UCTB: An Urban Computing Tool Box for Spatiotemporal Crowd Flow Prediction

Liyue Chen*
chenliyue2019@gmail.com
Key Lab of High Confidence Software Technologies, Ministry of Education
School of Computer Science, Peking University
Beijing, China

Di Chai*
dchai@cse.ust.hk
Hong Kong University of Science and Technology
Hong Kong, China

Leye Wang†
leyewang@pku.edu.cn
Key Lab of High Confidence Software Technologies, Ministry of Education
School of Computer Science, Peking University
Beijing, China

ABSTRACT

Spatiotemporal crowd flow prediction is one of the key technologies in smart cities. Currently, there are two major pain points that plague related research and practitioners. Firstly, crowd flow is related to multiple domain knowledge factors; however, due to the diversity of application scenarios, it is difficult for subsequent work to make reasonable and comprehensive use of domain knowledge. Secondly, with the development of deep learning technology, the implementation of relevant techniques has become increasingly complex; reproducing advanced models has become a time-consuming and increasingly cumbersome task. To address these issues, we design and implement a spatiotemporal crowd flow prediction toolbox called **UCTB** (Urban Computing Tool Box), which integrates multiple spatiotemporal domain knowledge and state-of-the-art models simultaneously. The relevant code and supporting documents have been open-sourced at https://github.com/uctb/UCTB.

CCS CONCEPTS

• **Information systems → Spatial-temporal systems**.

KEYWORDS

Spatiotemporal prediction, toolbox, crowd mobility

ACM Reference Format:
Liyue Chen, Di Chai, and Leye Wang. 2018. UCTB: An Urban Computing Tool Box for Spatiotemporal Crowd Flow Prediction. In *Proceedings of Make sure to enter the correct conference title from your rights confirmation emai (Conference acronym 'XX)*. ACM, New York, NY, USA, 9 pages. https://doi.org/XXXXXXX.XXXXXXX

*Both authors contributed equally to this research. D. Chai developed the first version of UCTB in 2018-2019. L. Chen has been the leading developer and maintainer of UCTB since 2020.
†Corresponding authors.

Conference acronym 'XX, June 03–05, 2018, Woodstock, NY

ACM ISBN 978-1-4503-XXXX-X/18/06…$15.00
https://doi.org/XXXXXXX.XXXXXXX

1 INTRODUCTION

Crowd flow prediction is a crucial aspect of urban computing, with numerous applications such as urban resource allocation, traffic planning, and safety management [52]. With the advancement of sensor networks, mobile intelligent terminals, and location acquisition technologies, vast amounts of data containing time and geographic information have become available. Such data includes vehicle speed, supply-demand intensity, and pedestrian mobility - collectively known as spatiotemporal data. These rich spatiotemporal datasets provide an excellent foundation for predicting crowd flows. Crowd flow prediction has a wide range of applications, such as scheduling idle vehicles to high-demand areas or predicting peak passenger flows at subway stations. Any application related to human migration and movement within the city can be considered a crowd flow prediction problem. However, due to their nonlinear dependence on various domain knowledge factors, accurate predictions are generally difficult.

Early studies in spatiotemporal crowd flow prediction treat this problem as a classic time series forecasting issue using linear models like ARIMA (AutoRegressive Integrated Moving Average) [12], which integrates moving average self-regressive model or historical mean method (HM). However, linear models like ARIMA and HM cannot model nonlinear mobility patterns. Later on, many nonlinear algorithms are proposed, such as Markov Random Field (MRF) [14], decision tree methods [21], etc.

With the development of deep learning technology, deep neural networks have been widely used in traffic forecasting [26]. RNN (Recurrent Neural Network) and its variants, mainly including LSTM (Long Short-Term Memory) and GRU (Gated Recurrent Network), have been widely used in traffic prediction due to their ability to capture sequence information [45]. However, the above models usually only consider the temporal dependence between predicted values and past values without making good use of spatial dependence.

Crowd flow exhibits spatial dependence, where adjacent stations and those with similar functions tend to have similar flow patterns [46, 47]. Convolutional neural networks (CNNs) have been successful in extracting features from data such as images. In the context of cities, scholars divide a city area into $H \times W$-sized images and use CNNs to capture their spatial dependencies [15, 48–50]. By stacking CNNs deeply through residual units, long-distance dependencies can be captured [48]. Later, GCN (Graph Convolutional Network) is developed for catching spatial dependencies in a more flexible manner. Many spatiotemporal models

based on GCN are applied to the crowd flow prediction problem [1, 2, 6, 8, 10, 11, 13, 16, 19, 25, 27, 29, 38].

The research mentioned above extensively explores the spatiotemporal domain knowledge of crowd flow patterns, which refers to their temporal and spatial dependence. This knowledge can help prediction models better capture crowd flow patterns, but there are many types of spatiotemporal knowledge difficult to fully utilize. While many novel models claim to be 'advanced', reproducing these models is still a challenging and time-consuming task. Although the deep learning community advocates for open-source code, these codes usually only include one or a few types of models scattered across different authors who use different frameworks and data organization methods. Therefore, it's difficult to directly use these codes for fair comparisons.

To address these issues, we develop UCTB (Urban Computing Tool Box), a toolbox designed to provide convenience for researchers and practitioners working on spatiotemporal crowd flow prediction applications. Our work aims to benefit researchers and practitioners by providing an accessible platform that includes various advanced models.

- UCTB incorporates commonly used spatiotemporal knowledge in crowd flow prediction fields and offers a unified data processing interface, making it easy to utilize various types of spatiotemporal prior knowledge. Additionally, the toolbox's unified interface design allows for a direct comparison of various existing spatiotemporal prediction models' performances.
- In addition to integrating classic prediction models, UCTB also includes advanced deep learning models that enable users to quickly apply existing models. The toolbox also provides reusable high-level model layers that accelerate the development of new customized modules.
- To help users get started with UCTB, we provide detailed documentation with tutorial examples. Everyone can access the UCTB toolbox at https://github.com/uctb/UCTB.

2 RELATED WORK

In recent years, we have witnessed many efforts toward building powerful and generalized spatiotemporal prediction models. In these pioneer works, both open-source datasets and traffic models benefit the research community a lot. We list some popular datasets related to crowd flow prediction in Table 1.

At the same time, researchers have implemented and released many open-source prediction models including *ST-ResNet* [48], *DCRNN* [19], *STGCN* [44], GraphWaveNet [38], *GMAN* [51], and *STMeta* [37]. A popular GitHub repository[2] has summarized various open-source deep learning models for traffic prediction. Please be aware that while these works offer open-source codes for traffic prediction, it is crucial to carefully review the documentation and follow the instructions provided in order to understand their specific usage and adapt them to meet target applications. Additionally, the pipelines of these prediction models vary (e.g., feature transformation and normalization techniques), which makes it difficult to make a fair comparison between existing spatiotemporal models.

