2023 IEEE 39th International Conference on Data Engineering (ICDE) | 979-8-3503-2227-9/23/$31.00 ©2023 IEEE | DOI: 10.1109/ICDE55515.2023.00086

Extreme-Aware Local-Global Attention for Spatio-Temporal Urban Mobility Learning

Huiqun Huang Suining He* Mahan Tabatabaie
{huiqun.huang, suining.he, mahan.tabatabaie}@uconn.edu
Department of Computer Science and Engineering
University of Connecticut

***Abstract*—The occurrence of special contexts or events (*e.g.,* extreme weather conditions, festival events, other urban anomalies) can significantly influence the movement patterns of urban mobility (*e.g.,* human crowds, transportation systems). Accurate mobility modeling and prediction under the occurrences of such anomaly events is therefore imperative for city management and urban resource allocation. In this study, we propose EALGAP, a novel Extreme-Aware Local-Global Attention urban mobility Prediction model. Specifically, EALGAP models the spatio-temporal *global* and *local* impacts of mobility in different regions and time steps for mobility prediction at various city regions. EALGAP takes into account the *global* impacts by extracting the overall or regular spatial dependencies and temporal patterns of mobility systems for different regions. We have designed a temporally-varying normalization and data-driven technique to quantify the *extreme degrees*, *i.e.,* how significantly the extreme events have impacted the *local* mobility trend, of the patterns within different regions and time steps. We have conducted extensive experimental studies upon four different mobility datasets (over 13 million trips in total) harvested from two metropolitan cities in U.S. with anomalous natural or social events (*e.g.,* hurricane events, other extreme weather conditions, and the Federal holidays). Our results have demonstrated the accuracy, effectiveness, and extreme-awareness of our proposed EALGAP with more than 44.12% error reduction on average compared with other state-of-the-art approaches.**

***Index Terms*—Extreme-aware prediction, local-global impact modeling, extreme degree quantification, urban anomaly, mobility prediction.**

I. Introduction

Urban mobility prediction, *e.g.,* forecasting the arrivals and departures of human crowd mobility or transportation systems, has attracted much attention recently due to the important social and business values [1], [2], [3], [4], [5], [6]. The relevant stakeholders such as the city emergency planners can benefit from the proactive and predictive mobility modeling and enhance their preparedness in essential responsive resource allocations.

Despite the prior studies on conventional deep learning (DL) based mobility prediction [1], [2], [7], [8], [9], [10] thanks to big mobility data, how to cope with and quantify the impacts of various anomaly and extreme events upon the predicted mobility patterns that are rarely observed within the historical data remain to be challenging and largely under-explored. In particular, our studies here will focus on approaching the following two important technical challenges:

- ***Local and Global Impacts on Mobility Patterns.*** As illustrated in Fig. 1, one may observe the occurrences of anomaly events, such as extreme weathers, abnormal traffic conditions, and other urban special events (say, a holiday parade), impact the urban mobility patterns. These events might significantly impact not only the short-term mobility patterns in the local neighborhood at certain city regions, but also pose long-range influence on other city regions' mobility trends in the subsequent time steps. Furthermore, via our further mobility data analytics with the anomaly events, we have observed the *local impacts*, *i.e.,* the instantaneous and *extreme* scales of mobility patterns of different regions incurred by the occurrences of the anomaly events. On the other hand, the *overall trend* dynamics of mobility patterns, as the *global impacts*, may demonstrate spatio-temporally variations across the city regions, due to the interactions of the events with different specific urban function zones (say, points-of-interest or POIs), time of the day, and day of the week. Given the resulting complex mobility patterns (*e.g.,* bike sharing, taxi trips), how to further factorize, quantify, and forecast these mixed impacts with both local and global scales remains challenging and under-explored.
- ***Under-modeling of Heavy-Tailed Mobility Distributions.*** Furthermore, the *extreme degrees* of the impacts from the extreme weather conditions and special urban events, *i.e.,* how the mobility surges or drops deviate from the past, sheds the light upon the under-modeling of extreme mobility patterns in the existing DL-based approaches [1], [2], [7], [8], [9], [10]. We note that the incurred extreme surges or sudden drops of mobility patterns (*e.g.,* the extreme high and low values shown in Fig. 1) can significantly deviate from the historical patterns in various scales, generating the heavy-tailed distributions [11] of the extreme mobility patterns. Many existing DL-based models [7], [8], [9], [10] that often rely on fitting upon the historical major patterns (*e.g.,* major repetitive patterns) might be affected severely, leading to the degraded mobility prediction performance and incorrect decision-making process of subsequent mobile and urban computing applications.

To address the aforementioned challenges, we propose EALGAP, a novel ***E***xtreme-***A***ware ***L***ocal-***G***lobal ***A***ttention mo-

*corresponding author

2375-026X/23/$31.00 ©2023 IEEE
DOI 10.1109/ICDE55515.2023.00086

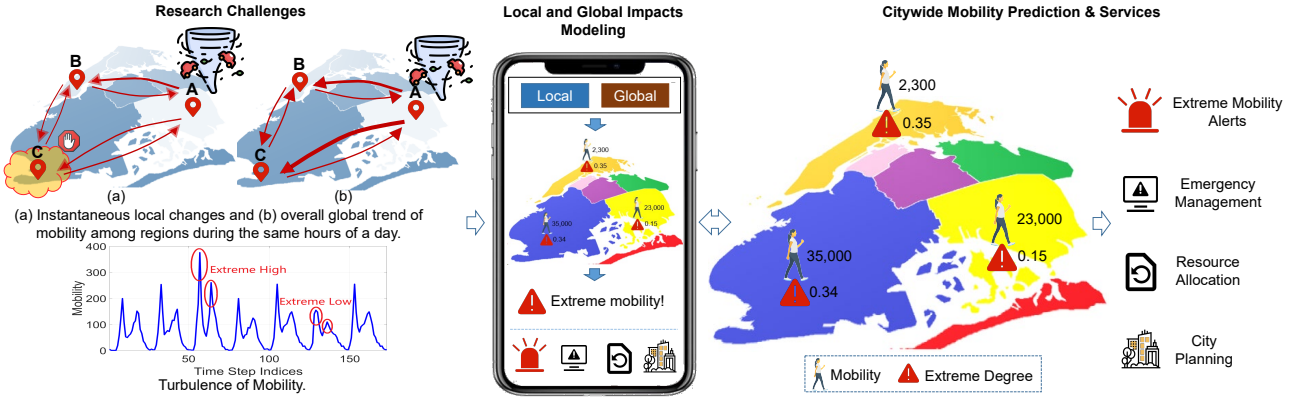


Fig. 1: Research motivations and potential applications of EALGAP.

bility ***P***rediction approach. As illustrated in Fig. 1, taking in the historical mobility system data (such as bike sharing rides or taxi trips) of different regions of a city, our EALGAP models the "local" and "global" impacts within the spatio-temporal dynamics of mobility data, and further quantifies the *extreme degree* of mobility patterns at each region in the target time step and predicts the corresponding mobility patterns. The resulting prediction results can be further utilized by the city emergency management departments and other stakeholders for extreme mobility alerts, emergency resource allocation, and many other city planning applications.

We have made the following three major contributions towards EALGAP:

1) ***Integrating Local-Global Impacts within Urban Mobility Modeling***: We have augmented the adaptivity of EALGAP through local and global impact modeling. Specifically, we have modeled the global impacts across the city regions in different time steps by modeling the statistical distributions (say, exponential distributions in this study) of the mobility patterns of city regions within the recent historical data, and then derive the overall spatio-temporal mobility patterns of each region. In addition, we have taken into account the temporally varying attention parameters for different regions to capture the corresponding temporal patterns for each region. We further model the local impacts of the mobility patterns, *i.e.,* the local fluctuation or turbulence due to the anomaly events (*e.g.,* extreme weather conditions and festival events), of each region in different time steps. The global and local impacts are jointly modeled towards the final mobility pattern prediction of different city regions.
2) ***Extreme-Aware Mobility Degree Formulation***: To further address the heavy-tailed distribution issue and quantify the local impacts of anomaly events, we have provided a data-driven formulation based on the Extreme Degree and Local Impact Modeling Module. In particular, our EALGAP first models the extreme degree of the mobility patterns (*e.g.,* the pick-up volumes of bike sharing) of each city region, *i.e.,* the statistical deviations from the historical time periods. Via capturing the characteristics of the extreme degrees, EALGAP further predicts the impacted mobility patterns by adapting them with the predicted extreme degrees, realizing the extreme-aware mobility prediction.
3) ***Extensive Data-driven Experimental Studies and Evaluations***: We have conducted extensive data analytics and experimental studies on four metropolitan-scale mobility datasets (13,721,726 trips in total from more than 26 months), *i.e.,* the Citi bike sharing and Yellow Taxi datasets in New York City (NYC), NY, and the Divvy bike sharing trips and taxi trips in Chicago, IL, to evaluate the effectiveness of our proposed model in predicting the citywide urban mobility under the occurrences of anomaly events. The experimental results demonstrate that our EALGAP achieves higher accuracy (by more than 44.12% on average) than the other baselines or state-of-the-art approaches (including ST-Norm [7], ST-ResNet [8], EVL [9], and CHAT [10]).

The rest of the paper is organized as follows. We first review the related works in Sec. II, followed by the data analysis, important concepts and motivations, and problem formulation in Sec. III. Afterwards, we present the details of EALGAP's core framework in Sec. V. We demonstrate the experimental results in Sec. VI, and finally conclude in Sec. VII.

II. Related Works

We review our related works in the following two categories.

