadnet data and vehicle trajectory data as input, we collect 1-h fine-grained data from the wildly-used traffic simulator of KDD-CUP2021 [29]. The roadnet data of these three cities is extracted from OpenStreetMap¹. The vehicle trajectory of Manhattan is processed real data from [40].

¹ <https://www.openstreetmap.org/>.

Table 1. Details of datasets

City	Intersections	Nodes	Connections between nodes
Nanchang	2,048	18,072	73,170
Hangzhou	3,819	32,772	76,788
Manhattan	3,938	40,026	107,442
Hangzhou-Small	211	1,944	4,770

The vehicle trajectories of Hangzhou and Nanchang are from the real information reported by the traffic police. The details of the four datasets are shown in Table 1. The running traffic screenshots of Nanchang, Hangzhou, and Manhattan in the traffic simulator are shown in Fig. 3. Code and data are released in <https://github.com/zhylu00/FDTI>.

Setup of Experiments

- *Data preprocessing*: The first 10 min are used to initialize the road network with sufficient vehicles. The remaining part of the data is split by the ratio of 6:2:2 in chronological order for training, validation, and testing.
- *Network Structure*: In mobility propagation, we use max-pooling as the function f . For the function g in mobility aggregation, we concatenate $H_{t,i}^{l-1}$ and $\hat{H}_{t,i}^l$ and feed them into a fully-connected layer. \tanh is used as the activation function. we set the hidden dimension of graph convolution as 256 and the graph convolution layers L as 4.
- *Training & Evaluating*: The model is optimized by Adam optimizer for at most 500 epochs. The learning rate is set to 0.0005. We evaluate the performance of related models by RMSE and MAPE.

$$RMSE = \sqrt{\frac{1}{s} \sum_{i=1}^s (y_i - \hat{y}_i)^2}, \quad MAPE = \frac{1}{s} \sum_{i=1}^s \left| \frac{y_i - \hat{y}_i}{y_i} \right|$$

Compared Methods. We compare FDTI with the following baselines. For the sake of fairness, all of the baselines except HA and ARIMA take the same node feature ($[v_t^i, p_t^i, l^i, \text{one hot coding for } d^i]$) and all of them are fine-tuned. Four types of baselines are compared.

- *Traditional methods*: **HA** is historical average method, and **ARIMA** [41] is a statistical time series analysis method.
- *Basic Machine Learning models*: **LSTM** [18] is a classic RNN-based model for series analysis. **LR** exploits the linear correlations between data. **XGBoost** [5] is a competitive method based on boosting-tree.
- *Convolution-Kernel-based STGNN*: **DCRNN** [28] and **STGCN** [46] use GCN, GRU, and diffusion techniques to model the spatial and temporal

Table 2. Performance comparison between FDTI and baselines on three large datasets. All adaptive methods can not run on these large-scale datasets due to the huge memory consumption. The lower the RMSE and MAPE are, the better. Horizon means the number of forecasting steps and one horizon means one minute. FDTI achieves the best performance

	Nanchang						Manhattan						Hangzhou					
	Horizon 1		Horizon 3		Horizon 5		Horizon 1		Horizon 3		Horizon 5		Horizon 1		Horizon 3		Horizon 5	
	RMSE	MAPE	RMSE	MAPE	RMSE	MAPE	RMSE	MAPE	RMSE	MAPE	RMSE	MAPE	RMSE	MAPE	RMSE	MAPE	RMSE	MAPE
HA	4.91	23.18	7.79	25.73	10.41	29.97	3.86	13.80	5.45	15.25	6.76	17.96	4.72	21.41	7.47	23.36	10.06	27.21
ARIMA	4.42	24.30	6.90	26.00	9.58	30.77	3.78	14.02	5.26	15.23	6.47	17.95	4.36	22.76	6.62	23.95	9.07	27.75
LSTM	3.71	17.17	5.45	22.40	7.20	28.10	3.68	12.30	5.04	16.81	5.41	18.60	3.55	16.52	5.07	22.00	6.79	27.83
LR	3.19	18.78	5.02	24.70	6.80	30.07	2.68	12.93	4.43	18.56	5.70	22.61	3.24	19.31	4.91	25.45	6.73	32.40
XGBoost	2.85	14.74	4.91	20.70	6.86	26.23	2.53	9.87	4.31	15.17	5.45	19.33	2.94	14.92	4.84	20.95	6.79	27.48
DCRNN	3.98	18.91	5.48	26.42	7.38	31.94	3.92	14.31	5.42	21.48	7.01	29.56	3.85	20.19	5.44	27.27	7.49	35.14
STGCN	11.35	33.35	12.59	37.18	13.71	38.90	8.51	25.40	9.29	27.85	9.94	30.03	10.78	33.30	12.00	36.52	13.22	39.94
STDEN	15.05	37.43	16.11	39.18	17.34	40.37	7.38	25.97	12.26	35.32	15.79	43.21	9.06	28.92	10.14	32.14	11.78	35.01
FDTI	1.30	6.55	4.17	19.18	6.50	25.34	1.20	4.84	3.62	13.63	5.22	17.75	1.46	7.20	4.40	20.44	6.65	25.95

