Fairness-enhancing deep learning for ride-hailing demand prediction

Yunhan Zheng, Qingyi Wang, Dingyi Zhuang, Shenhao Wang*, and Jinhua Zhao

***Abstract*—Short-term demand forecasting for on-demand ride-hailing services is a fundamental issue in intelligent transportation systems. However, previous research predominantly focused on improving prediction accuracy, ignoring fairness issues such as systematic underestimations of travel demand in disadvantaged neighborhoods. This study investigates how to measure, evaluate, and enhance prediction fairness between disadvantaged and privileged communities in spatial-temporal demand forecasting of ride-hailing services. We developed a socially-aware neural network (SA-Net) that integrates socio-demographics and ridership information for fair demand prediction, and introduced a bias-mitigation regularization to reduce the prediction error gap between black and non-black, and low-income and high-income communities. The experimental results, using Chicago Transportation Network Company (TNC) data, demonstrate that our de-biasing SA-Net model outperforms other models in both prediction accuracy and fairness. Notably, the SA-Net exhibits a significant improvement in prediction accuracy, reducing 2.3% in Mean Absolute Error (MAE) compared to state-of-the-art models. When coupled with the bias-mitigation regularization, the de-biasing SA-Net effectively bridges the mean percentage prediction error (MPE) gap between the disadvantaged and privileged groups, and protects the disadvantaged regions against systematic underestimation of TNC demand. Specifically, our approach reduces the MPE gap between black and non-black communities by 67% without compromising overall prediction accuracy.**

***Index Terms*—Spatial-temporal travel demand prediction, algorithmic fairness, demand forecasting, ride-hailing service**

I. Introduction

In recent years, on-demand ride-hailing services have grown rapidly. Transportation network companies (TNCs) such as Uber and Lyft provide the ride-hailing services by connecting passengers with drivers based on real-time information [1, 2]. Reliable and accurate short-term travel demand forecasting is a promising tool to balance vehicle supply and demand with low cost and high quality of service [3, 4, 5]. Researchers have developed a series of data-driven approaches to predict travel demand in real-time, including time series analysis methods [6, 7], machine learning methods [8, 9] and deep learning models [3, 10, 11]. These approaches typically divide the study region into small areas, use the past travel requests in a time interval as the historical demand, and then seek to enhance the prediction accuracy of the future travel demand as a function of the historical demand (assuming certain spatial and temporal correlations among them) and exogenous features such as the weather and holiday.

However, a narrow focus on prediction accuracy ignores the crucial social consequences underlying the prediction tasks, such as unfairness in travel demand forecasting. For instance, since the transport operators depend on the predicted passenger demand to dispatch vehicles, if travel demand in disadvantaged neighborhoods is systematically under-predicted, the resulting service provision may be inadequate. The existing literature has the following two limitations: first, most previous studies evaluated the performance of the demand predictions by the average prediction accuracy across the whole study region, while research into the disparity of predictive performance between the disadvantaged and privileged areas is very scarce. This raises an equity concern because if the ride-hailing demand for the disadvantaged neighborhoods is systematically underestimated, the vehicles allocated to these neighborhoods may not be enough to serve the actual demand. Second, most previous models did not consider the socioeconomic and demographic information of the areas when making travel demand predictions. Areas with different socioeconomic and demographic makeup could have very different spatial-temporal dependencies. Failure to account for the heterogeneity of these spatial-temporal dependencies can lead to biased model results.

To overcome these limitations, this paper presents a novel approach that aims to enhance prediction fairness without compromising high prediction accuracy in zone-level ride-hailing demand forecasts. This strategy is comprised of a new deep learning architecture, named the *socially aware network* (SA-Net), and a bias-mitigation regularization method, to achieve fairness-aware travel demand predictions. While previous research typically adopted spatially-invariant convolutional kernels to capture spatial dependencies, this new network incorporates a novel Socially-Aware Convolution (SAC) module that adapts the standard invariant kernel at each area of the study region based on the socio-demographic makeup of that area, which is highly flexible and thus can better capture the spatial-temporal dependencies across different locations. The bias-mitigation regularization method modifies the traditional objective functions in deep learning travel demand predictions by adding a fairness regularization term, thus facilitating fair travel demand predictions.

The main contributions of this paper can be summarised as

Yunhan Zheng, Qingyi Wang, Dingyi Zhuang, and Jinhua Zhao are with the Department of Civil and Environmental Engineering, Massachusetts Institute of Technology, Cambridge, MA 02139, USA .