[2]https://github.com/aptx1231/Traffic-Prediction-Open-Code-Summary

Table 1: Datasets related to crowd flow prediction.

Dataset	Type
Node-based Spatiotemporal Dataset	
METR-LA, PEMS-BAY [19]	Highway Speed
PEMSD3, PEMSD7 [31]	Highway Speed
PEMSD4, PEMSD8 [11]	Highway Speed
PEMS7(L), PEMS7(M) [44]	Highway Speed
Melbourne Pedestrian[1]	Pedestrian Count
Grid-based Spatiotemporal Dataset	
NYCTaxi [42]	Taxi Trip
NYCBike [42]	Bike Trip
TaxiBJ2014, TaxiBJ2015 [48]	Taxi Trip
External Dataset	
Foursquare NYC and Tokyo [40]	Foursquare Check-in
Gowalla [5]	Friendship Network

LibCity is an open-source library that offers a wide range of tools, datasets, and models for urban computing tasks [35, 36]. The primary objective of LibCity is to simplify the application and evaluation of algorithms related to urban data analysis, traffic prediction, and intelligent transportation systems. The toolbox includes various traffic prediction models such as time series models and graph neural networks. These models can be utilized for predicting travel time, traffic congestion, etc.

While UCTB shares many commonalities with LibCity, one key difference is that UCTB focuses on managing and integrating domain knowledge in a unified way for various crowd flow prediction scenarios, no matter which spatiotemporal prediction model is used. In other words, given a specific crowd flow prediction task, rather than directly using existing spatiotemporal prediction models, developers can use UCTB to easily adapt existing spatiotemporal prediction models (e.g., creating spatial graphs useful for the target task, even if these graphs are not used in the model's original paper). Specifically, UCTB provides both spatial and temporal feature transformation interfaces (i.e., *ST_MoveSample* and *GraphGenerator*) which may help developers to flexibly generate useful temporal sequence features and spatial graphs for the target prediction task.

3 UCTB DESIGN AND METHODOLOGY

3.1 Tool Box Overview

The process of executing crowd flow prediction models generally involves three steps: (1) data loading and processing, (2) model construction, and (3) training and evaluation. Accordingly, UCTB consists of three fundamental modules: *data processing*, *model*, and *training & evaluation* (Figure 1).

In the data processing module, a standardized dataset format has been defined that can be applied to various crowd flow prediction applications. Based on this dataset format, we design spatiotemporal feature transformation interfaces to leverage temporal and spatial domain knowledge. In the model module, we decouple some reusable advanced modeling layers and construct prediction models (including both existing methods and newly designed ones) based on them. Finally, in the training and evaluation module, we implement strategies for both model training and evaluation. The

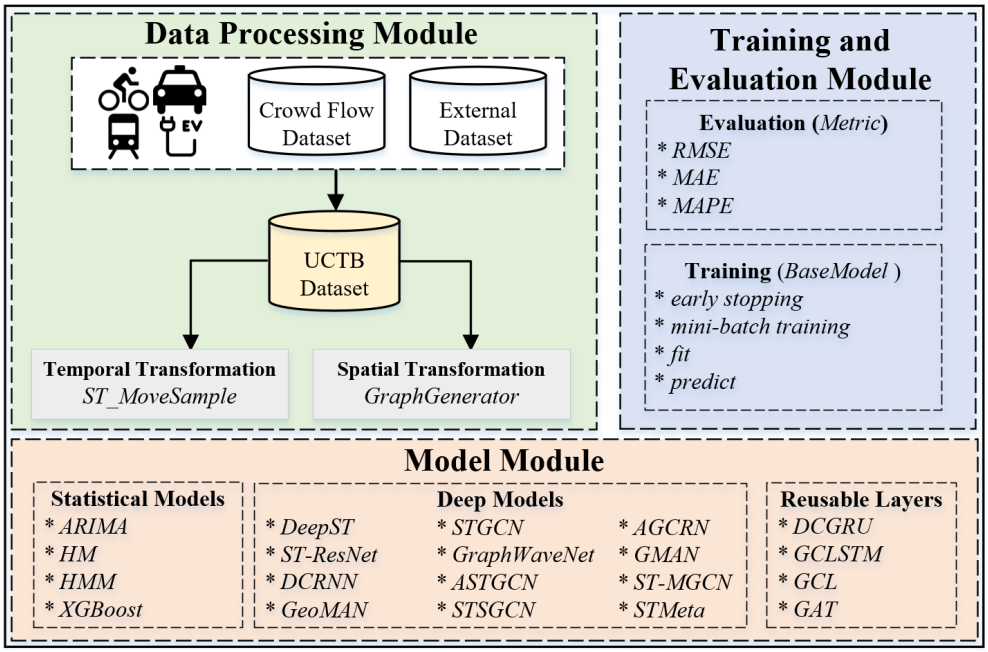


Figure 1: The framework of UCTB

Table 2: UCTB dataset format.

Key	Description
TimeRange	The time range of the dataset
TimeFitness	The time interval of data
Node	2-D (node) crowd flow data
Grid	3-D (grid) crowd flow data
ExternalFeature	Store external features

training strategy will focus on optimizing the model parameters to improve prediction accuracy. The evaluation strategy will assess the model performance based on certain metrics (e.g. MAE, RMSE) using separate test datasets.

3.2 Data Processing Module

UCTB performs feature transformation on raw crowd flow data in two stages. The first stage requires users to convert their application-specific raw data into UCTB's standardized dataset format. In the second stage, UCTB applies spatiotemporal domain knowledge to transform various types of spatiotemporal features from the standardized dataset.

3.2.1 Dataset Format. The standardized dataset format functions as an intermediary between raw data and the interface that UCTB can accept. Users only need to preprocess their raw data into this format before using UCTB for their applications. The UCTB standardized dataset is structured as key-value pairs using the pickle protocol. The primary information for each key-value pair is listed in Table 2.

3.2.2 Feature Transformation. Feature transformation utilizes prior knowledge to convert crowd flow data into various features. This process helps predictive models capture different patterns of crowd flow more effectively.

Temporal knowledge plays a crucial role in this feature transformation process by projecting future traffic values based on past traffic values sampled at different time intervals (refer to Figure 2). There are three common types of temporal knowledge:

(1) *Closeness*: Crowd flows at adjacent moments typically do not experience substantial changes, implying that past traffic values are related to future predicted values.
(2) *Daily Periodicity*: Future traffic values often correspond to the same moment on previous days, reflecting the daily periodicity of human activities.
(3) *Weekly Periodicity*: The flow on Saturday resembles that on the previous Saturday compared to other weekdays.