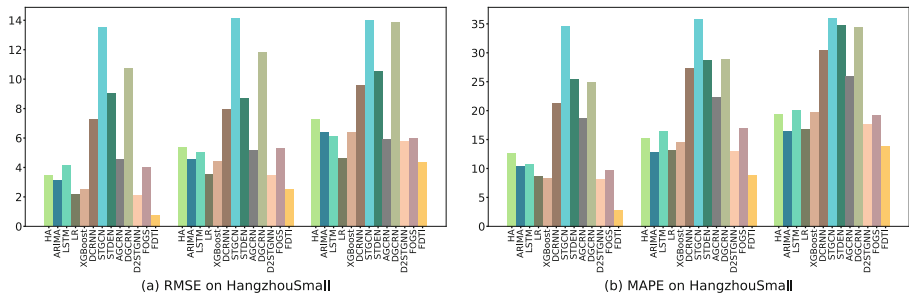


Fig. 4. The performance of different methods w.r.t. RMSE and MAPE on HangzhouSmall under horizon 1, 3 and 5. Horizon means the number of forecasting steps and one horizon means one minute. The lower the RMSE and MAPE are, the better. FDTI achieves the best performance.

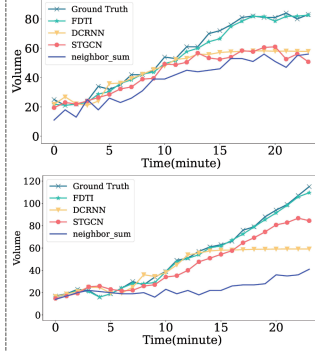
dependencies. **STDEN** [20] is a physics-based ODE method that models the traffic flow.

- *Adaptive-based STGNN*: **AGCRN** [2], **DGCRN** [26], **D2STGNN** [37] focuses on learning the dynamic graph by various methods such as node embeddings and learnable traffic pattern matrix. **FOGS** [36] utilize node2vec-based methods to learn the graph. However, these methods could not run on three large-scale datasets due to the out-of-memory error. They are only evaluated on the HangzhouSmall dataset.

5.2 Overall Performance

The results of the comparison between FDTI and baselines are shown in Table 2 and Fig. 4, where Table 2 shows the performance on three large-scale traffic datasets and Fig. 4 shows the performance on a small traffic dataset. On all four datasets with different scales, our proposed model outperforms all baseline methods in both single-step inference and multi-step inference. The good performance

Nanchang					
GND	0.7274	0.7458	0.7512	0.7506	0.7526
FDTI	0.7251	0.7653	0.7771	0.7830	0.7886
DCRNN	0.0177	0.0179	0.0168	0.0161	0.0157
STGCN	0.0104	0.0127	0.0122	0.0117	0.0114
Hangzhou					
GND	0.7640	0.7886	0.7982	0.8010	0.8047
FDTI	0.5885	0.6472	0.6541	0.6547	0.6589
DCRNN	0.0120	0.0135	0.0138	0.0135	0.0133
STGCN	0.0121	0.0178	0.0186	0.0185	0.0183
Manhattan					
GND	0.8963	0.9335	0.9410	0.9401	0.9412
FDTI	0.6334	0.6962	0.7084	0.7094	0.7124
DCRNN	0.0138	0.0145	0.0133	0.0131	0.0130
STGCN	0.0025	0.0041	0.0045	0.0045	0.0045
	1	2	3	4	5
	Hops				

(a) The k-hop *STMAD* comparison.