• ***Spatio-Temporal Mobility Learning.*** Thanks to the big mobility and advances in DL, various spatio-temporal mobility learning approaches have been proposed [12], [13], [14], [15], [16], [17], [18], [19], [20], [21]. Jiang *et al.* [1] designed a multi-task ConvLSTM encoder-decoder framework to capture the deep trend of the crowd density in different regions. Yan *et al.* [22] utilized a 3D convolutional neural network as the core building block to capture spatial temporal dynamics of the mobility. These prior approaches largely consider the temporal dynamics of the data in different regions with the same set of the parameters for all the regions, while the situations that the temporal patterns of mobility at different city regions may

vary differently even in the same time step remain largely overlooked or under-characterized.

Cirstea *et al.* [17] considered the traffic dynamics by modeling the regular distribution of the traffic in different regions. Chen *et al.* [23] utilized the meta learning to generate the parameters of the task network for sequence modeling. Li *et al.* [12] designed the dynamic spatial graph convolution network to model the dynamic spatial correlation to reflect the dynamic change correlation of the road. Zheng *et al.* [24] proposed both the spatial and temporal embeddings to encode the spatial correlation in each time step into a vector. However, despite the incorporation of spatial and temporal dependencies for accurate mobility or traffic prediction, the above studies have not carefully taken into account the distinct overall spatio-temporal patterns, such as the global impacts, of different regions.

Different from these studies, we have further designed a spatio-temporally varying attention mechanism in the Global Impact Modeling Module. Our designs capture the distinct global impacts for different regions using a data-driven approach, *i.e.,* leveraging the distinct parameters generated from the estimated mobility data distribution (say, the exponential distribution in this study) of the mobility of each region within near history. The experimental studies have demonstrated that our approach outperforms the state-of-the-art approaches [7], [8], [9], [10].

• ***Extreme Value Distribution Modeling.*** Irregular spatial and temporal dynamics due to the exogenous factors (such as extreme weather conditions and festival events) often generate the extremely high/low values, leading to the imbalanced distribution issue. How to detect the happening of extreme events has attracted much attention given the growing applications in geography, internet of things, and many other fields [25], [26], [27]. Recent studies have considered setting dynamic thresholds [25], utilizing the spectral residual algorithm [26], and stochastic recurrent neural network [27] to detect the extreme events. However, these studies have not thoroughly considered *predicting* the spatio-temporal mobility patterns *ahead of time* given the complex event impacts.

In order to model the distributions of the extreme values and alleviate their impacts on the prediction models, Ding *et al.* [9] categorized the data into high, normal, and low categories based on the predefined thresholds and proposed a loss function design based on the extreme value theory [28]. Deng *et al.* [7] further modeled the high-frequency and low-frequency components within the time-series data to understand the distributions of normal and long-tailed samples. Huang *et al.* [10] studied a multilayer perceptron-based temporal attention module to capture the relevance of human mobility in different time steps for urban anomaly prediction. Li *et al.* [29] predicted the bike sharing pick-ups/drop-offs from the clustered regions under the anomalous periods by quantifying the weather and time similarity among near history and the target time step, and considering the transition and trip duration among clusters.

Different from the above studies, we have proposed the extreme degree and local impact modeling designs to quantify the extreme dynamics of the mobility incurred by the anomaly events. Specifically, instead of considering the extreme data by simply dividing the data into categories (*e.g.,* high, normal, and low) by predefined thresholds (say, with extreme value theory [9]) or measuring the similarity with the extreme impacts from the historical event/weather data, our `EALGAP` provides a novel method to quantify the extreme degrees of the mobility in different regions and time steps based on the deviations from the historical data. This way, our `EALGAP` incorporates the quantified impacts of the extreme events within the deep learning formulation. Furthermore, our `EALGAP` adaptively parameterizes the spatio-temporal pattern of each region, providing more resilience and flexibility that is essential in characterizing the extreme events (such as extreme weather conditions).

III. Overview of Extreme Mobility Patterns

In this section, we motivate our extreme mobility modeling based on the local-global impact data analysis in Sec. III-A and extreme mobility patterns in Sec. III-B, respectively. In this research study, we leverage the Citi bike sharing and Yellow Taxi datasets in NYC and the Divvy bike sharing trips and taxi trips in Chicago for experimental evaluations (details of the datasets and the processing can be referred to Sec. VI).

A. Local and Global Impacts on Mobility Patterns

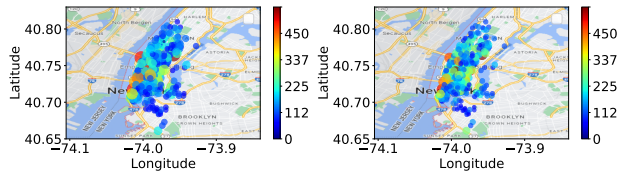
Taking the bike sharing trips in NYC before and during the Hurricane Isaias[1] (the Category 1 hurricane in 2020) as an example, we first illustrate the Citi bike sharing usage on 08/03/2020 (Monday) in NYC right before the tropical storm (Fig. 2a) and during its arrival (Fig. 2b). We can observe that the overall spatio-temporal patterns of the bike pick-up distributions declined with the advent of the hurricane (Fig. 2b), demonstrating the instantaneous impacts of the events upon the overall spatio-temporal patterns of the bike sharing usage.

Furthermore, when we further partition the bike stations (similar to the prior efforts in [30], [29], [31]) into multiple regions (based on k-means clustering and their geographic closeness), we can also see that the bike mobility (pick-ups) from Region A (highlighted in Figs. 3a and 3b) demonstrated more noticeable drop from 10am to 9pm in the face of the Hurricane Isaias, implying different local impacts.

B. Impact of Extreme Mobility Patterns

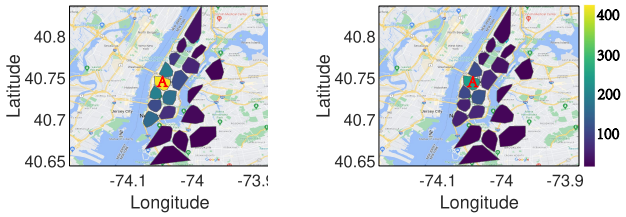
With the partitioned regions (Sec. VI), Fig. 4 further compares the hourly bike pick-ups in the historical average records and during the arrival of the Hurricane Isaias in NYC. Specifically, we find the historical hourly average pick-ups of the region during all weekdays of the previous three months before the hurricane arrived. As illustrated in Fig. 4, there are clearly two bike pick-up peaks within a day in each region due to the commutes of the morning and evening rush hours. However, these peaks might vary across regions in

[1]https://storymaps.arcgis.com/stories/38f75a3c66e44e758a6b094f6b39e1b3



(a) Pick-ups on 08/03/2020. **(b)** Pick-ups on 08/04/2020.

Fig. 2: Illustration of impacts due to the hurricane. Bike sharing pick-ups at all the stations on (a) 08/03/2020 and (b)



(a) Historical weekday aver **(b)** With the Hurricane Isaias.

Fig. 3: Illustration of local impacts due to the hurricane. Bike sharing pick-ups at all the regions on (a) historical weekday average; and (b) 08/04/2020 (with the hurricane event).

terms of start time, duration, and scale. For instance, the bike pick-ups around *W 57th St, Manhattan* (Region 1, Fig. 4a), *West End Ave, Manhattan* (Region 6, Fig. 4b), and *30th Ave, Astoria* (Region 11, Fig. 4c) demonstrate dominant ride surge at 9am on regular weekdays. On the other hand, the morning surge in the *Empire State Building, Manhattan* (Region 20, Fig. 4d) may happen and last from 7am to 10am, likely due to the complex traffic flow and commutes in this business and tourism-intensive area.

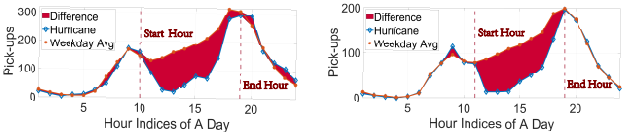
Despite the regular patterns during the morning and evening rush hours, one may observe from Fig. 4 that all the regions experience significant drops of bike usage between the two regular peak hours. We also note that the different regions may have different start and end timestamps for the bike sharing trip drop periods, likely due to their different closeness to the hurricane impacts. In terms of all regions studied, Fig. 5 compares the pick-ups of all regions during the hurricane and those historical daily average records during all weekdays of the three months prior to the tropical storm. We can observe the mobility drops of each region relative to the historical daily average can vary from 19% to 34%.

IV. Concepts and Problem Definitions

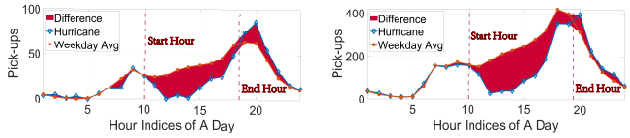
Table I summarizes the important symbols and their definitions presented in this work. We first summarize the important concepts and motivations in Sec. IV-A, followed by the global and local impacts in Sec. IV-B.

A. Problem Formulation

In this study, we divide each day into T time steps, and hence the length of each time step is $24/T$ hours ($T = 24$



(a) W 57th St, Manhattan (Region 1). **(b)** West End Ave, Manhattan (Region 6).



(c) 30th Ave, Astoria (Region 11). **(d)** Empire State Building, Manhattan (Region 20).

Fig. 4: Comparison of historical average hourly bike pick-ups and that during the Hurricane Isaias of four regions (illustrated in Fig. 10).