Spatial knowledge, which reflects the relationship between different locations, has been extensively studied in previous research [37, 49]. In deep prediction models, various techniques are employed to extract spatial features from different types of data. For Euclidean data (e.g., grids), convolutional techniques such as ST-ResNet [48, 50] are typically used to capture spatial dependencies. On the other hand, non-Euclidean data requires constructing different adjacency graphs based on prior knowledge. Graph convolutional techniques [2, 19] such as ChebNet [7] and DCRNN [19] are then applied to extract spatial dependence. Building appropriate graphs may be the most critical step in capturing spatial correlations.

After transforming the features, the data processing module then conducts various operations such as feature normalization

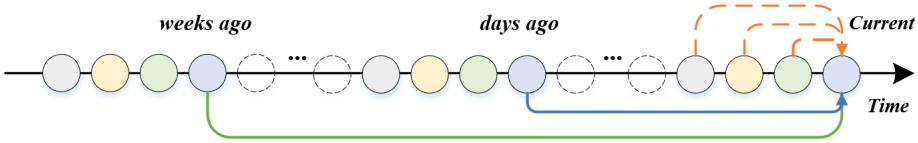


Figure 2: Temporal transformation: generating temporal features by sampling time series data.

and splitting datasets into training, validation and test sets to facilitate subsequent module calls. Apart from spatiotemporal features, external factors like weather conditions affect crowd flow. For example, heavy rain and strong winds can reduce taxi demand [21]. To incorporate external features, UCTB reserves the 'ExternalFeature' key in its standardized dataset format and could expand its capability to utilize these external features.

3.3 Model Module

UCTB integrates two main categories of prediction models, namely statistical learning models and deep learning models. Although these types of models differ in construction and training processes, it is convenient for the user if UCTB encapsulates them and provides similar user interfaces. For instance, all integrated models in UCTB may have internally implemented the 'fit' training method and 'predict' prediction method.

Besides, when implementing deep learning models, some training and prediction interfaces as well as model storage interfaces are similar. To maximize code reuse, UCTB needs to design a basic model class that includes functions such as training, prediction, breakpoint continuation training, etc. Therefore, specific deep learning models only need to inherit this basic class while defining their model structure and setting corresponding feature input functions.

3.4 Training and Evaluation Module

Training statistical learning models is relatively straightforward. When it comes to training deep learning models, it's common practice to divide the data into batches and select small batches for gradient updates during training. This process needs to be integrated into the UCTB toolbox in a reusable way. Additionally, early stopping mechanisms are often used as they help train converged models with fewer epochs.

Once model training and testing are finished, prediction results must undergo appropriate evaluation by comparing actual values with predicted ones. In spatiotemporal crowd flow prediction problems, RMSE (Root Mean Square Error), MAE (Mean Absolute Error), and MAPE (Mean Absolute Percentage Error) are widely used for evaluation purposes. Therefore, corresponding evaluation interfaces need to be implemented by UCTB.

4 IMPLEMENTATION AND INTERFACE

In this section, we will introduce the implementation and the interfaces of UCTB. We mainly focus on three major modules including data processing, model, and training & evaluation.

4.1 Data Processing Interface

We summary the interface of the data processing module in Table 3.

Table 3: Feature Transformation Interface.

Interface	Description
GridTrafficLoader	Load and preprocess grid data
NodeTrafficLoader	Load and preprocess node data
ST_MoveSample	Sample features at different intervals
GraphGenerator	Generate different types of graphs

The *GridTrafficLoader* and *NodeTrafficLoader* interfaces are used to load Euclidean and non-Euclidean data, respectively. These data loaders transform the raw crowd flow data into features based on spatiotemporal prior knowledge. The *ST_MoveSample* interface samples traffic data at different time intervals to obtain temporal features, while the *GraphGenerator* interface generates various spatial graphs to obtain spatial features. Users can also inherit from the *GraphGenerator* interface and implement new graphs for their applications.

ST_MoveSample has three main sampling intervals that represent the number of sampled time features with different interval times. For example, adjacency similarity is 6; daily similarity is 7; weekly similarity is 4. This means that flow at six previous moments before prediction time, along with flow at the same moment during the last seven days and four weeks, are jointly considered as temporal features for prediction.

GraphGenerator receives a graph name and generates its adjacency matrix and Laplacian matrix based on a corresponding threshold value. For instance, for a distance graph, we can set the threshold parameter to 6,500 meters. *GraphGenerator* will generate an adjacency matrix based on Euclidean distances between each station in which stations less than or equal to this threshold are associated (set as 1) while those greater than it are not associated (set as 0). Selecting an appropriate threshold determines whether spatial knowledge can be well extracted or not. Based on our experimental experience, a better threshold generally allows each node to have connections with approximately 20% of other nodes' average connections.

4.2 Model Interface

UCTB offers two types of model interfaces: complete predictive models and reusable model layers that are widely used in the field of crowd flow prediction. The latter enables users to easily create new custom models.

Currently, UCTB has integrated 17 prediction models (Table 4), including 5 statistical learning models (*ARIMA* [12], *HM, XGBoost* [4], etc.) and 12 deep learning models (*ST-ResNet* [48], *DCRNN* [19], *ST-MGCN* [10], *STMeta* [37], etc.). These models have been encapsulated into model classes with internal implementations of the

Table 4: Prediction models in UCTB.

Model	Temporal View	Spatial View
statistical models		
ARIMA [12]	✓	
HM	✓	
GBRT	✓	
HMM	✓	
XGBoost	✓	
deep models		
DeepST [49]	✓	✓
ST-ResNet [48]	✓	✓
DCRNN [19]	✓	✓
GeoMAN [22]	✓	✓
STGCN [44]	✓	✓
GraphWaveNet [38]	✓	✓
ASTGCN [11]	✓	✓
ST-MGCN [10]	✓	✓
GMAN [51]	✓	✓
STSGCN [31]	✓	✓
AGCRN [1]	✓	✓
STMeta [37]	✓	✓

training function ('fit' methods in every model class) and prediction function ('predict' methods in every model class). We also summarize the spatiotemporal domain knowledge utilized by these models in Table 4.

In the statistical learning methods, the historical flow is used as input for the widely-used crowd flow prediction model, *ARIMA*. Historical mean (*HM*) generates predictions by averaging past flows. It considers not only recent crowd flows but also same-time flows from previous days and weeks. Both *GBRT* and *XGBoost* utilize features similar to *HM* that reflect various types of temporal knowledge such as closeness, daily periodicity, and weekly periodicity. In deep learning methods, *DCRNN* utilizes diffusion convolution and RNN to capture spatial and temporal dependencies. *DCRNN* constructs a distance graph reflecting spatial proximity correlation while *ST-MGCN* builds multiple graphs to capture various kinds of spatial dependencies. Finally, *STMeta* is a spatiotemporal knowledge fusion framework focused on leveraging spatiotemporal knowledge and benefiting from developments in spatial or temporal modeling techniques. For details on other methods, readers can refer to the UCTB documentation.[3]

Table 5: Reusable model layers (T: Temporal, S: Spatial).