(b) Two cases of predictions

Fig. 5. (a) The k-hop *STMAD* comparison between GND (Ground Truth) and the predictions of several methods. (b) Cases of the predicted volume of several methods and ground truth.

indicates that the dynamic correlation modeling and the conservative traffic state inference based on flow-in and flow-out volume help the model grasp the intrinsic pattern of traffic. Note that other deep learning baselines perform worse than the traditional statistic methods and regression-based methods. This indicates that the dynamic correlation between traffic nodes could not be captured by simply stacking convolutional, recurrent, or adaptive mechanisms. Another reason for the bad performance is that these GNN-based methods tend to yield smooth predictions on the nonsmooth dataset. We will discuss the smoothness in detail later. Besides, all of the adaptive methods are not able to run on large-scale datasets for their huge cost. Hence, they are only evaluated on Hangzhou-Small.

5.3 Graph Smooth Analysis

In this part, we explain the reason why previous GCN-based methods yield unsatisfying results by analyzing the smoothness of datasets and prediction results.

To quantitatively measure the smoothness over spatial temporal graphs, we leverage the *STMAD* (Spatial-Temporal Mean Average Distance) based on MAD [4]. The MAD evaluates the smoothness of a given static graph with node features, and lower MAD indicates the graph is smoother. Formally, given a spatial temporal graph in which each node contains a long time series data with length \mathcal{T} , we cut the time series data over a sliding window with length \mathcal{P} . After that, the spatial temporal graph is cut into $\frac{\mathcal{T}}{\mathcal{P}}$ subgraphs. The feature H of each subgraph is the aligned partial time series data with length \mathcal{P} , i.e., $H \in \mathbb{R}^{N \times \mathcal{P}}$. Then, we define the k -hop *STMAD* as follows.

$$STMAD^k = \frac{1}{\mathcal{T}} \sum_{m=1}^{\frac{\mathcal{T}}{\mathcal{P}}} MAD_m^k \quad (15)$$

$$MAD_m^k = \frac{1}{N} \sum_{i=1}^N \frac{\sum_{j \in \mathcal{N}_k(i)} (1 - \frac{H_m^i \cdot H_m^j}{|H_m^i| \cdot |H_m^j|})}{|\mathcal{N}_k(i)|} \quad (16)$$

Here $STMAD^k$ means k -hop $STMAD$ and it is the average of the k -hop MAD of all subgraphs. MAD_m^k is the k -hop MAD of m -th subgraph and it is essentially the average cosine distance between nodes and their k -hop neighbors. The k -hop neighbors set of node i of m -th subgraph and its corresponding time series are denoted as $\mathcal{N}_k(i)$ and H_m^i respectively.

We show the comparison of $STMAD$ between the ground truth data and the prediction result yielded by several methods in Fig. 5(a). Among all the three datasets, We can observe that the $STMAD$ of ground truth is large due to its non-smoothness. The non-smoothness could also be observed in the prediction result of FDTI, indicating that FDTI preserves the original traffic pattern of the ground truth data, thus making accurate predictions. On the contrary, the $STMAD$ of STGCN and DCRNN is much smaller than the $STMAD$ of ground truth data. Furthermore, the $STMAD$ of STGCN and DCRNN is similar to the previous smooth datasets (METR-LA, PEMS-BAY, PEMS7) as shown in Fig. 1(b). This result explains that previous methodologies such as STGCN and DCRNN could yield satisfying results on the previous smooth datasets since these methodologies have a high tendency to make smooth predictions despite the smoothness of the input data. However, when it comes to unsmooth datasets, they make predictions with high errors.

Two examples of the smoothness of STGCN and DCRNN are shown in Fig. 5(b). It shows the ground truth and prediction volume of a node along with the sum volume of its neighbors. We could observe that STGCN and DCRNN make relatively reasonable predictions at the beginning since the ground truth volume is similar to the sum volume of neighbors. As the traffic flow goes on, the gap between Ground Truth and the sum volume of neighbors increases, while STGCN and DCRNN fail to follow the Ground Truth. Being consistent with the fact that smoothing is the essential nature of the GCN design [4], this phenomenon indicates that STGCN and DCRNN tend to average the volume of a node and its neighbors and use the result as the prediction. As a result, the prediction is smooth and a big error exists. On the contrary, FDTI keeps consistent with the ground truth value, which is similar to the $STMAD$ comparison.