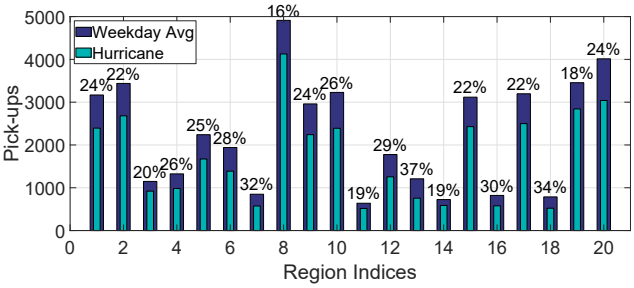


Fig. 5: Comparison of historical daily average and those during the Hurricane Isaias of all regions (illustrated in Fig. 10).

in our current studies). We further parse the mobility system time series data (*e.g.,* bike sharing pick-ups and taxi pick-ups) based on the discretized time steps. Specifically, in each time step t at a region n, we find that the total volume of mobility system (bike sharing or taxi pick-ups) usage as $\mathbf{X}[n, t]$. Then

TABLE I: Important symbols and their definitions.

Symbols	Definitions	Symbols	Definitions
T	Number of the time steps within a day.	L	Length of the time steps of the near history citywide mobility data.
N	Number of the regions.	M	Number of the windows.
$\mathbf{X}[n, l]$	The mobility of region n in time step l.	$\mathbf{F}$	Window of the citywide mobility data.
$\mathbf{X}^g[n, l]$	Global impacts of the mobility of region n in time step l.	λ_n	Rate parameter of the fitted exponential distribution from the region n.
$\mathbf{Z}$	Probability densities of the near history citywide mobility data.	$\mathbf{D}[n, l]$	Extreme degree of the mobility of region n in time step l.
$\widehat{\mathbf{X}}[n, t+1]$	Mobility prediction of region n in time step $(t+1)$.		

we have $\mathbf{X}[n, t-L+1:t] \in \mathbb{R}^L$, whicl patterns of a region $n \in \{1, 2, \ldots, N$ $\{t-L+1, t-L+2, \ldots, t\}$. In addit regions, we will find a total of M slidin slide with one time step), each of whic L consecutive time steps. We will leve to derive the statistical distributions o patterns. Specifically, each time windc is given by $\mathbf{F}_m \in \mathbb{R}^{N\times L} = \mathbf{X}[:, t-$ $t-T(M-m)]$, *i.e.,* the mobility patte during the time periods of $\{t-T(M-$ $m)-L+2, \ldots, t-T(M-m)\}$. Bi can then form a multi-window mobili $[\mathbf{F}_1, \mathbf{F}_2, \ldots, \mathbf{F}_M] \in \mathbb{R}^{M\times N\times L}$ of M our extreme mobility pattern analysis.

The goal of EALGAP is to take in the bike sharing pick-ups, taxi pick-ups) f time steps, $\mathbf{X}[:, t-L+1:t] \in \mathbb{R}^{N\times I}$ window mobility pattern tensor from windows, denoted as $\mathbf{F}$, and forecast th $\mathbb{R}^N$ of each region in the incoming tin

B. Preliminaries of Local and Global

Towards local-global impact modelin the mixture of the spatio-temporally v impacts that can reveal the mobility m Given the mobility pattern $\mathbf{X}[n, l] \in$ time step $l \in \mathbb{R}$, we first consider de major aspects:

$$\begin{aligned}\mathbf{X}[n,l] &= \mathbf{X}^g[n,l] + \mathbf{X}^\lambda[n,\\ &= \mathbf{X}^g[n,l] + \mathbf{X}^g[n,l] \times \mathbf{D}[n,l],\end{aligned} \tag{1}$$

where $\mathbf{X}^g[n, l] \in \mathbb{R}$ measures the global impacts of the mobility of region n in time step l, $\mathbf{D}[n, l]$ represents the extreme degrees, *i.e.,* how the mobility patterns of a region in a specific time step deviate from the overall trends, and $\mathbf{X}^\lambda[n, l] = \mathbf{X}^g[n, l] \times \mathbf{D}[n, l] \in \mathbb{R}$ quantifies the local impacts, *i.e.,* given either positive or negative effect from the extreme events (compared to the regular mobility of this region in the same time step of a day). Towards the prediction of $\widehat{\mathbf{X}}[:, t+1]$, our EALGAP needs to jointly characterize $\mathbf{X}^g[n, l]$ and $\mathbf{X}^\lambda[n, l]$.

V. Core Framework of EALGAP

In this section, we introduce the detailed designs of our proposed model EALGAP as shown in Fig. 6. We first model the global impacts of the near historical mobility from different regions in the Global Impact Modeling Module (Sec. V-A). Then, EALGAP models the extreme degrees of the historical mobility, predicts the extreme degree of each region, and generates the final citywide mobility prediction in the Extreme Degree and Local Impact Modeling Module (Sec. V-B).

A. Global Impact Modeling Module

A-1) *Global Dominant Spatial Dependencies Generation*. We first discuss how to derive the global impacts $X^g[n, l]$

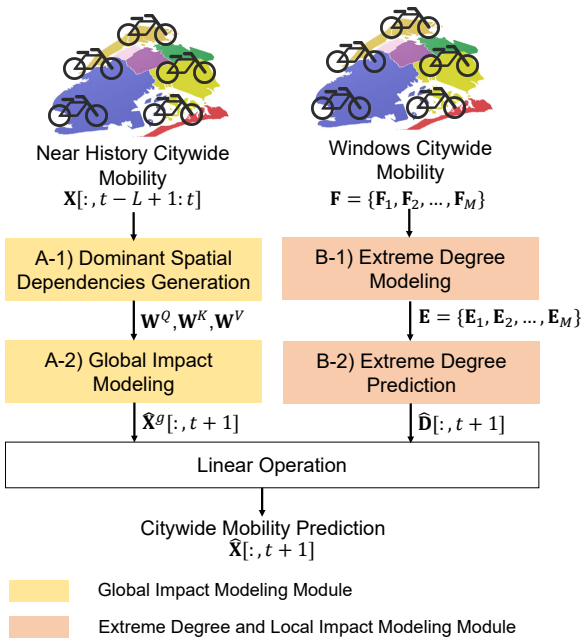


Fig. 6: Architecture overview of EALGAP.

based on the Global Impact Modeling Module in Fig. 8. We have designed a novel data-driven approach with self-attention to generate the adaptive temporally-varying spatial dependencies from the mobility pattern distribution from the recent time steps for each city region.

Specifically, given the mobility data $\mathbf{X}[:, t-L+1:t] \in \mathbb{R}^{N\times L}$ from N regions of L time steps during time periods $\{t-L+1, t-L+2, \ldots, t\}$, we first design the self-attention mechanism to transform the mobility patterns into the matrices of *query* $\mathbf{Q} \in \mathbb{R}^{N\times L\times J}$, *key* $\mathbf{K} \in \mathbb{R}^{N\times L\times J}$, and *value* $\mathbf{V} \in \mathbb{R}^{N\times L\times J}$, respectively, *i.e.,*

$$\begin{aligned}\mathbf{Q} &= \mathbf{X}[:, t-L+1:t] \cdot \mathbf{W}^Q,\\ \mathbf{K} &= \mathbf{X}[:, t-L+1:t] \cdot \mathbf{W}^K,\\ \mathbf{V} &= \mathbf{X}[:, t-L+1:t] \cdot \mathbf{W}^V,\end{aligned} \tag{2}$$

where $\mathbf{W}^Q \in \mathbb{R}^{N\times I\times J}$, $\mathbf{W}^K \in \mathbb{R}^{N\times I\times J}$, and $\mathbf{W}^V \in \mathbb{R}^{N\times I\times J}$ are the learnable parameter matrices for the query, key, and value matrices in the self-attention mechanism. The basic idea of self-attention is to capture the inter-dependencies across the mobility patterns with themselves, yielding learnability in the complex mobility data.

Conventional attention designs [32] often consider static dependencies that characterized by $\mathbf{W}^Q$, $\mathbf{W}^K$, and $\mathbf{W}^V$ across all the time steps once the model is trained. However, in the complex urban mobility prediction, the global impacts upon the mobility patterns may change over time across all the regions. The parameter matrices in the self-attention design may not necessarily hold in the spatio-temporal settings, and require a dynamic and adaptive mechanism to provide flexible mobility prediction.

To overcome such a restriction, we design a data-driven approach within EALGAP to learn the temporally-varying spatial dependencies $\mathbf{W}^Q$, $\mathbf{W}^K$, and $\mathbf{W}^V$ of each region during these L time steps based on the probability density function (PDF) of fitted exponential distributions.

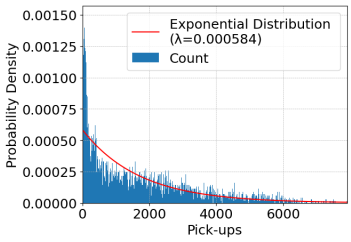


Fig. 7: Actual bike sharing pick-up probability density and the fitted exponential distribution (red curve).