Interface	Description
DCGRU [19]	Diffusion Convolution (S) + GRU (T)
GCLSTM [2]	ChebNet (S) + LSTM (T)
GCL [7]	ChebNet (S)
GAT [33]	Graph Attention Layers (S)

Table 5 displays the reusable model layers implemented in UCTB. Two spatiotemporal modeling units, *DCGRU* [19] and *GCLSTM*[2], are employed to capture both temporal and spatial features. They substitute matrix product operations inside traditional RNNs with graph convolution operations. ChebNet [7] implements graph convolutional operations grounded in the Chebyshev polynomial. It leverages the Laplacian matrix to discern spatial interdependencies. Graph Attention Layer (GAL) [33] introduces an attention mechanism for graph learning.

[3]https://uctb.github.io/UCTB/md_file/static/current_supported_models.html

Table 6: Training and evaluation interface.

Module	Interface	Description
Training	*MiniBatchFeedDict*	Mini-batch training mechanism
	EarlyStopping	Early stopping mechanism
Evaluation	*RMSE/MAE/MAPE*	Evaluation metrics

4.3 Training and Evaluation Interface

Table 6 outlines the training and evaluation interfaces in UCTB. When applying deep learning models to large training datasets, the full dataset cannot be read into memory simultaneously. UCTB thus employs mini-batch data to update gradients incrementally over multiple epochs. The *MiniBatchFeedDict* interface implements this functionality by continuously invoking the internal *get_batch* method to generate batch data for training.

UCTB implements two early stopping mechanisms: the naive method and the *t-test* approach. The naive method permits several steps without achieving a lower validation set error. In contrast, the *t-test* approach partitions recent validation set errors into two independent samples, each comprising n samples. An independent samples t-test is then conducted. The null hypothesis is that the means of both samples are equivalent. If the p-value of the hypothesis test falls below a threshold (typically 0.1 or 0.05), this hypothesis is rejected, indicating that the model convergence criteria have not yet been met. Otherwise, model convergence is achieved, and training ceases.

In addition, UCTB integrates three commonly used evaluation metrics in crowd flow prediction including RMSE (Roost Mean Square Error), MAE (Mean Absolute Error), and MAPE (Mean Absolute Percentage Error).

5 BENCHMARK EXPERIMENT

Utilizing UCTB, we implement two benchmark experiments evaluating both current spatiotemporal models as well as external context modeling methods. All experimental code has been made openly available. We briefly summarize the primary results and conclusions here; for further details, see [3, 37].

5.1 Spatiotemporal Knowledge Modeling

5.1.1 Motivation. Researchers today are inundated with a plethora of spatiotemporal prediction papers continuously published in top-tier conferences and journals [13, 16–18, 24, 25, 27, 29, 30, 39]. However, most efforts focus on developing sophisticated application-specific models and evaluating them on limited data from one or a few specialized applications, such as ride-sharing, bike-sharing [2, 21], and highway traffic speed [19]. Although these proposed models could theoretically generalize to other spatiotemporal prediction scenarios, their performance in new domains remains unclear. Determining if a crowd flow prediction model can rapidly

Table 7: 30-minute prediction error. The best two results are in bold, and the top one is marked with '*'. (*TC: Temporal Closeness; TM: Multi-Temporal Factors; SP: Spatial Proximity; SM: Multi-Spatial Factors; SD: Data-driven Spatial Knowledge Extraction*)

	Bikesharing			Ridesharing				Metro		EV	Speed		Overall	
	NYC	*CHI*	*DC*	*XA-gr.*	*CD-gr.*	*XA-di.*	*CD-di.*	*SH*	*CQ*	*BJ*	*LA*	*Bay*	*AvgNRMSE*	*WstNRMSE*
Temporal														
HM (TC)	3.206	2.458	2.304	5.280	6.969	19.893	32.098	269.16	221.39	0.768	9.471	4.155	1.865	1.909
ARIMA (TC)	3.178	2.428	2.228	5.035	6.618	19.253	26.131	212.01	180.53	0.755	9.230	3.936	1.667	3.687
LSTM (TC)	3.018	2.493	2.212	4.950	6.444	18.150	23.075	195.60	104.61	0.755	7.866	3.683	1.463	2.596
HM (TM)	2.686	2.230	1.956	4.239	4.851	16.281	17.264	108.59	74.55	0.864	9.560	3.965	1.235	1.523
XGBoost (TM)	2.704	2.376	1.956	4.172	4.915	15.040	16.766	81.82	69.50	0.686	8.298	3.253	1.134	1.420
GBRT (TM)	2.682	2.355	1.928	4.135	4.873	16.202	**14.924***	83.94	72.99	0.689	8.269	3.370	1.139	1.491
TMeta-LSTM-GAL (TM)	2.511	**2.133***	1.927	3.847	4.678	**12.687**	15.324	85.19	53.18	0.686	7.436	3.231	1.047	1.130
Temporal & Spatial														
DCRNN (TC+SP)	2.618	2.246	2.118	4.529	6.258	19.487	22.945	116.15	65.72	0.757	8.562	6.198	1.350	2.051
STGCN (TC+SP)	2.841	2.482	2.067	3.992	5.051	14.139	17.777	91.29	58.34	0.694	7.871	3.136	1.130	1.211
GMAN (TC+SP)	2.792	2.336	**1.836***	4.026	5.293	13.994	20.157	97.58	51.37	0.764	7.276	3.688	1.142	1.351
Graph-WaveNet (TC+SP+SD)	2.666	2.158	1.874	3.986	5.097	13.682	17.170	92.88	52.52	0.719	**6.809***	3.589	1.092	1.232
ST-ResNet (TM+SP)	—	—	—	3.903	4.673	—	—	—	—	—	—	—	—	—
ST-MGCN (TM+SM)	2.513	2.177	1.903	3.886	4.732	13.107	15.404	88.76	50.96	0.691	8.079	**3.042**	1.056	1.186
AGCRN-CDW (TM+SD)	2.830	2.565	2.074	3.958	4.753	16.921	17.982	238.99	131.55	0.688	8.575	**3.022***	1.440	3.171
STMeta-GCL-GAL (TM+SM)	**2.410***	2.170	1.856	**3.808**	4.650	**12.679***	15.307	**75.36***	**49.47**	**0.670**	7.156	3.116	**1.014***	**1.051***
STMeta-GCL-CON (TM+SM)	**2.411**	**2.133***	1.859	**3.772***	**4.613***	12.737	**15.227**	80.69	50.01	**0.667***	**6.889***	3.204	**1.017**	1.071
STMeta-DCG-GAL (TM+SM)	**2.411**	2.182	**1.852**	3.833	**4.635**	12.703	15.398	**77.49**	**48.96***	**0.670**	7.184	3.187	1.019	**1.055**

generalize across various scenarios is challenging. Generalizability is a fundamental model property that substantially impacts its potential scope.