Shenhao Wang is with the Department of Urban and Regional Planning, University of Florida, Gainesville, FL 32611, USA (email: shenhaowang@ufl.edu), corresponding author.

follows:

- We propose a new model (SA-Net) that adopts location-specific modification to the standard spatially-invariant convolutional filters. The proposed network can flexibly capture the spatial heterogeneity by incorporating the local socioeconomic and demographic information into the prediction process.
- We propose a fairness metric, the *mean percentage error gap* (MPE Gap), which measures the gap of mean percentage prediction error between the disadvantaged and privileged groups. A positive MPE indicates that the model has underestimated the demand, whereas a negative MPE indicates an overestimation of the demand.
- We develop a bias-mitigation regularization method that allows the network to learn fair predictions by bringing down the MPE Gap between the disadvantaged and privileged groups.
- Experiments on Chicago TNC data reveal the risk of generating spatially unfair demand prediction with the state-of-the-art spatial-temporal deep learning predictions, and show that our proposed new method can not only reduce the fairness gap between the disadvantaged and privileged groups, but also increase the overall prediction accuracy.

The rest of the paper is organized as follows. Section II reviews the existing literature on ride-hailing demand prediction and fairness in machine learning. Section III defines the research problem. Section IV describes the model architecture of the proposed SA-Net, the fairness evaluation metrics, as well as the bias mitigation regularization method. Section V shows the experiment results, which compare the prediction accuracy and fairness between the proposed de-biasing SA-Net and the benchmark models on the Chicago TNC dataset. Section VIII concludes the paper.

II. Literature Review

A. Spatial-temporal travel demand forecasting

Spatial-temporal travel demand forecasting has been a fundamental issue in intelligent transportation systems [12, 10, 13]. It involves predicting the demand for travel in specific locations over time. Early research in this field focused on traditional time-series regression models such as ARIMA [6] and Kalman Filter Lippi et al. [14]. These models were effective in capturing temporal patterns and variations in travel demand.

In recent years, there has been a shift towards machine learning-based approaches due to the availability of large-scale data and increased computing power [15, 16, 17, 18]. Machine learning models, such as support vector regression [8] and regression trees [9], have been employed for travel demand forecasting. More recently, deep learning methods have gained significant popularity due to their capabilities of approximating human's decision functions and their capability of capturing complex spatial-temporal correlations in the data [19, 20].

Deep learning architectures, including recurrent neural networks (RNNs) [21] and long short-term memory (LSTM) [22, 23], have been widely used to capture sequential dependencies and long-range temporal dependencies in travel demand data. These models can effectively handle the temporal dynamics of travel demand. Additionally, convolutional neural networks (CNNs) have been adopted to capture spatial correlations in grid-based travel demand predictions. By using localized kernels, CNNs can identify local and global spatial patterns in the data [24]. Furthermore, graph neural networks (GCNs) have been explored to capture non-Euclidean spatial correlations in network-structured data, such as station-based or traffic network scenarios [25, 26, 27, 28]. More recently, GCNs have been applied to region-based scenarios, such as urban areas [29].

B. Ride-hailing demand prediction

Ride-hailing demand prediction is a specific application of spatial-temporal travel demand forecasting, focusing on predicting the demand for ride-hailing services in particular. This area has garnered significant attention in recent years, given the rise of ride-hailing platforms. Researchers have developed various techniques to accurately forecast ride-hailing demand.

Early ride-hailing demand prediction models utilized traditional time-series regression models, such as ARIMA and its variants [30, 31]. However, with the advancements in machine learning, researchers have increasingly turned to more sophisticated approaches. Deep learning methods, in particular, have gained prominence in this domain [4].

Several studies have proposed deep learning architectures specifically designed for ride-hailing demand prediction. For instance, Ke et al. [3] developed a Conv-LSTM network that combines CNNs and LSTMs to capture spatial, temporal, and exogenous dependencies simultaneously. Huang et al. [32] proposed a GAN framework-based dynamic multi-graph convolutional network tailored for origin-destination-based ride-hailing demand prediction. Rahman and Rifaat [33] employed spatio-temporal deep learning techniques to forecast demand and supply-demand gaps in ride-hailing systems, incorporating anonymized spatial adjacency information. Chen et al. [34] developed a multivariate deep learning convolutional neural network that incorporates socio-demographic information to forecast ride-hailing demand.