First, we fit the mobility data during time periods $\{t-L+1, t-L+2, \ldots, t\}$ of each region into the exponential distribution. We have conducted empirical studies, and Fig. 7 shows the count distribution and the fitted probability density function (PDF) of the hourly bike sharing pick-ups in NYC of all regions. We can observe that the hourly bike sharing pick-ups reflect the exponential distribution. We have also studied other probability distributions (*e.g.,* normal distribution), and we select the exponential distribution based on our empirical analysis. Then EALGAP returns the rate parameters $\boldsymbol{\lambda} = \{\lambda_1, \lambda_2, \ldots, \lambda_N\} \in \mathbb{R}^N$ for each region based on the maximum likelihood estimations $\mathcal{E}(\cdot)$, *i.e.,* the reciprocal of the average mobility records (say, taxi pick-ups) of each region $n \in \{1, 2, \ldots, N\}$ during $\{t-L+1, t-L+2, \ldots, t\}$,

$$\boldsymbol{\lambda} = \mathcal{E}(\mathbf{X}[:, t-L+1:t]). \tag{3}$$

Second, having obtained the rate parameters $\boldsymbol{\lambda}$ of each region, we then calculate all the probability densities $\mathbf{Z} \in \mathbb{R}^{N \times L}$ of mobility $\mathbf{X}[:, t-L+1:t]$ during time periods $\{t-L+1, t-L+2, \ldots, t\}$ based on the PDF of the exponential distribution, denoted as $\mathcal{F}(\cdot;\cdot)$, by

$$\mathbf{Z} \sim \mathcal{F}(\mathbf{X}[:, t-L+1:t]; \boldsymbol{\lambda}). \tag{4}$$

Based on the nature of the learned PDF, the output $\mathbf{Z}$ represents the mobility dynamics of each region over time. We then utilize three dense or fully-connected layers (FC) interleaved with the Softmax activation functions as a decoder to process $\mathbf{Z}$ into the spatial dependencies of mobility patterns of N regions during $\{t-L+1, t-L+2, \ldots, t\}$ by

$$\left[\mathbf{W}^Q, \mathbf{W}^K, \mathbf{W}^V\right] = \texttt{FC}(\texttt{FC}(\texttt{FC}(\mathbf{Z}))). \tag{5}$$

This way, we do not need to fix the parameter matrices $\mathbf{W}^Q$, $\mathbf{W}^K$, and $\mathbf{W}^V$ for the query, key, and value matrices over all the time steps, and then the output global impact of a region in a specific time step represents the weighted sum of the mobility of the same region in all the recent historical time steps (say, in the most recent L time steps).

A-2) *Modeling Global Impacts via Self-Attention*. We have further designed the self-attention mechanism to model the global impacts. We have the spatial dependencies $\mathbf{W}^Q$, $\mathbf{W}^K$, and $\mathbf{W}^V$ over the time periods of $\{t-L+1, t-L+2, \ldots, t\}$. Here we utilize the self-attention to model the global impacts $\mathbf{X}^g[:, t-L+1:t] \in \mathbb{R}^{N \times L}$ of the mobility patterns of each region over L historical time steps, *i.e.,*

$$\mathbf{X}^g[:, t-L+1:t] = \texttt{Softmax}\left(\frac{\mathbf{Q}\mathbf{K}^\top}{\sqrt{J}}\right)\mathbf{V}, \tag{6}$$

where the query, key, and value are given by Eq. (2). The basic idea is to leverage the dot product between the query $\mathbf{Q}$ and key $\mathbf{K}$ to (both are given by the parameterized mobility patterns in Eq. (2)) to capture the matching and interactions between them, and the resulting importance is further used to weigh the value $\mathbf{V}$ (also parameterized in Eq. (2) based on mobility patterns). We note that Eq. (6) represents the attention score matrix which learns the pairwise similarities of the mobility between any two time steps within the same region. Then the Softmax activation function is applied on the last dimension of the attention score matrix, quantifying the different importance of the mobility in each historical time step of a region, and the importance of all regions sums up to 1. Having the entire score matrix of each region from $\texttt{Softmax}\left(\frac{\mathbf{Q}\mathbf{K}^\top}{\sqrt{J}}\right)$, we will then re-scale the historical mobility of each region by Eq. (6). We set $I = 1$ and $J = 1$ in this study.

Afterwards, we use three fully-connected layers interleaved with the ReLU activation to predict the global impacts $\widehat{\mathbf{X}}^g[:, t+1] \in \mathbb{R}^N$ of N regions in time step $(t+1)$ from the historical global impacts $\widehat{\mathbf{X}}^g[:, t-L+1:t]$ by

$$\widehat{\mathbf{X}}^g[:, t+1] = \texttt{FC}(\texttt{FC}(\texttt{FC}(\mathbf{X}^g[:, t-L+1:t]))). \tag{7}$$

B. Extreme Degree and Local Impact Modeling Module

We further present the Extreme Degree and Local Impact Modeling Module for prediction of the extreme degree $\mathbf{D}[n, l]$ for $\mathbf{X}^\lambda[n, l]$ in Eq. (1), whose designs are illustrated in Fig. 9.

B-1) *Extreme Degree Modeling Based on Historical Same Time Periods*. Normalization is an useful way to ensure that data with very diverse ranges will proportionally impact the model performance, without ignoring the impacts of any long-tailed data [7]. In this study, inspired by the *instance normalization* proposed by Ulyanov *et al.* [33], we consider the instance normalization and evaluate the degree that the mobility of a region in a specific time step varies from the historical data of this region. Given the historical mobility data $\mathbf{X}[n, t-L+1:t] \in \mathbb{R}^L$ of region $n \in \{1, 2, \ldots, N\}$ during time periods $\{t-L+1, t-L+2, \ldots, t\}$, one may consider normalizing the data directly by the mean $\mu_n \in \mathbb{R}$ and the standard deviation $\sigma_n \in \mathbb{R}$ [7] of all the historical data of region n during L time steps, and obtain the extreme degree $\mathbf{D}[n, l] \in \mathbb{R}$:

$$\mathbf{D}[n, l] = \frac{\mathbf{X}[n, l] - \mu_n}{\sqrt{\sigma_n^2 + \varepsilon_n}}, \tag{8}$$

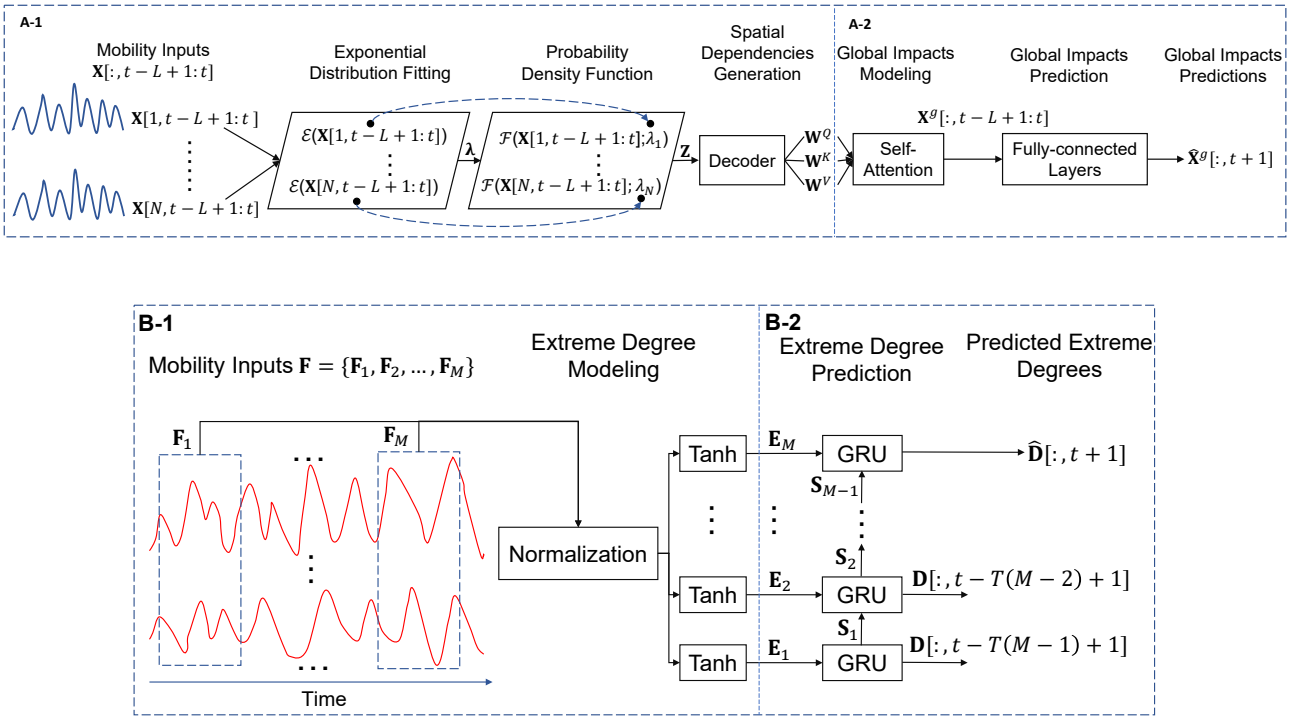


Fig. 9: The detailed designs of the Extreme Degree and Local Impact Modeling Module.

where $\varepsilon_n \in \mathbb{R}$ is set as a learnable constant for a region n. Eq. (8) measures the extreme degree based on the mobility of all the input time steps of the same region, which is applicable to the time series data whose extreme degree does not vary in time. However, for the mobility data, this may cause, for instance, the regular mobility patterns during rush hours of a day to be considered as an *extreme value* if they are compared to the low ones during late night. To overcome such a temporal variation problem, we adapt Eq. (8) by quantifying the extreme degree of mobility of a region based on *the same historical time period* of each region.