To address this research gap, in our previous work [37], we proposed an analytical framework called *STAnalytic* to evaluate spatiotemporal models based on the high-level spatial and temporal factors they consider. Moreover, we proposed a spatiotemporal meta-model, *STMeta*, with a hierarchical structure that flexibly and efficiently integrates the generalizable temporal and spatial knowledge identified by *STAnalytic*. UCTB can thus be used to assess the generalizability of *STMeta* and other existing models across diverse scenarios.

5.1.2 Results. In Table 7, we present the evaluation results for 30-minute crowd flow prediction. Additional experimental results are available in [37]; dataset and implementation details can be found in the online Appendix[4]. The *avgNRMSE* and *WstNRMSE* metrics were calculated to assess overall performance across diverse applications. For each approach, the temporal and/or spatial factors considered are specified.

According to Table 7, the *STMeta* variants demonstrate superior generalizability across datasets compared to other benchmark approaches. Specifically, in all three experiments with varying time slot durations, *STMeta* consistently ranks first for both *AvgNRMSE* and *WstNRMSE*. This suggests *STMeta* could serve as a crowd flow prediction meta-model for various scenarios. However, the optimal variant depends on the specific application as component implementations can differ. We also evaluated the generalizability of benchmark modeling techniques by comparing those utilizing similar spatial and temporal knowledge. Our findings indicate advanced modeling techniques improve model generalizability to diverse datasets by only approximately 10-20%, far less than improvements from extra temporal and spatial knowledge.

Furthermore, we found that data-driven spatial knowledge modeling (the methods denoted 'SD') lacked consistency across datasets. Extracting spatial knowledge solely from data introduces unique challenges. Therefore, caution is advised when adopting such methods as performance is highly application-dependent.

5.2 External Context Modeling

5.2.1 Motivation. Contextual factors (e.g., weather, holidays) have proven useful as features for various spatiotemporal crowd flow prediction tasks, such as bike-sharing [2, 20, 21, 41] and ride-sharing [15, 28, 32, 34, 43, 53]. For example, warmer temperatures increase bike-sharing usage [21], while heavy rain reduces both bike-sharing and ride-hailing [14]. Prior work has focused on leveraging specific contextual features for certain applications, but the generalizability of contextual features remains understudied. For instance, POIs (Points-Of-Interest) data improves taxi demand prediction [32], but whether POIs data benefits other scenarios is unclear.

Furthermore, while pioneering studies propose various context modeling techniques, such as *Adding* [49], *Embedding* [2], and *Gating* [50], selecting appropriate techniques for a given problem is challenging as their generalizability is unknown (details of these techniques are in Section 5.2.2).

Overall, analyzing the generalizability of contextual features and context modeling techniques holds significant value for developing

[4]https://ieeexplore.ieee.org/document/9627543/

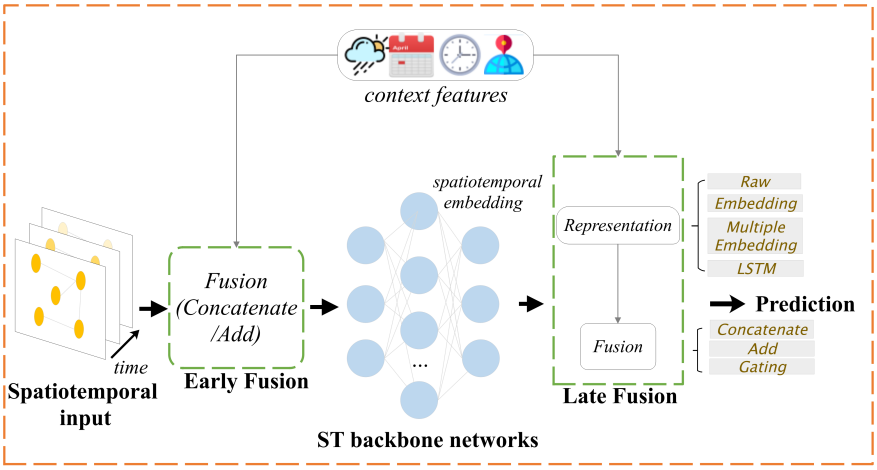


Figure 3: The context modeling analytic framework that introduces two primary categories of context modeling techniques, namely Early Fusion and Late Fusion. In the former, context features are fused with spatiotemporal inputs, while in the latter, context representation is fused with spatiotemporal representation.

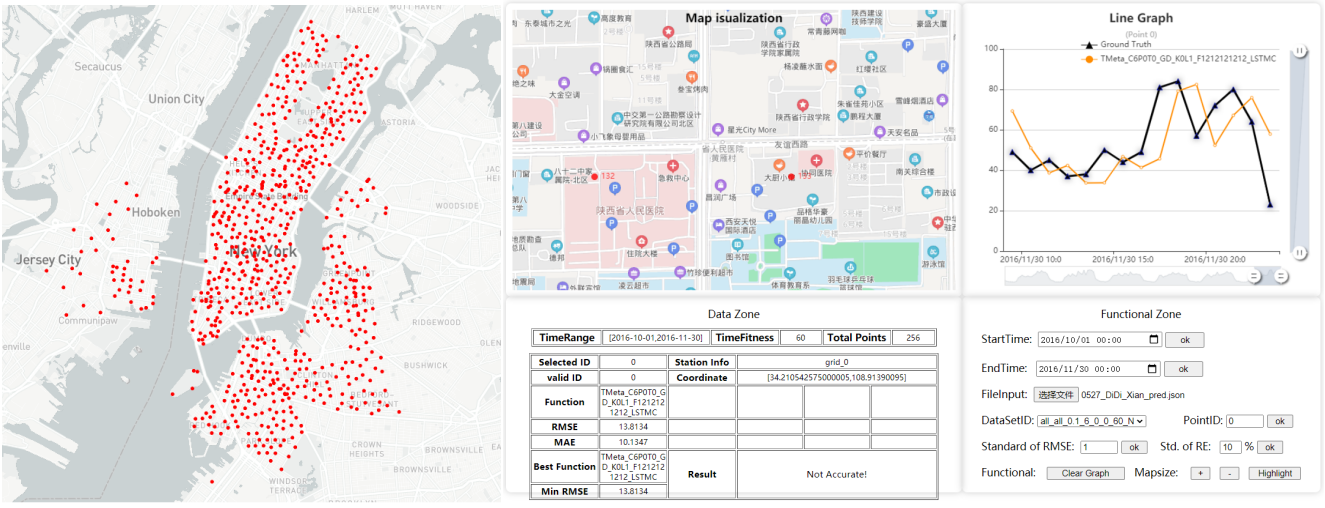
Table 8: 30/60/120-minute RMSE of context modeling techniques. The best results are in bold. *No Context* does not incorporate contextual features. The modeling techniques with better *avgNRMSE* than *No Context* are marked with *.