Although the abovementioned methods have made remarkable progress on improving the prediction accuracy, most of them do not consider fairness when making predictions. Fairness essentially involves the evaluation of a predictive model regarding its social consequences, so without incorporating any socio-demographic information, the predictive models are hardly aware of the social consequences. Motivated by this research gap, this paper aims to evaluate and improve fairness in TNC travel demand forecasting by incorporating socio-demographics, proposing fairness metrics, and developing an fairness-enhancing prediction method.

C. Fairness in machine learning

There exists extensive machine learning literature showing that a model can act discriminatorily on one population in a variety of settings such as criminal risk assessment [35, 36], clinical care [37, 38] and credit risk evaluation [39, 40]. These studies made significant contributions in terms of formalizing fairness in machine learning [41, 42], designing fairness-enhancing algorithms [43, 44, 45] and solving fairness concerns in real-world industries [46, 47].

However, literature that investigated the algorithmic fairness issue in transportation research was very scarce. In the domain of travel behavior modeling, Zheng et al. [48] demonstrated prediction disparities regarding race, income, medical condition and region in travel behavior modeling using the 2017 National Household Travel Survey (NHTS) and the 2018-2019 My Daily Travel Survey in Chicago. The authors adopted an absolute correlation regularization method to mitigate the prediction biases. In the spatial-temporal travel demand modeling setting, to the best of our knowledge, very few studies tried to tackle the fairness issue. Yan and Howe [49] modified the loss function in deep learning to reduce the gap of per capita predicted bikeshare demand between the disadvantaged and advantaged regions. The modification is based on the fairness assumption that the per capita predicted demand should be the same across regions. Yan and Howe [50] leveraged adversarial learning to mitigate the gap in prediction errors of bikeshare demand between the advantaged and disadvantaged groups.

Although much progress has been made in addressing algorithmic bias, there are still several research limitations that need to be addressed. First, one critical source of bias is feature selection, where selected variables fail to capture sufficient details that affect different outcomes [51, 52]. To combat this, it is crucial to develop strategies to integrate sociodemographic information into the modeling process. Another limitation is that previous research on fairness has measured it based on the absolute value of demand, which can lead to errors in disadvantaged groups being considered insignificant. To address this, we propose a new measure of fairness based on the relative value of demand. This measure compares errors with the typical demand of the region, which is based on the concept of algorithmic fairness known as "equality of odds"[53]. This principle requires that all individuals who have a TNC demand should have an equal chance of having it reflected in the prediction, regardless of their social and demographic characteristics. By using this new measurement of fairness, we can better understand and mitigate algorithmic bias in ride-sharing platforms. To address these limitations, we build upon the state-of-the-art spatial-temporal travel demand models, and propose a novel method for fair predictions of TNC travel demand.

III. Problem Description and Preliminaries

The goal of this study is to predict the zone-level short-term TNC demand in the study area. Based on the method proposed by Ke et al. [3] and Guo and Zhang [10], the study area is partitioned into $I \times J$ grids with each grid referring to a zone. The temporal dimension considered is 1 hour. It is assumed that future TNC demand is correlated with the TNC demand in the past. It is also influenced by seasonality (time-of-day, day-of-week, etc.), and exogenous variables such as weather conditions and the level of transit service. The variables examined in this study are defined as follows:

1) **TNC demand**
   The TNC demand at the tth time slot across the whole region is denoted as $\mathcal{D}_t \in R^{I*J}$, which is defined as the number of TNC orders happened during that time interval. The TNC demand in grid (i, j) is then denoted as $(\mathcal{D}_t)_{i,j}$
2) **Time-of-day, day-of-week, holiday**
   By examining the Chicago traffic index data[1], we categorize 24h in each day into three periods: the peak hours (7am - 9am and 3pm - 7pm in workdays), the mid-peak hours (9am -3pm in workdays and 11am - 7pm in weekends), the off-peak hours (7pm - 7am in workdays and 7pm - 11am in weekends). We use tod_t to indicate this time-of-day variable, which takes values 0, 1, 2 if t belongs to the off-peak hours, the mid-peak hours and the peak hours, respectively. dow_t is the day-of-week variable, which takes value 1 if t is among the weekdays and 0 if t is among the weekends. The dummy variable h_t is used to indicate whether t is in a holiday or not.
3) **Weather**
   We consider precipitation as the weather variable, which is denoted as p_t. The precipitation data is obtained from the website of National Centers for Environmental Information [54].
4) **Socio-demographic data**
   The study examines various socio-demographic variables, including total population, population per squared kilometers, employment count, percentage of African-American population, percentage of female population, percentage of spanish speakers, percentage of foreign-born population, median household income, percentage of population with 2019 household income lower than $25,000, percentage of college graduates, percentage of population with age between 25 and 34, percentage of population with age over 65, percentage of transit commuters and percentage of population with no household vehicles. Employment data at the census tract level is collected from LEHD employment statistics [55]. The remaining census tract level socio-demographic data derived from the 2019 American Community Survey (ACS) 5-year estimates data. We use Z^p to represent the p^{th} socio-demographic variable and use P to denote the total number of socio-demographic variables.
5) **Transit service level**
   To account for the supply-demand interaction, transit service level represented by the frequency of bus and