Specifically, to evaluate the extreme degree of mobility $\mathbf{X}[n, l]$ of region n in a time step l, we calculate the mean $\mu[n, l] \in \mathbb{R}$ and standard deviation $\sigma[n, l] \in \mathbb{R}$ from $\mathbf{X}[n, l]$ and previous M mobility records of region n, which are in the same time step of day in weekday/weekend as the target time step l. Then, the means and the standard deviations used to quantify the extreme degrees of mobility $\mathbf{X}[n, t-L+1:t]$ vary across different time steps $\{t-L+1, t-L+2, \ldots, t\}$. We then rewrite the Eq. (8) into

$$\mathbf{D}[n, l] = \frac{\mathbf{X}[n, l] - \mu[n, l]}{\sqrt{\sigma^2[n, l] + \varepsilon_n}} \gamma_n, \tag{9}$$

where $\gamma_n \in \mathbb{R}$ is a learnable parameter used to adjust the scales of the extreme degrees of the mobility patterns in each region n. Since we focus on the urban mobility patterns (*e.g.,* bike sharing pick-ups, taxi trips) in this study, $\mathbf{X}[n, l]$ is considered non-negative, and the resulting range of $\mathbf{D}[n, l]$ lies in $[-1, 1]$. Therefore, different $\mathbf{D}[n, l]$ will be considered to introduce varying effects upon the actual trends of mobility patterns within the mobility patterns. After modeling the extreme degree of the mobility patterns $\mathbf{X}[n, t-L+1:t]$, we apply a `tanh` activation function to bound the extreme degree of each region into the range of $[-1, 1]$.

B-2) *Extreme Degree Prediction*. Recall in Sec. IV-A we have defined the multi-window mobility pattern tensor $\mathbf{F}$. To predict the extreme degree $\widehat{\mathbf{D}}[:, t+1] \in \mathbb{R}^N$ of N regions in time step $(t+1)$, we first construct $M \in \mathbb{R}$ windows of mobility data $\mathbf{F} = \{\mathbf{F}_1, \mathbf{F}_2, \ldots, \mathbf{F}_M\} \in \mathbb{R}^{M \times N \times L}$ from N regions. The length of the time steps of each window is also L time steps. We note that $\mathbf{F}_m = \mathbf{X}[:, t-T(M-m)-L+1 : t-T(M-m)] \in \mathbb{R}^{N \times L}$ represents the mobility data of window $m \in \{1, 2, \ldots, M\}$ from the N regions during time periods $\{t-T(M-m)-L+1, t-T(M-m)-L+2, \ldots, t-T(M-m)\}$. For instance, if we set $T = 24$ (24 hours backward) and $L = M = 2$ for prediction of 8:00–9:00am, Thursday (a weekday), our $\mathbf{F}_1$ and $\mathbf{F}_2$ respectively correspond the mobility patterns of all regions on the Tuesday and Wednesday right before the target date.

Then, we evaluate the extreme degrees for M windows of mobility data $\mathbf{F}$ by Eq. (9) and a `tanh` activation after it, and represent the extreme degrees as $\mathbf{E} = \{\mathbf{E}_1, \mathbf{E}_2, \ldots, \mathbf{E}_M\} \in \mathbb{R}^{M \times N \times L}$, where $\mathbf{E}_m = \mathbf{D}[:, t-T(M-m)-L+1 : t-T(M-m)] \in \mathbb{R}^{N \times L}$. We then apply a Gated Recurrent Units (`GRU`) module [34] to embed each window of extreme degree into the feature space, and predict the extreme degree of N regions in the one following time step of each window. In each `GRU` module, we use a `tanh` activation function. The output

states from GRU of a window are utilized as the initial states of the GRU when operating the data of the next window. The embedding and the subsequent prediction are then formulated as

$$[\mathbf{D}[:, t - T(M-m)+1], \mathbf{S}_m] = \texttt{GRU}(\mathbf{E}_m, \mathbf{S}_{m-1}), \quad (10)$$

where $\mathbf{D}[:, t - T(M-m)+1] \in \mathbb{R}^N$ and $\mathbf{S}_m \in \mathbb{R}^N$ are respectively the predicted extreme degrees and hidden states when operating the extreme degree data $\mathbf{E}_m \in \mathbb{R}^{N \times L}$ of a window m. $\mathbf{S}_{m-1} \in \mathbb{R}^N$ represents the hidden states of GRU from the $(m-1)^{\text{th}}$ window. After embedding all windows of extreme degree data, we denote the final extreme degrees prediction of N regions at the time step $(t+1)$ as $\widehat{\mathbf{D}}[:, t+1] \in \mathbb{R}^N$.

• **Final Mobility Prediction Output.** Finally, given the global impact prediction $\widehat{\mathbf{X}}^g[:, t+1] \in \mathbb{R}^N$, and the extreme degree prediction $\widehat{\mathbf{D}}[:, t+1] \in \mathbb{R}^N$, we consider both the global impacts and extreme degrees with Eq. (1) to predict the final mobility pattern of N regions in time step t, $\widehat{\mathbf{X}}[:, t+1] =$

$$\texttt{ReLU}(\widehat{\mathbf{X}}^g[:, t+1] + \widehat{\mathbf{X}}^g[:, t+1] \times \widehat{\mathbf{D}}[:, t+1]). \quad (11)$$

VI. Experimental Studies

In this section, we first present the details of the datasets in Sec. VI-A, followed by the baseline models and the experimental settings in Sec. VI-B. After that, we present the experimental results of this study in Sec. VI-C.

A. Details of Mobility Data Studied

• **Bike Mobility Data:** We consider the Citi bike sharing data in NYC, NY (10,064,558 trips in total from 347 bike stations are studied)[2] during the selected time periods (about 7.5 months) in 2020. In addition, we have also included the Divvy bike sharing in Chicago (1,281,745 trips in total from 799 bike stations are studied)[3] during selected time periods (about 7 months) in 2021.

• **Taxi Mobility Data:** We also utilize the NYC Yellow Taxi data (2,359,903 trips in total) during selected time periods (about 5 months) in 2016 and the Chicago taxi data (15,520 trips in total)[4] during selected time periods (about 7 months) in 2021. It also serves as another angle to look at the regional connectivity and the impact of the events.

B. Baselines & Experimental Settings

• **Baseline Models:** In this study, we implement the following models as the baselines for comparison.

1) *Auto-Regressive Integrated Moving Average* (ARIMA): We utilize ARIMA to predict the urban mobility of N regions in the time step $(t+1)$. We implement ARIMA to model the nonseasonal patterns within the time series of mobility.
2) GRU/LSTM/RNN: We utilize the Gated Recurrent Unit (GRU), Long Short-Term Memory (LSTM) neural network, and Recurrent Neural Networks (RNN) to predict the mobility patterns of N regions in time step $(t+1)$ from the near history mobility data $\mathbf{X}[:, t-L+1:t] \in \mathbb{R}^{N \times L}$.
3) ST-Norm [7]: which implements the Spatial and Temporal Normalization-based (ST-Norm) framework. Here ST-Norm models the high- and low-frequency components within the mobility patterns.
4) ST-ResNet [8]: ST-ResNet is adapted to take in the historical citywide mobility data in terms of short-term (close), mid-term (period), and long-term (trend) sequences to predict the mobility in time step $(t+1)$.
5) EVL [9]: which implements the Extreme Value Loss (EVL). Here EVL categorizes the mobility patterns into high/normal/low categories based on the predefined thresholds and utilizes an extreme value theory based loss function for multi-region mobility prediction.
6) CHAT [10]: which leverages the Cross-Interaction Hierarchical Attention (CHAT) network for extreme mobility prediction. CHAT takes in the spatial, temporal, and contextual information, and retrieves the joint representations of anomaly or extreme patterns from the mobility data across these three aspects.

• **Evaluation Metrics:** In this study, we use the *Error Rate* (ER), *Mean Squared Logarithmic Error* (MSLE), and *R-squared* (R^2) as evaluation metrics. Taking the mobility prediction of N regions in time step $(t+1)$ as an example, the evaluation metrics are formally given by

$$\text{ER} = \frac{\sum_{n=1}^{N} \left| \mathbf{X}_{n,t+1} - \widehat{\mathbf{X}}[n, t+1] \right|}{\sum_{n=1}^{N} \mathbf{X}[n, t+1]},$$

$$\text{MSLE} = \frac{1}{N} \times \sum_{n=1}^{N} | \log_2 \left(\widehat{\mathbf{X}}[n, t+1] + 1 \right) - \log_2 \left(\mathbf{X}[n, t+1] + 1 \right) |,$$

and

$$\text{R}^2 = 1 - \frac{\sum_{n=1}^{N} \left(\mathbf{X}[n, t+1] - \widehat{\mathbf{X}}[n, t+1] \right)^2}{\sum_{n=1}^{N} \left(\mathbf{X}[n, t+1] - \bar{\mathbf{X}} \right)^2},$$

where $\widehat{\mathbf{X}}[n, t+1]$ and $\mathbf{X}[n, t+1]$ denote the predicted and ground-truth mobility of region n in time step $(t+1)$, and $\bar{\mathbf{X}}$ denotes the mean value of all the ground-truth values in the predicted time steps.

• **Details of Mobility Data and Events Studied.** Given the trained models from the preceding time periods, we will further test our model by the following time periods:

(a) For Citi bike sharing in NYC, the time periods studied include: *1)* 09/18/2020–09/27/2020 without the occurrence of anomaly/extreme events, *2)* 07/31/2020–08/09/2020 when NYC was affected by the Hurricane Isaias on 08/04/2020, and *3)* 12/22/2020–12/31/2020 with the public holidays of

[2] https://ride.citibikenyc.com/system-data
[3] https://ride.divvybikes.com/system-data
[4] https://data.cityofchicago.org/Transportation/Taxi-Trips-2021/9kgb-ykyt

NYC for Christmas holidays on 12/24/2020–12/25/2020. **(b)** For the Yellow Taxi trip data in NYC, we studied the following three time periods: *1)* 04/21/2016–04/30/2016 without the occurrence of anomaly/extreme events, *2)* 04/01/2016–04/10/2016 when the city was significantly under influence by wind gust and freezing rain during the 04/03/2016–04/04/2016, and *3)* 05/05/2016–06/04/2016 with the Memorial Day on 05/30/2016. **(c)** For the Divvy bike sharing and taxi trip data in Chicago, we studied the following three time periods: *1)* 06/01/2021–06/10/2021 without the occurrence of anomaly/extreme events, *2)* 10/22/2021–10/31/2021 when the city was significantly influenced by heavy rainstorm during the 10/24/2021–10/25/2021, and *3)* 11/21/2021–11/30/2021 with the Thanksgiving Holidays.