	Bike			Metro			EV			*avgNRMSE*		
	30	60	120	30	60	120	30	60	120	30	60	120
No Context	2.211	2.740	3.830	77.62	154.5	339.62	0.672	0.818	0.955	1.047	1.058	1.053
Early Fusion												
EarlyConcat	3.173	3.027	4.731	119.5	243.9	421.5	0.967	1.552	1.653	1.541	1.615	1.481
EarlyAdd	2.485	2.703	3.991	78.99	184.5	361.1	0.718	0.786	1.322	1.121	1.109	1.227
Late Fusion												
Raw-Concat	2.229	2.665	3.834	83.51	173.3	544.8	0.658	0.783	0.955	1.069	1.077	1.260
Raw-Add	2.205	2.632	3.797	84.63	162.1	556.3	0.676	0.935	1.045	1.080	1.112	1.302
*Raw-Gating**	2.173	2.598	3.741	**74.40**	**145.5**	334.9	**0.640**	0.783	0.906	**1.010***	**1.004***	1.021*
Emb-Concat	2.199	2.630	3.840	77.35	170.6	375.8	0.662	0.785	0.899	1.039*	1.067	1.069
Emb-Add	2.124	2.701	3.794	80.04	162.6	345.3	0.669	0.788	0.902	1.043*	1.059	1.035*
Emb-Gating	2.189	2.608	3.758	91.76	193.2	381.1	0.656	0.787	0.899	1.099	1.117	1.067
MultiEmb-Concat	2.133	2.593	3.800	78.91	179.6	**330.6**	0.670	0.793	0.929	1.040*	1.086	1.031*
MultiEmb-Add	2.254	2.634	3.776	88.05	175.7	388.1	0.665	**0.778**	0.914	1.097	1.076	1.081
MultiEmb-Gating	2.208	2.690	3.788	85.03	231.9	371.3	0.670	0.780	**0.884**	1.079	1.213	1.054
*LSTM-Concat**	2.116	2.594	3.737	77.23	163.4	350.6	0.665	0.789	0.902	1.027*	1.048*	1.035*
*LSTM-Add**	**2.109**	**2.580**	3.691	76.96	162.8	343.4	0.656	0.787	0.889	1.020*	1.043*	**1.019***
*LSTM-Gating**	2.167	2.585	**3.648**	78.99	156.4	352.4	0.657	0.784	0.893	1.039*	1.028*	1.025*

effective spatiotemporal prediction models. UCTB could also be utilized to benchmark such external context modeling techniques.

5.2.2 Results. In our previous work [3], we introduced a general context modeling analytical framework (Figure 3) by decoupling the context-aware traffic prediction models into spatiotemporal networks and context modeling techniques. The ST networks, applied to capture spatiotemporal dependencies, have been widely investigated [1, 10, 44, 48]. Our framework introduces two primary categories of context modeling techniques, namely Early Fusion (2 variants) and Late Fusion (12 variants) as shown in Table 8.

Early Fusion involves the fusion of context features with spatiotemporal inputs, enabling the spatiotemporal networks to simultaneously capture the dependencies related to flow input and context. The fusion of raw spatiotemporal input and contextual features commonly utilizes the techniques of *Adding* and *Concatenate*



(a) Station visualization

(b) Prediction results visualization

Figure 4: Visualization tools in UCTB

[9, 23, 43]. In contrast, Late Fusion employs separate networks to learn informative context representations during the representation stage. These representations are then further fused with spatiotemporal representations in the fusion stage. The representation stage can utilize techniques such as *LSTM* [15], *Embedding* [2] , and *Multiple Embedding* [34]. As for the fusion stage, techniques such as *Adding*, *Concatenate*, and *Gating* [50] are commonly employed.

Using UCTB, we compared 14 context modeling techniques by leveraging combinations of contextual features, including weather, holidays, temporal position, and POIs. Table 8 shows the results. The *avgNRMSE* metric assessed generalizability; values closer to 1 indicate superior performance across datasets and thus greater generalizability. We found that most modeling techniques did not consistently outperform *No Context*, highlighting the importance of evaluating technique generalizability. Notably, *Raw-Gating* had consistently lower *avgNRMSE* than *No Context* for both *STMGCN* and *STMeta*, demonstrating strong generalizability.

Moreover, we found *Late Fusion* approaches outperformed *Early Joint Modeling* methods. Compared to the *No Context* baseline, *EarlyConcat* and *EarlyAdd* techniques showed inferior performance. These results suggest incorporating contextual features through *Early Joint Modeling* may be suboptimal. Notably, *Late Fusion* methods employing gating mechanisms, such as *Emb-Gating* and *Raw-Gating*, consistently achieved strong performance across applications and models, indicating robustness and generalizability. These findings have significant implications for developing effective spatiotemporal prediction models and can guide future research.

6 VISUALIZATION

To help users better understand datasets and experimental results, we have designed two types of visualization interfaces. The first type is a site visualization interface for displaying the spatial locations of each site. We have integrated this function into *NodeTrafficLoader* and *GridTrafficLoader*. By calling their st_map method, we could get the station visualization as shown in Figure 4(a). Besides, UCTB also provide experimental result visualization interfaces that are used for displaying different model or different station experiment results. We implemented these interfaces using JavaScript and the demo UI pages are shown in Figure 4(b). Currently, we are developing visualization tools 2.0, which mainly focuses on prediction error display and easy human-machine interface. More details are in our repository[5] and our preliminary demo is deployed at http://39.107.116.221/.

7 CONCLUSION

We have developed and released an open-source toolkit, UCTB, to enable cutting-edge model design and optimize leveraging of domain expertise for spatiotemporal forecasting. Through UCTB, we aim to foster sophisticated model development, facilitate comprehensive utilization of domain knowledge, and promote reproducible research in the spatiotemporal forecasting domain. By providing an array of predictive models and tools for knowledge incorporation in an open and well-documented framework, we look to lower barriers to progress and catalyze innovative solutions to predictive challenges. We view UCTB as an evolving community resource and welcome user comments and contributions to help advance, expand, and strengthen the toolkit.

ACKNOWLEDGMENTS

This work was supported by National Science Foundation of China (NSFC) Grant No. 61972008.