To sum up, these two comparisons show FDTI performs admirably in the non-smooth situation. We owe this excellent property to the conservative traffic transitions as shown in Eq. (13). This equation shows that FDTI predicts the first-order derivative of the ground truth and is hence resistant to oversmoothness.

5.4 Ablation Study

For FDTI, there are four main designs including FTSTG that models the traffic dynamics, roadnet-enriched features (denoted as R) that help model capture traffic dynamics, the discount mechanism (denoted as D) that reduces the accumulated error of volume inference, and dynamic mobility convolution (denoted as C)

that simulates the flow of vehicles. To validate these components, we design four variants by adding blocks sequentially: FTSTG, FTSTG+R, FTSTG+R+D, and FTSTG+R+D+C. Specifically, FTSTG+R+D+C equals FDTI because it has all of these four components.

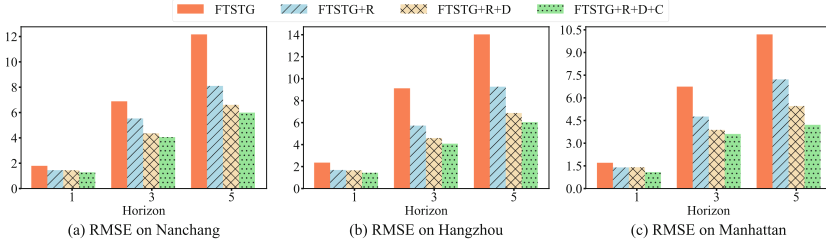


Fig. 6. Performance on the multi-step inference of different variants of FDTI on three datasets. Horizon means the number of forecasting steps.

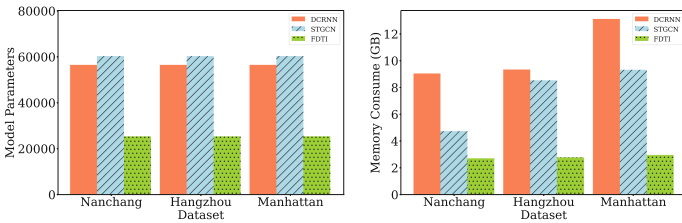


Fig. 7. Model parameter size and memory cost.

Results are shown in Fig. 6 from which we could observe that adding each module can induce further improvement. The improvement induced by adding the roadnet-enriched features (R) is due to that adding traffic-dynamic-related features helps the model aggregate richer information. The performance of multi-step inference is improved by adding the discount mechanism (D), which indicates that the cumulative error could not be neglected and the discount mechanism tackles this error well. Adding dynamic mobility convolution (C) also brings notable performance gain. This demonstrates that considering the dynamic edges contributes to the fine-grained traffic dynamics between nodes.

5.5 Scalability

In this part, we explore the scalability of datasets and models. Then, we explain why previous adaptive methods fail to run on our datasets.

City-scale Datasets and Experiments. To the best of our knowledge, we are the first to complete the city-level traffic state inference. These three city-level traffic datasets Nanchang, Hangzhou, and Manhattan cover more than 2,000 intersections and 18,000 nodes as shown in Table 1. In comparison, we have summarized the datasets used in previous literature as in Table 3. It is easy to observe that our datasets are at least 10 times larger than previously wildly-used datasets in terms of the number of nodes.

Table 3. The scale of datasets. The upper 8 datasets are wildly used by previous research. The lower 4 datasets downside are the datasets of this paper.

Dataset	# of nodes	Dataset	# of nodes
PeMSD7(M) [46]	228	METR-LA [28]	207
PeMSD7(L) [46]	1,026	PEMS-BAY [28]	325
PeMSD4 [14]	307	Xiamen [47]	95
PeMSD8 [14]	170	PeMS3 [38]	358
Nanchang	18,072	Manhattan	40,026
Hangzhou	32,772	Hangzhou-Small	1,944

Model Scalability. The space complexity of FDTI is $O(d \times d)$ and thus the model size is independent of the input graph scale. Furthermore, FDTI is an inductive graph learning method due to the special construction of FTSTG that limited hops of neighbors are required for predicting future traffic. Benefiting from this, our model could deal with large-scale graphs with decent parameter efficiency.