Including each of the ten days of the predicted time periods as mentioned above, from each mobility dataset, we retrieve three months to predict the mobility of each region during the occurrence of each selected situation. In each experiment, the last
i.e.,
pric

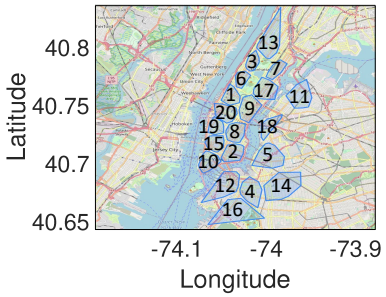


Fig. 10: Region indices of bike sharing trips in NYC (2020).

We process each dataset as follows. For the NYC Citi bike sharing Data (2020) and the Chicago Divvy bike sharing Data (2021), we remove *1)* the bike stations who do not exist in the whole year of 2020 and 2021, respectively, *2)* the bike stations whose average hourly bike pick-ups are less than 1, *3)* the trips with errors in the timestamps, and *4)* the trips with duration less than 1 minute. For the NYC yellow taxi trip data (2016) and the Chicago taxi trip Data (2021), we remove the trips *1)* with incorrect timestamps, and *2)* with duration of less than 1 minute.

Parameter Settings: Following the practices of improving bike distributions for analysis [29], [35], We partition NYC into 20 regions and Chicago into 18 regions by the K-means [36] algorithm using both the geographical locations of the bike stations and the pick-up/drop-off locations of all the taxi trips of each city. We illustrate the clustered and generated regions of NYC in Fig. 10. For each region n at time step t, we aggregate the bike sharing usage (pick-ups) of all the bike stations in region n and form $\mathbf{X}[n,t]$. We aggregate the taxi trips and find $\mathbf{X}[n,t]$ in the same manner. For Citi bike sharing and the Yellow Taxi trips in NYC, the number of the training epochs is set to 500 with a learning rate of 0.0002 and a batch size of 128. Also, we set L and M to 5 and 3 respectively.

TABLE II: Prediction results and performance comparison on the Citi bike sharing data in NYC (2020).

Scheme	Normal			Hurricane			Christmas		
	ER	MSLE	R^2	ER	MSLE	R^2	ER	MSLE	R^2
ARIMA	0.611	3.583	0.374	0.640	3.344	0.335	0.947	5.682	-0.042
GRU	0.263	3.916	0.797	0.390	7.024	0.387	0.392	4.921	0.777
LSTM	0.284	3.856	0.845	0.569	13.589	0.332	0.632	9.712	0.335
RNN	0.312	6.200	0.731	0.348	6.789	0.669	0.373	4.546	0.785
ST-Norm	0.295	1.146	0.837	0.339	1.381	0.773	0.318	1.465	0.858
ST-ResNet	0.320	6.728	0.773	0.682	20.122	0.202	0.715	12.274	0.030
EVL	0.247	5.001	0.719	0.417	9.275	0.421	0.306	3.345	0.733
CHAT	0.276	1.283	0.838	0.349	2.502	0.745	0.546	2.641	0.470
EALGAP	**0.205**	**0.661**	**0.920**	**0.256**	**0.932**	**0.833**	**0.282**	**1.194**	**0.879**

TABLE III: Prediction results and performance comparison on the Divvy bike sharing data in Chicago (2021).

Scheme	Normal			Rainstorm			Thanksgiving		
	ER	MSLE	R^2	ER	MSLE	R^2	ER	MSLE	R^2
ARIMA	0.605	1.569	0.517	1.025	2.978	0.193	0.736	1.434	0.373
GRU	0.327	3.855	0.900	0.665	4.812	0.593	0.483	2.053	0.790
LSTM	0.313	2.670	0.896	0.545	1.604	0.715	0.474	1.244	0.802
RNN	0.326	2.801	0.894	0.481	2.084	0.736	0.673	1.435	0.344
ST-Norm	0.315	1.119	0.840	0.474	1.834	0.780	0.552	2.016	0.795
ST-ResNet	0.543	9.151	0.739	0.542	4.661	0.817	0.606	4.368	0.798
EVL	0.327	2.702	0.869	0.504	2.422	0.217	0.428	1.441	0.756
CHAT	0.414	1.122	0.825	0.622	0.872	0.469	0.516	0.842	0.656
EALGAP	**0.284**	**0.871**	**0.907**	**0.404**	**0.866**	**0.817**	**0.414**	**0.786**	**0.826**

For Divvy bike sharing and taxi trips in Chicago, we set the number of the training epochs to 1000 with a learning rate of 0.0002 and a batch size of 128. Also, we set the values of L and M to 2.

In this study, all the experiments are conducted upon a desktop server with Intel i7-9700 CPU, NVIDIA GeForce RTX 2060 SUPER GPU, 16.0 GB RAM, and Windows 10. The proposed model is implemented in Python with Tensorflow-GPU-2.3.0. For all the baselines, the batch size, the number of regions N, the length of time steps of the near history mobility data L, and the number of windows M are the same as EALGAP when predicting the same mobility data. The average training time of each deep learning method per time step and epoch (on our desktop server) are: 0.023ms for GRU, 0.024ms for LSTM, 0.011ms for RNN, 0.574ms for ST-Norm, 0.119ms for ST-ResNet, 1.028ms for EVL, 0.135ms for CHAT, and 0.307ms for EALGAP, respectively.

C. Evaluation Results

• **Overall Performance:** We illustrate the results of all the experiments in Tables II, III, IV, and V. The prediction results show that our proposed model EALGAP outperforms all other selected baseline models in urban mobility prediction even with the existence of strong turbulence in the mobility data. Furthermore, our results show that accounting for the extreme degree of mobility due to the occurrence of anomaly events can significantly enhance the model's urban mobility prediction accuracy. In particular, with respect to each of the four datasets, *i.e.,* Citi in NYC, Divvy in Chicago, Yellow Taxi in NYC, and taxi trips in Chicago, EALGAP can achieve 61.61%, 40.65%, 30.93%, and 43.29% lower errors (on average for all metrics) than the other state-of-the-art approaches.

Compared with ARIMA, GRU, LSTM, RNN, ST-ResNet, and CHAT, our EALGAP can model the dynamics of mobility patterns in different regions during anomaly events, thereof

TABLE IV: Prediction results and performance comparison on the yellow taxi trip data in NYC (2016).

Scheme	Normal			Wind Gust			Memorial Day		
	ER	MSLE	R^2	ER	MSLE	R^2	ER	MSLE	R^2
ARIMA	0.485	1.849	0.499	0.499	1.857	0.498	0.504	1.818	0.473
GRU	0.354	3.029	0.600	0.334	3.147	0.719	0.349	2.971	0.696
LSTM	0.343	3.092	0.750	0.342	3.146	0.745	0.345	2.174	0.745
RNN	0.305	2.829	0.807	0.330	3.266	0.696	0.331	2.984	0.752
ST-Norm	0.286	0.949	0.802	0.269	0.830	0.844	0.335	1.734	0.806
ST-ResNet	0.288	1.633	0.825	0.281	1.675	0.836	0.308	2.333	0.804
EVL	0.289	1.635	0.763	0.262	1.388	0.796	0.292	1.501	0.773
CHAT	0.293	1.027	0.814	0.301	1.067	0.801	0.306	1.136	0.795
EALGAP	**0.275**	**0.890**	**0.840**	**0.263**	**0.732**	**0.854**	**0.286**	**0.786**	**0.822**

TABLE V: Prediction results and performance comparison on the taxi trip data in Chicago (2021).

Scheme	Normal			Rainstorm			Thanksgiving		
	ER	MSLE	R^2	ER	MSLE	R^2	ER	MSLE	R^2
ARIMA	0.404	0.731	0.766	0.514	1.153	0.674	0.653	1.969	0.577
GRU	0.318	1.796	0.916	0.336	3.999	0.931	0.368	4.559	0.881
LSTM	0.332	1.954	0.891	0.276	3.233	0.956	0.371	4.741	0.920
RNN	0.305	2.007	0.918	0.330	4.251	0.933	0.296	3.632	0.944
ST-Norm	0.258	0.631	0.944	0.246	0.889	0.947	0.301	0.937	0.874
ST-ResNet	0.596	2.182	0.199	0.514	3.368	0.670	0.564	2.684	0.528
EVL	0.350	2.979	0.856	0.300	3.998	0.944	0.338	4.145	0.914
CHAT	0.584	1.688	0.626	0.484	1.508	0.762	0.556	2.219	0.692
EALGAP	**0.247**	**0.598**	**0.948**	**0.226**	**0.777**	**0.962**	**0.251**	**0.769**	**0.953**

improving the prediction accuracy and corroborating the effectiveness of the Extreme Degree Modeling Module. ST-Norm considers that the global impacts of the data in different regions is constant over a period of time, which nevertheless may not hold well in dynamic spatio-temporal mobility data. It is mainly because the overall temporal patterns of mobility in different time periods are different. In contrast, our EALGAP considers the temporally-varying global impacts, and captures the overall spatial dependencies and the regular temporal patterns for the mobility of different regions for different time steps, which validates our design of the Global Impact Modeling Module. Furthermore, EVL captures both the regular temporal patterns and the extreme degrees of the data in each region. However, it does not account for the spatial correlations among the region for more accurate urban mobility prediction.