REFERENCES

[1] Lei Bai, Lina Yao, Can Li, Xianzhi Wang, and Can Wang. 2020. Adaptive Graph Convolutional Recurrent Network for Traffic Forecasting. In *Proceedings of the 34th International Conference on Neural Information Processing Systems.*

[5] https://github.com/uctb/visualization-tool-UCTB

[2] Di Chai, Leye Wang, and Qiang Yang. 2018. Bike Flow Prediction with Multi-Graph Convolutional Networks. In *Proceedings of the 26th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems*.

[3] Liyue Chen, Xiaoxiang Wang, and Leye Wang. 2021. Exploring the Context Generalizability in Spatiotemporal Crowd Flow Prediction: Benchmark and Guideline. *arXiv e-prints* (2021).

[4] Tianqi Chen and Carlos Guestrin. 2016. XGBoost: A Scalable Tree Boosting System. In *KDD '16*.

[5] Eunjoon Cho, Seth A. Myers, and Jure Leskovec. 2011. Friendship and Mobility: User Movement in Location-Based Social Networks. In *KDD '11*.

[6] Rui Dai, Shenkun Xu, Qian Gu, Chenguang Ji, and Kaikui Liu. 2020. Hybrid Spatio-Temporal Graph Convolutional Network: Improving Traffic Prediction with Navigation Data. In *KDD '20*.

[7] Michaël Defferrard, Xavier Bresson, and Pierre Vandergheynst. 2016. Convolutional Neural Networks on Graphs with Fast Localized Spectral Filtering. In *Proceedings of the 30th International Conference on Neural Information Processing Systems*.

[8] Zheng Fang, Qingqing Long, Guojie Song, and Kunqing Xie. 2021. Spatial-Temporal Graph ODE Networks for Traffic Flow Forecasting. In *KDD '21*.

[9] Jie Feng, Yong Li, Ziqian Lin, Can Rong, Funing Sun, Diansheng Guo, and Depeng Jin. 2021. Context-Aware Spatial-Temporal Neural Network for Citywide Crowd Flow Prediction via Modeling Long-Range Spatial Dependency. *ACM Trans. Knowl. Discov. Data* (2021).

[10] Xu Geng, Yaguang Li, Leye Wang, Lingyu Zhang, Qiang Yang, Jieping Ye, and Yan Liu. 2019. Spatiotemporal Multi-Graph Convolution Network for Ride-Hailing Demand Forecasting. *AAAI* (2019).

[11] Shengnan Guo, Youfang Lin, Ning Feng, Chao Song, and Huaiyu Wan. 2019. Attention Based Spatial-Temporal Graph Convolutional Networks for Traffic Flow Forecasting. *AAAI* (2019).

[12] Mohammad M. Hamed, Hashem R. Al-Masaeid, and Zahi M. Bani Said. 1995. Short-Term Prediction of Traffic Volume in Urban Arterials. *Journal of Transportation Engineering* 121, 3 (1995), 249–254.

[13] Liangzhe Han, Bowen Du, Leilei Sun, Yanjie Fu, Yisheng Lv, and Hui Xiong. 2021. Dynamic and Multi-Faceted Spatio-Temporal Deep Learning for Traffic Speed Forecasting. In *KDD '21*.

[14] Minh X. Hoang, Yu Zheng, and Ambuj K. Singh. 2016. FCCF: Forecasting Citywide Crowd Flows Based on Big Data. In *Proceedings of the 24th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems*.

[15] Jintao Ke, Hongyu Zheng, Hai Yang, and Xiqun (Michael) Chen. 2017. Short-term forecasting of passenger demand under on-demand ride services: A spatio-temporal deep learning approach. *Transportation Research Part C: Emerging Technologies* 85 (2017), 591–608.

[16] Xiaoliang Lei, Hao Mei, Bin Shi, and Hua Wei. 2022. Modeling Network-Level Traffic Flow Transitions on Sparse Data. In *KDD '22*.

[17] He Li, Duo Jin, Xuejiao Li, Jianbin Huang, Xiaoke Ma, Jiangtao Cui, Deshuang Huang, Shaojie Qiao, and Jaesoo Yoo. 2023. DMGF-Net: An Efficient Dynamic Multi-Graph Fusion Network for Traffic Prediction. *ACM Trans. Knowl. Discov. Data* (2023).

[18] Haoran Li, Zhiqiang Lv, Jianbo Li, Zhihao Xu, Yue Wang, Haokai Sun, and Zhaoyu Sheng. 2023. Traffic Flow Forecasting in the COVID-19: A Deep Spatial-Temporal Model Based on Discrete Wavelet Transformation. *ACM Trans. Knowl. Discov. Data* (2023).

[19] Yaguang Li, Rose Yu, Cyrus Shahabi, and Yan Liu. 2018. Diffusion Convolutional Recurrent Neural Network: Data-Driven Traffic Forecasting. In *ICLR '18*.

[20] Yexin Li and Yu Zheng. 2020. Citywide Bike Usage Prediction in a Bike-Sharing System. *IEEE Transactions on Knowledge and Data Engineering* (2020).

[21] Yexin Li, Yu Zheng, Huichu Zhang, and Lei Chen. 2015. Traffic Prediction in a Bike-Sharing System. In *Proceedings of the 23rd SIGSPATIAL International Conference on Advances in Geographic Information Systems*.

[22] Yuxuan Liang, Songyu Ke, Junbo Zhang, Xiuwen Yi, and Yu Zheng. 2018. GeoMAN: Multi-level Attention Networks for Geo-sensory Time Series Prediction. In *IJCAI-18*.

[23] Ziqian Lin, Jie Feng, Ziyang Lu, Yong Li, and Depeng Jin. 2019. DeepSTN+: Context-Aware Spatial-Temporal Neural Network for Crowd Flow Prediction in Metropolis. *AAAI* (2019).

[24] Shuai Ling, Zhe Yu, Shaosheng Cao, Haipeng Zhang, and Simon Hu. 2022. STHAN: Transportation Demand Forecasting with Compound Spatio-Temporal Relationships. *ACM Trans. Knowl. Discov. Data* (2022).

[25] Dachuan Liu, Jin Wang, Shuo Shang, and Peng Han. 2022. MSDR: Multi-Step Dependency Relation Networks for Spatial Temporal Forecasting. In *KDD '22*.

[26] Yisheng Lv, Yanjie Duan, Wenwen Kang, Zhengxi Li, and Fei-Yue Wang. 2015. Traffic Flow Prediction With Big Data: A Deep Learning Approach. *IEEE Transactions on Intelligent Transportation Systems* (2015).

[27] Huiling qin, Xianyuan Zhan, Yuanxun Li, Xiaodu Yang, and Yu Zheng. 2021. Network-Wide Traffic States Imputation Using Self-Interested Coalitional Learning. In *KDD '21*.