Most of the previous deep spatial temporal methods [13, 24, 33, 36, 37] based on adaptive mechanism fail to run on large-scale datasets. Firstly, they have at least $O(N \times d)$ parameters as node embeddings, which is not parameter-efficient. Secondly, the space complexity of calculating the similarity matrix of these methods is $O(N^2)$, which is unacceptable for a large graph. To validate their performance in the fine-grained setting, we extract Hangzhou-Small dataset and implement some of them as baselines.

For the rest deep learning based methods, we select DCRNN (best performance), STGCN (most efficient), and FDTI (this paper) and compare the model efficiency on the large-scale datasets with the same number of hidden states and layers as shown in Fig. 7. We can observe that FDTI has a similar number of parameters while FDTI consumes much less memory.

6 Conclusion

In this paper, we have worked on a brand-new fine-grained traffic volume prediction problem. We demonstrate that the traffic signal significantly influences the correlation between neighboring roads. To address the problems, We propose a novel method FDTI that models the influence of traffic signal and capture the

fine-grained traffic dynamics. Extensive experiments are conducted on large-scale traffic datasets, where FDTI outperforms other baselines and shows good properties such as resistance to oversmoothness. We believe that FDTI can better support real-world downstream applications such as traffic policy making.

Acknowledgement. This work was sponsored by National Key Research and Development Program of China under Grant No.2022YFB3904204, National Natural Science Foundation of China under Grant No. 62102246, 62272301, and Provincial Key Research and Development Program of Zhejiang under Grant No. 2021C01034.

Ethical Statement. The data used in this paper is collected from the widely-used traffic simulator of KDDCUP2021 and does not contain any personal or sensitive data. The authors ensured that the data was collected in an ethical and legal manner. Hence, no personally identifiable information was obtained and people can not infer personal information through the data. The potential use of this work is accurate traffic prediction and better support of downstream tasks such as traffic signal control. This work is not potentially a part of policing or military work. The authors of this paper are committed to ethical principles and guidelines in conducting research and have taken measures to ensure the integrity and validity of the data. The use of the data in this study is in accordance with ethical standards and is intended to advance knowledge in the field of traffic prediction.

References

1. Akagi, Y., Nishimura, T., Kurashima, T., Toda, H.: A fast and accurate method for estimating people flow from spatiotemporal population data. In: IJCAI, pp. 3293–3300 (2018)
2. Bai, L., Yao, L., Li, C., Wang, X., Wang, C.: Adaptive graph convolutional recurrent network for traffic forecasting. arXiv preprint [arXiv:2007.02842](https://arxiv.org/abs/2007.02842) (2020)
3. Bengio, S., Vinyals, O., Jaitly, N., Shazeer, N.: Scheduled sampling for sequence prediction with recurrent neural networks. arXiv preprint [arXiv:1506.03099](https://arxiv.org/abs/1506.03099) (2015)
4. Chen, D., Lin, Y., Li, W., Li, P., Zhou, J., Sun, X.: Measuring and relieving the over-smoothing problem for graph neural networks from the topological view. In: Proceedings of the AAAI Conference on Artificial Intelligence, vol. 34, pp. 3438–3445 (2020)
5. Chen, T., Guestrin, C.: XGBoost: a scalable tree boosting system. In: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 785–794 (2016)
6. Chen, Y., Segovia, I., Gel, Y.R.: Z-GCNets: time zigzags at graph convolutional networks for time series forecasting. In: International Conference on Machine Learning, pp. 1684–1694. PMLR (2021)
7. Choi, J., Choi, H., Hwang, J., Park, N.: Graph neural controlled differential equations for traffic forecasting. In: Proceedings of the AAAI Conference on Artificial Intelligence, vol. 36, pp. 6367–6374 (2022)
8. Cirstea, R.G., Yang, B., Guo, C., Kieu, T., Pan, S.: Towards spatio-temporal aware traffic time series forecasting. In: 2022 IEEE 38th International Conference on Data Engineering (ICDE), pp. 2900–2913. IEEE (2022)