• **Ablation Studies:** To validate the designs of each component of our proposed model, taking NYC Citi bike sharing data during the Hurricane Isaias, we have carried out the following experiments: *(i)* complete EALGAP; *(ii)* EALGAP with only the Global Impact Modeling Module; *(iii)* EALGAP with only the Extreme Degree and Local Impact Modeling Module; *(iv)* replacing the exponential distribution in the Global Impact Modeling Module with the normal distribution; *(v)* region partitioning with Density-Based Spatial Clustering of Applications with Noise (DBSCAN) [37]; and *(vi)* region partitioning with Ordering Points To Identify the Clustering Structure (OPTICS) [38] based on the geological distance. We note that in *(iii)* we use two Dense layers with ReLU activation function to predict the global impacts $\mathbf{X}^g[:, t+1] \in \mathbb{R}^N$ of the mobility patterns in time step $(t+1)$. The parameters in each experiment are the same as EALGAP.

As shown in Fig. 11, both of the Global Impact Modeling Module and the Extreme Degree and Local Impact Modeling Module can model the dynamic of mobility from different regions. Especially, since the Extreme Degree and Local Impact Modeling Module can capture the extreme degree of the mobility of each region under the occurrence of anomaly events, its results are better than the models solely using the Global Impact Modeling Module. Combing both of the components can achieve the best modeling of the mobility of different regions in various time steps. In addition, compared with the normal distribution, the exponential distribution can better represent the dynamics of the mobility in different regions and help more accurate urban mobility predictions. We can also observe that using our default region partitioning method EALGAP suffices to achieve good mobility prediction accuracy.

• **Sensitivity Studies:** We have conducted sensitivity studies for our proposed model on two parameters, the length of near history mobility data, L, and the number of windows, M. In particular, both L and M are set from 2 to 6. L and M represent the length of time steps of the most relevant near history instantaneous mobility dynamic and the regular dynamic of the mobility when predicting the mobility of a region in the next time step, respectively. Leveraging the mobility during these most relevant time steps will greatly reduce the computational cost and improve the prediction accuracy of the model.

As shown in Fig. 12, our proposed model EALGAP is consistent in accurately predicting the mobility of each region in the city during the occurrence of anomaly events. However, we can still find that setting L as 5 and M as 3 can achieve the best prediction results. Therefore, to predict the urban mobility of a future time step, the mobility data belonging to the previous 3 time steps along with the mobility of the same time step for the previous 5 days contain the most relevant information for bike sharing trips in NYC. In the experiments, we also find that using L and M greater than 6 or more may cause the prediction accuracy to significantly drop. We speculate the reason as, considering the local instantaneous mobility dynamics incurred by the anomaly events only last for a short time period and the regular dynamic of urban mobility is also influenced by time, using large L and M may impact the characterization of both the local and global impacts.

• **Result Visualization:** We visualize the predictions (Pre) and ground-truths (Gro) of the Citi bike sharing pick-ups of NYC in 2020 during the normal situations (a) and two anomaly events (b and c) in Fig. 13. We can see that EALGAP achieves high accuracy in capturing the instantaneous temporal dynamics of mobility patterns under both normal and anomalous periods. We can also observe that the urban mobility deviated from the overall global trends due to the Hurricane Isaias on 08/04/2020 and the Christmas holidays during 12/24/020–12/25/2020. Our EALGAP can still capture the instantaneous dynamics of the mobility of each region by capturing the extreme degree.

In addition, we illustrate the ground-truth and prediction heatmaps of the urban bike sharing pick-ups of NYC for one of the predicted time steps during the occurrences of hurricane and the public holidays (Christmas) in Figs. 14 and 15. For

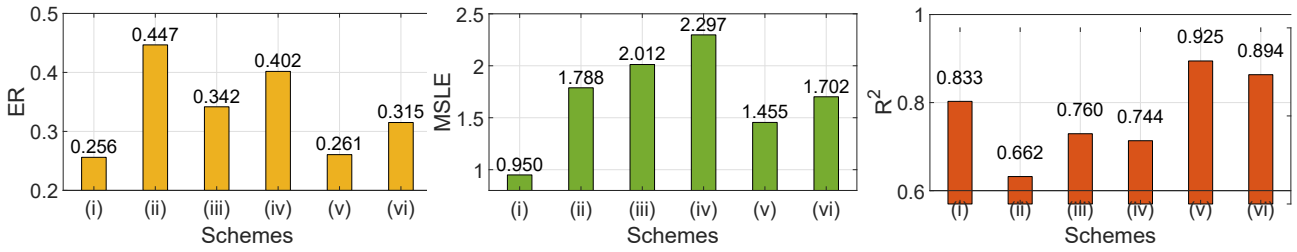


Fig. 11: Ablation studies with (a) ER; (b) MSLE; and (c) R^2 of predicting the bike pick-ups of 20 regions of NYC during the hurricane on 08/04/2020.

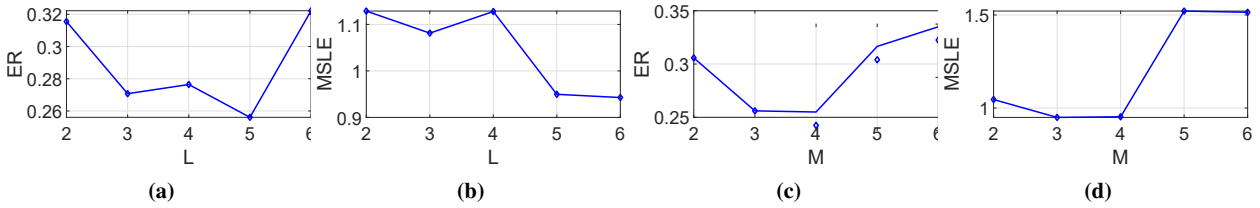


Fig. 12: Sensitivity studies on EALGAP. (a) and (b): ER and MSLE of EALGAP on L; (c) and (d): ER and MSLE of EALGAP on M. We studied the bike sharing pick-up prediction in NYC during 08/04/2020.

each region, we show the aggregate bike usage of the stations within that region with the same colors, and the warmer colors imply more bike sharing pick-ups for a region. By incorporating the overall global trend and the instantaneous local changes of the mobility, this demonstrates the effectiveness of EALGAP to address the data imbalance issue of the mobility prediction, which can benefit the applications in city management, related resource allocation, and emergency managements.

VII. Conclusion

We propose EALGAP, a novel extreme-aware local-global attention urban mobility prediction model, which takes in the spatio-temporal local and global impacts of mobility in different regions and time steps for prediction at various city regions. In particular, we have taken into account the global impacts by extracting the overall spatial dependencies and the regular temporal patterns of mobility systems (*e.g.,* bike sharing, taxicab) for different city regions. We have designed a normalization-based technique to retrieve the extreme degrees of mobility patterns locally within different regions and time steps given the occurrences of anomaly events. We have conducted extensive experimental studies upon four mobility datasets harvested from two metropolitan cities in U.S. with anomaly natural/social events (*e.g.,* hurricane events, Federal holidays). Our results have demonstrated the accuracy and extreme-awareness of our proposed EALGAP compared with other state-of-the-art approaches.

Acknowledgments

We would like to thank the anonymous reviewers for their constructive comments. This project was supported, in part, by the National Science Foundation (NSF) under Grant 2118102. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the funding agencies.

References

[1] R. Jiang, Z. Cai, Z. Wang, C. Yang, Z. Fan, Q. Chen, X. Song, and R. Shibasaki, "Predicting Citywide Crowd Dynamics at Big Events: A Deep Learning System," *ACM Transactions on Intelligent Systems and Technology (TIST)*, vol. 13, no. 2, pp. 1–24, 2022.

[2] Y. Liang, K. Ouyang, J. Sun, Y. Wang, J. Zhang, Y. Zheng, D. Rosenblum, and R. Zimmermann, "Fine-grained Urban Flow Prediction," in *Proc. the Web Conference*, 2021, pp. 1833–1845.

[3] Z. Lin, S. Lyu, H. Cao, F. Xu, Y. Wei, H. Samet, and Y. Li, "HealthWalks: Sensing Fine-grained Individual Health Condition via Mobility Data," *Proc. ACM IMWUT*, vol. 4, no. 4, pp. 1–26, 2020.

[4] S. Ruan, J. Bao, Y. Liang, R. Li, T. He, C. Meng, Y. Li, Y. Wu, and Y. Zheng, "Dynamic Public Resource Allocation Based on Human Mobility Prediction," *Proc. ACM IMWUT*, vol. 4, no. 1, pp. 1–22, 2020.

[5] S. A. Pedersen, B. Yang, and C. S. Jensen, "Anytime Stochastic Routing with Hybrid Learning," *Proc. VLDB Endowment*, vol. 13, no. 9, pp. 1555–1567, 2020.

[6] K. D'Silva, K. Jayarajah, A. Noulas, C. Mascolo, and A. Misra, "The Role of Urban Mobility in Retail Business Survival," *Proc. ACM IMWUT*, vol. 2, no. 3, pp. 1–22, 2018.

[7] J. Deng, X. Chen, R. Jiang, X. Song, and I. W. Tsang, "ST-Norm: Spatial and Temporal Normalization for Multi-variate Time Series Forecasting," in *Proc. ACM SIGKDD*, 2021, pp. 269–278.