[28] Amal Saadallah, Luís Moreira-Matias, Ricardo Sousa, Jihed Khiari, Erik Jenelius, and João Gama. 2020. BRIGHT—Drift-Aware Demand Predictions for Taxi Networks. *IEEE Transactions on Knowledge and Data Engineering* (2020).

[29] Zezhi Shao, Zhao Zhang, Fei Wang, and Yongjun Xu. 2022. Pre-Training Enhanced Spatial-Temporal Graph Neural Network for Multivariate Time Series Forecasting. In *KDD '22*.

[30] Zezhi Shao, Zhao Zhang, Wei Wei, Fei Wang, Yongjun Xu, Xin Cao, and Christian S. Jensen. 2022. Decoupled Dynamic Spatial-Temporal Graph Neural Network for Traffic Forecasting. *Proc. VLDB Endow.* (2022).

[31] Chao Song, Youfang Lin, Shengnan Guo, and Huaiyu Wan. 2020. Spatial-temporal synchronous graph convolutional networks: A new framework for spatial-temporal network data forecasting. In *AAAI 2020*.

[32] Yongxin Tong, Yuqiang Chen, Zimu Zhou, Lei Chen, Jie Wang, Qiang Yang, Jieping Ye, and Weifeng Lv. 2017. The Simpler The Better: A Unified Approach to Predicting Original Taxi Demands based on Large-Scale Online Platforms. In *KDD '17*. 1653–1662.

[33] Petar Veličković, Guillem Cucurull, Arantxa Casanova, Adriana Romero, Pietro Liò, and Yoshua Bengio. 2018. Graph Attention Networks. *ICLR* (2018).

[34] Dong Wang, Wei Cao, Jian Li, and Jieping Ye. 2017. DeepSD: Supply-Demand Prediction for Online Car-Hailing Services Using Deep Neural Networks. In *ICDE 2017*.

[35] Jingyuan Wang, Jiawei Jiang, Wenjun Jiang, Chengkai Han, and Wayne Xin Zhao. 2023. Towards Efficient and Comprehensive Urban Spatial-Temporal Prediction: A Unified Library and Performance Benchmark. *arXiv preprint* (2023).

[36] Jingyuan Wang, Jiawei Jiang, Wenjun Jiang, Chao Li, and Wayne Xin Zhao. 2021. LibCity: An Open Library for Traffic Prediction. In *Proceedings of the 29th International Conference on Advances in Geographic Information Systems*.

[37] Leye Wang, Di Chai, Xuanzhe Liu, Liyue Chen, and Kai Chen. 2021. Exploring the Generalizability of Spatio-Temporal Traffic Prediction: Meta-Modeling and an Analytic Framework. *IEEE Transactions on Knowledge and Data Engineering* (2021).

[38] Zonghan Wu, Shirui Pan, Guodong Long, Jing Jiang, and Chengqi Zhang. 2019. Graph Wavenet for Deep Spatial-Temporal Graph Modeling. In *IJCAI*.

[39] Tong Xia, Junjie Lin, Yong Li, Jie Feng, Pan Hui, Funing Sun, Diansheng Guo, and Depeng Jin. 2021. 3DGCN: 3-Dimensional Dynamic Graph Convolutional Network for Citywide Crowd Flow Prediction. *ACM Trans. Knowl. Discov. Data* (2021).

[40] Dingqi Yang, Daqing Zhang, Vincent. W. Zheng, and Zhiyong Yu. 2015. Modeling User Activity Preference by Leveraging User Spatial Temporal Characteristics in LBSNs. *IEEE Transactions on Systems, Man, and Cybernetics: Systems* (2015).

[41] Zidong Yang, Ji Hu, Yuanchao Shu, Peng Cheng, Jiming Chen, and Thomas Moscibroda. 2016. Mobility Modeling and Prediction in Bike-Sharing Systems. In *MobiSys '16*.

[42] Huaxiu Yao, Xianfeng Tang, Hua Wei, Guanjie Zheng, and Zhenhui Li. 2019. Revisiting Spatial-Temporal Similarity: A Deep Learning Framework for Traffic Prediction. In *AAAI*.

[43] Huaxiu Yao, Fei Wu, Jintao Ke, Xianfeng Tang, Yitian Jia, Siyu Lu, Pinghua Gong, Jieping Ye, and Zhenhui Li. 2018. Deep Multi-View Spatial-Temporal Network for Taxi Demand Prediction. *AAAI* (2018).

[44] Bing Yu, Haoteng Yin, and Zhanxing Zhu. 2018. Spatio-temporal Graph Convolutional Networks: A Deep Learning Framework for Traffic Forecasting. In *IJCAI*.

[45] Rose Yu, Yaguang Li, Cyrus Shahabi, Ugur Demiryurek, and Yan Liu. 2017. Deep Learning: A Generic Approach for Extreme Condition Traffic Forecasting. In *Proceedings of the 2017 SIAM International Conference on Data Mining*.

[46] Jing Yuan, Yu Zheng, and Xing Xie. 2012. Discovering Regions of Different Functions in a City Using Human Mobility and POIs. In *KDD '12*.

[47] Nicholas Jing Yuan, Yu Zheng, Xing Xie, Yingzi Wang, Kai Zheng, and Hui Xiong. 2015. Discovering Urban Functional Zones Using Latent Activity Trajectories. *IEEE Transactions on Knowledge and Data Engineering* (2015).

[48] Junbo Zhang, Yu Zheng, and Dekang Qi. 2017. Deep Spatio-Temporal Residual Networks for Citywide Crowd Flows Prediction. In *AAAI*.

[49] Junbo Zhang, Yu Zheng, Dekang Qi, Ruiyuan Li, and Xiuwen Yi. 2016. DNN-based prediction model for spatio-temporal data. *Proceedings of the 24th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems* (2016).

[50] Junbo Zhang, Yu Zheng, Junkai Sun, and Dekang Qi. 2020. Flow Prediction in Spatio-Temporal Networks Based on Multitask Deep Learning. *IEEE Transactions on Knowledge and Data Engineering* (2020).

[51] Chuanpan Zheng, Xiaoliang Fan, Cheng Wang, and Jianzhong Qi. 2020. GMAN: A Graph Multi-Attention Network for Traffic Prediction. In *AAAI*.

[52] Yu Zheng, Licia Capra, Ouri Wolfson, and Hai Yang. 2014. Urban Computing: Concepts, Methodologies, and Applications. *ACM Trans. Intell. Syst. Technol.* (2014).

[53] L. Zhu and N. Laptev. 2017. Deep and Confident Prediction for Time Series at Uber. In *2017 IEEE International Conference on Data Mining Workshops*.