9. Diao, Z., Wang, X., Zhang, D., Liu, Y., Xie, K., He, S.: Dynamic spatial-temporal graph convolutional neural networks for traffic forecasting. In: Proceedings of the AAAI Conference on Artificial Intelligence, vol. 33, pp. 890–897 (2019)
10. Fang, S., Zhang, Q., Meng, G., Xiang, S., Pan, C.: GSTNet: global spatial-temporal network for traffic flow prediction. In: IJCAI, pp. 2286–2293 (2019)
11. Fang, Z., Long, Q., Song, G., Xie, K.: Spatial-temporal graph ode networks for traffic flow forecasting. In: Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery and Data Mining, pp. 364–373 (2021)
12. Fang, Z., Pan, L., Chen, L., Du, Y., Gao, Y.: MDTP: a multi-source deep traffic prediction framework over spatio-temporal trajectory data. *Proc. VLDB Endow.* **14**(8), 1289–1297 (2021)
13. Guo, K., Hu, Y., Sun, Y., Qian, S., Gao, J., Yin, B.: Hierarchical graph convolution network for traffic forecasting. In: Proceedings of the AAAI Conference on Artificial Intelligence, vol. 35, pp. 151–159 (2021)
14. Guo, S., Lin, Y., Feng, N., Song, C., Wan, H.: Attention based spatial-temporal graph convolutional networks for traffic flow forecasting. In: Proceedings of the AAAI Conference on Artificial Intelligence, vol. 33, pp. 922–929 (2019)
15. Hamilton, W.L., Ying, R., Leskovec, J.: Inductive representation learning on large graphs. In: Proceedings of the 31st International Conference on Neural Information Processing Systems, pp. 1025–1035 (2017)
16. Han, L., Du, B., Sun, L., Fu, Y., Lv, Y., Xiong, H.: Dynamic and multi-faceted spatio-temporal deep learning for traffic speed forecasting. In: Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery and Data Mining, pp. 547–555 (2021)
17. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 770–778 (2016)
18. Hochreiter, S., Schmidhuber, J.: Long short-term memory. *Neural Comput.* **9**(8), 1735–1780 (1997)
19. Huang, R., Huang, C., Liu, Y., Dai, G., Kong, W.: LSGCN: long short-term traffic prediction with graph convolutional networks. In: IJCAI, pp. 2355–2361 (2020)
20. Ji, J., Wang, J., Jiang, Z., Jiang, J., Zhang, H.: STDEN: towards physics-guided neural networks for traffic flow prediction (2022)
21. Kipf, T.N., Welling, M.: Semi-supervised classification with graph convolutional networks. arXiv preprint [arXiv:1609.02907](https://arxiv.org/abs/1609.02907) (2016)
22. Koonce, P., Rodegerdts, L.: Traffic signal timing manual. Technical report, United States. Federal Highway Administration (2008)
23. Lan, S., Ma, Y., Huang, W., Wang, W., Yang, H., Li, P.: DSTAGNN: dynamic spatial-temporal aware graph neural network for traffic flow forecasting. In: International Conference on Machine Learning, pp. 11906–11917. PMLR (2022)
24. Lee, H., Jin, S., Chu, H., Lim, H., Ko, S.: Learning to remember patterns: pattern matching memory networks for traffic forecasting. arXiv preprint [arXiv:2110.10380](https://arxiv.org/abs/2110.10380) (2021)
25. Lei, X., Mei, H., Shi, B., Wei, H.: Modeling network-level traffic flow transitions on sparse data. In: Proceedings of the 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining, pp. 835–845 (2022)
26. Li, F., et al.: Dynamic graph convolutional recurrent network for traffic prediction: benchmark and solution. *ACM Trans. Knowl. Discov. Data (TKDD)* (2021)
27. Li, M., Zhu, Z.: Spatial-temporal fusion graph neural networks for traffic flow forecasting. In: Proceedings of the AAAI Conference on Artificial Intelligence, vol. 35, pp. 4189–4196 (2021)