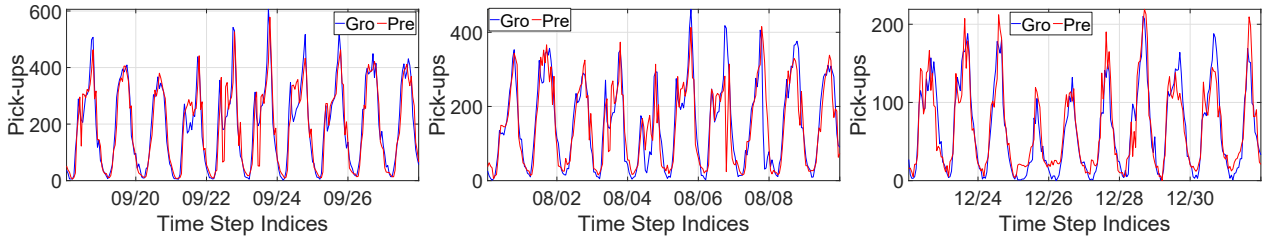
[8] J. Zhang, Y. Zheng, and D. Qi, "Deep Spatio-temporal Residual Networks for Citywide Crowd Flows Prediction," in *Proc. AAAI*, 2017.

[9] D. Ding, M. Zhang, X. Pan, M. Yang, and X. He, "Modeling Extreme Events in Time Series Prediction," in *Proc. ACM SIGKDD*, 2019, pp. 1114–1122.

[10] C. Huang, C. Zhang, P. Dai, and L. Bo, "Cross-interaction Hierarchical Attention Networks for Urban Anomaly Prediction," in *Proc. IJCAI*, 2021, pp. 4359–4365.

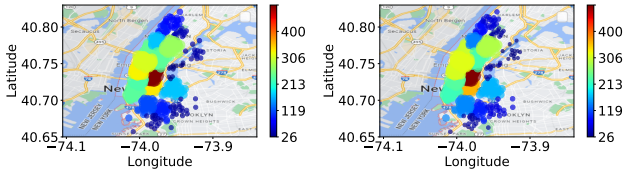
[11] S. Foss, D. Korshunov, S. Zachary *et al.*, *An Introduction to Heavy-tailed and Subexponential Distributions*. Springer, 2011, vol. 6.

[12] H. Li, S. Zhang, X. Li, L. Su, H. Huang, D. Jin, L. Chen, J. Huang, and J. Yoo, "DetectorNet: Transformer-enhanced Spatial Temporal Graph Neural Network for Traffic Prediction," in *Proc. ACM SIGSPATIAL*, 2021, pp. 133–136.



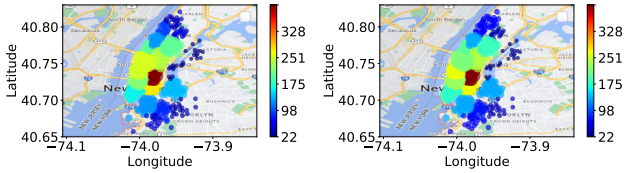
(a) Normal. **(b)** Hurricane. **(c)** Christmas.

Fig. 13: Predictions (Pre) and ground-truths (Gro) of bike sharing pick-up prediction during (a) normal situation from 09/18/2020 to 09/27/2020; (b) ten days from 07/31/2020 to 08/09/2020 when hurricane occurred in NYC on 08/04/2020; and (c) ten days from 21/22/2020 to 12/31/2020 during the public holidays of NYC for Christmas.



(a) **(b)**

Fig. 14: The ground-truth (a) and predicted heatmaps (b) of the bike pick-ups of NYC of one selected time step during the occurrence of the hurricane.



(a) **(b)**

Fig. 15: The ground-truth (a) and predicted heatmaps (b) of the bike pick-ups of NYC of one of the predicted time steps during the public holidays of Christmas.

[13] C. Jiang, J. Li, W. Wang, and W.-S. Ku, "Modeling Real Estate Dynamics using Temporal Encoding," in *Proc. ACM SIGSPATIAL*, 2021, pp. 516–525.

[14] H. Li, D. Jin, X. Li, J. Huang, and J. Yoo, "Multi-Task Synchronous Graph Neural Networks for Traffic Spatial-Temporal Prediction," in *Proc. ACM SIGSPATIAL*, 2021, pp. 137–140.

[15] H. Huang, X. Yang, and S. He, "Multi-Head Spatio-Temporal Attention Mechanism for Urban Anomaly Event Prediction," *Proc. ACM IMWUT*, vol. 5, no. 3, pp. 1–21, 2021.

[16] S. He and K. G. Shin, "Towards fine-grained flow forecasting: A graph attention approach for bike sharing systems," in *Proc. WWW*, 2020, p. 88–98.

[17] R.-G. Cirstea, B. Yang, C. Guo, T. Kieu, and S. Pan, "Towards Spatio-Temporal Aware Traffic Time Series Forecasting," in *Proc. IEEE ICDE*, 2022.

[18] J. Sun, J. Zhang, Q. Li, X. Yi, Y. Liang, and Y. Zheng, "Predicting Citywide Crowd Flows in Irregular Regions using Multi-view Graph Convolutional Networks," *IEEE Transactions on Knowledge and Data Engineering*, 2020.

[19] Z. Pan, W. Zhang, Y. Liang, W. Zhang, Y. Yu, J. Zhang, and Y. Zheng, "Spatio-temporal Meta Learning for Urban Traffic Prediction," *IEEE Transactions on Knowledge and Data Engineering*, 2020.

[20] R. Li, H. He, R. Wang, Y. Huang, J. Liu, S. Ruan, T. He, J. Bao, and Y. Zheng, "JUST: JD Urban Spatio-temporal Data Engine," in *Proc. IEEE ICDE*. IEEE, 2020, pp. 1558–1569.

[21] S. He and K. G. Shin, "Information fusion for (re)configuring bike station networks with crowdsourcing," *IEEE TKDE*, vol. 34, no. 2, pp. 736–752, 2022.

[22] A. Yan and B. Howe, "FairST: Equitable Spatial and Temporal Demand Prediction for New Mobility Systems," in *Proc. ACM SIGSPATIAL*, 2019, pp. 552–555.

[23] J. Chen, X. Qiu, P. Liu, and X. Huang, "Meta Multi-task Learning for Sequence Modeling," in *Proc. AAAI*, vol. 32, no. 1, 2018.

[24] C. Zheng, X. Fan, C. Wang, and J. Qi, "GMAN: A Graph Multi-attention Network for Traffic Prediction," in *Proc. AAAI*, vol. 34, no. 01, 2020, pp. 1234–1241.

[25] A. Siffer, P.-A. Fouque, A. Termier, and C. Largouet, "Anomaly Detection in Streams with Extreme Value Theory," in *Proc. ACM SIGKDD*, 2017, pp. 1067–1075.

[26] H. Ren, B. Xu, Y. Wang, C. Yi, C. Huang, X. Kou, T. Xing, M. Yang, J. Tong, and Q. Zhang, "Time-series Anomaly Detection Service at Microsoft," in *Proc. ACM SIGKDD*, 2019, pp. 3009–3017.

[27] Y. Su, Y. Zhao, C. Niu, R. Liu, W. Sun, and D. Pei, "Robust Anomaly Detection for Multivariate Time Series through Stochastic Recurrent Neural Network," in *Proc. ACM SIGKDD*, 2019, pp. 2828–2837.

[28] R. L. Smith, "Extreme Value Theory," *Handbook of applicable mathematics*, vol. 7, pp. 437–471, 1990.

[29] Y. Li, Y. Zheng, H. Zhang, and L. Chen, "Traffic Prediction in a Bike-sharing System," in *Proc. ACM SIGSPATIAL*, 2015, pp. 1–10.

[30] L. Chen, D. Zhang, L. Wang, D. Yang, X. Ma, S. Li, Z. Wu, G. Pan, T.-M.-T. Nguyen, and J. Jakubowicz, "Dynamic Cluster-based Over-demand Prediction in Bike Sharing Systems," in *Proc. ACM UbiComp*, 2016, pp. 841–852.

[31] X. Yang, S. He, B. Wang, and M. Tabatabaie, "Spatio-Temporal Graph Attention Embedding for Joint Crowd Flow and Transition Predictions: A Wi-Fi-based Mobility Case Study," *Proc. ACM IMWUT*, vol. 5, no. 4, pp. 1–24, 2021.

[32] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin, "Attention is All You Need," *Proc. NeurIPS*, vol. 30, 2017.

[33] D. Ulyanov, A. Vedaldi, and V. Lempitsky, "Improved Texture Networks: Maximizing Quality and Diversity in Feed-forward Stylization and Texture Synthesis," in *Proc. IEEE CVPR*, 2017, pp. 6924–6932.

[34] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *nature*, vol. 521, no. 7553, pp. 436–444, 2015.

[35] Y. Li and Y. Zheng, "Citywide Bike Usage Prediction in a Bike-sharing System," *IEEE Transactions on Knowledge and Data Engineering*, vol. 32, no. 6, pp. 1079–1091, 2019.

[36] J. A. Hartigan and M. A. Wong, "Algorithm AS 136: A K-means Clustering Algorithm," *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, vol. 28, no. 1, pp. 100–108, 1979.

[37] M. Ester, H.-P. Kriegel, J. Sander, X. Xu *et al.*, "A Density-based Algorithm for Discovering Clusters in Large Spatial Databases with Noise," in *Proc. ACM KDD*, vol. 96, no. 34, 1996, pp. 226–231.

[38] M. Ankerst, M. M. Breunig, H.-P. Kriegel, and J. Sander, "OPTICS: Ordering Points to Identify the Clustering Structure," *ACM SIGMOD Record*, vol. 28, no. 2, pp. 49–60, 1999.