28. Li, Y., Yu, R., Shahabi, C., Liu, Y.: Diffusion convolutional recurrent neural network: data-driven traffic forecasting. arXiv preprint [arXiv:1707.01926](https://arxiv.org/abs/1707.01926) (2017)
29. Liang, C., et al.: CBLAB: scalable traffic simulation with enriched data supporting. arXiv preprint [arXiv:2210.00896](https://arxiv.org/abs/2210.00896) (2022)
30. Lippi, M., Bertini, M., Frasconi, P.: Short-term traffic flow forecasting: an experimental comparison of time-series analysis and supervised learning. *IEEE Trans. Intell. Transp. Syst.* **14**(2), 871–882 (2013)
31. Nikravesh, A.Y., Ajila, S.A., Lung, C.H., Ding, W.: Mobile network traffic prediction using MLP, MLPWD, and SVM. In: 2016 IEEE International Congress on Big Data (BigData Congress), pp. 402–409. IEEE (2016)
32. Okutani, I., Stephanedes, Y.J.: Dynamic prediction of traffic volume through Kalman filtering theory. *Transport. Res. Part B: Methodol.* **18**(1), 1–11 (1984)
33. Oreshkin, B.N., Amini, A., Coyle, L., Coates, M.: FC-GAGA: fully connected gated graph architecture for spatio-temporal traffic forecasting. In: Proceedings of the AAAI Conference on Artificial Intelligence, vol. 35, pp. 9233–9241 (2021)
34. Ouyang, K., et al.: Fine-grained urban flow inference. *IEEE Trans. Knowl. Data Eng.* **34**(6), 2755–2770 (2020)
35. Qu, H., Gong, Y., Chen, M., Zhang, J., Zheng, Y., Yin, Y.: Forecasting fine-grained urban flows via spatio-temporal contrastive self-supervision. *IEEE Trans. Knowl. Data Eng.* (2022)
36. Rao, X., Wang, H., Zhang, L., Li, J., Shang, S., Han, P.: Fogs: first-order gradient supervision with learning-based graph for traffic flow forecasting. In: Proceedings of International Joint Conference on Artificial Intelligence, IJCAI. ijcai.org (2022)
37. Shao, Z., et al.: Decoupled dynamic spatial-temporal graph neural network for traffic forecasting. arXiv preprint [arXiv:2206.09112](https://arxiv.org/abs/2206.09112) (2022)
38. Song, C., Lin, Y., Guo, S., Wan, H.: Spatial-temporal synchronous graph convolutional networks: a new framework for spatial-temporal network data forecasting. In: Proceedings of the AAAI Conference on Artificial Intelligence, vol. 34, pp. 914–921 (2020)
39. Veličković, P., Cucurull, G., Casanova, A., Romero, A., Lio, P., Bengio, Y.: Graph attention networks. arXiv preprint [arXiv:1710.10903](https://arxiv.org/abs/1710.10903) (2017)
40. Wei, H., Zheng, G., Gayah, V., Li, Z.: A survey on traffic signal control methods. arXiv preprint [arXiv:1904.08117](https://arxiv.org/abs/1904.08117) (2019)
41. Williams, B.M., Hoel, L.A.: Modeling and forecasting vehicular traffic flow as a seasonal Arima process: theoretical basis and empirical results. *J. Transp. Eng.* **129**(6), 664–672 (2003)
42. Wu, Z., Pan, S., Long, G., Jiang, J., Chang, X., Zhang, C.: Connecting the dots: multivariate time series forecasting with graph neural networks. In: Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 753–763 (2020)
43. Wu, Z., Pan, S., Long, G., Jiang, J., Zhang, C.: Graph wavenet for deep spatial-temporal graph modeling. arXiv preprint [arXiv:1906.00121](https://arxiv.org/abs/1906.00121) (2019)
44. Xie, P., Li, T., Liu, J., Du, S., Yang, X., Zhang, J.: Urban flow prediction from spatiotemporal data using machine learning: A survey. *Inf. Fusion* **59**, 1–12 (2020)
45. Yao, H., Tang, X., Wei, H., Zheng, G., Li, Z.: Revisiting spatial-temporal similarity: a deep learning framework for traffic prediction. In: Proceedings of the AAAI Conference on Artificial Intelligence, vol. 33, pp. 5668–5675 (2019)
46. Yu, B., Yin, H., Zhu, Z.: Spatio-temporal graph convolutional networks: a deep learning framework for traffic forecasting. arXiv preprint [arXiv:1709.04875](https://arxiv.org/abs/1709.04875) (2017)

47. Zheng, C., Fan, X., Wang, C., Qi, J.: GMAN: a graph multi-attention network for traffic prediction. In: Proceedings of the AAAI Conference on Artificial Intelligence, vol. 34, pp. 1234–1241 (2020)
48. Zheng, G., et al.: Learning phase competition for traffic signal control. In: Proceedings of the 28th ACM International Conference on Information and Knowledge Management, pp. 1963–1972 (2019)
49. Zhu, Z., Peng, B., Xiong, C., Zhang, L.: Short-term traffic flow prediction with linear conditional gaussian Bayesian network. *J. Adv. Transp.* **50**(6), 1111–1123 (2016)