[1] https://www.tomtom.com/en_gb/traffic-index/chicago-traffic/

rail serving each grid cell is used. Frequency of transit service is obtained from the General Transit Feed Specification (GTFS) [56], which details the transit schedule service each stop. For any grid cell (i, j), we count the total number of bus and rail visits to all stops located in the grid cell during time period t, and denote them as B_t and R_t respectively. The symbol $\mathcal{G}_t$ represents the transit service level at a specific time t. $\mathcal{G}_t$ is equal to the concatenation of B_t and R_t, denoted as $\mathcal{G}_t = [B_t, R_t]$.

The target of this study is to predict the TNC demand at time t ($\mathcal{D}_t$), given the historical TNC demand, the transit service supply, the time series features and the socio-demographic variables: $\{\mathcal{D}_s, \mathcal{G}_s, p_s | s = 0, ..., t-1\}$, $\{tod_s, dow_s, h_s | s = 0, ..., t\}$ and $\{Z^p | p = 1, ..., P\}$. This research focuses on two objectives: prediction accuracy and fairness. Prediction accuracy refers to the goal of minimizing the overall prediction errors. Prediction fairness refers to the goal of reducing the gap in mean percentage errors between the disadvantaged and privileged groups.

IV. Methodology

This research designs a novel SA-Net to predict the short-term TNC demand with enhanced fairness. We first introduce the Socially-Aware Convolution (SAC), a base module that is repeatedly used in SA-Net, and describe how SAC is adapted from the standard CNN. We will then introduce SAC-LSTM which combines SAC and LSTM. After that, we will explain the complete model architecture used in this study.

A. CNN and SAC

In this section, we will start with a formulation of the standard convolution neural network, and then extend it to the Socially-Aware Convolution (SAC). The concept of SAC is illustrated in Figure 1. We start from a standard convolution, which can be written as:

$$Y[m,p,q] = \sigma\Big(\sum_{i,j,n} W[m,n,i,j] * X[n, p+\hat{i}, q+\hat{j}]\Big), \quad (1)$$

where $Y \in R^{O \times S \times S}$ denotes the output tensor, $X \in R^{I \times S \times S}$ is the input tensor, $W \in R^{O \times I \times S \times S}$ denotes the filter weight. O, I, S and V represent the output channel size, input channel size, image size, and kernel size. $[p,q]$ denotes the pixel coordinates. m and n are the indices for the output and input channels. $\hat{i} = i - [V/2]$, $\hat{j} = j - [V/2]$. σ is the activation function. From Equation 1, we can see that the filter weight $W[m,n,i,j]$ is invariant to image locations. Therefore, the standard convolution is content-agnostic. To account for the local information, we use the Socially-Aware Convolution (SAC) which was built upon the work by Su et al. [57]. A SAC modifies the spatially invariant filter W with an adapting kernel K, which can be expressed as follows:

$$\begin{aligned} Y[m,p,q] = &\sigma\Big(\sum_{i,j,n} K(F[p+\hat{i}, q+\hat{j}], F[p,q]) * W[m,n,i,j] \\ &* X[n, p+\hat{i}, q+\hat{j}]\Big) \end{aligned} \quad (2)$$

where $F \in R^{S \times S}$ is the feature map, which will be explained in the following subsection. K represents the Gaussian kernel function: $K(f_1, f_2) = exp(-\frac{1}{2}(f_1 - f_2)^T(f_1 - f_2))$. The kernel values are higher for regions with similar feature values. The SAC operation represented by Equation 3 adapts the standard convolution filter W at each pixel by multiplying the spatially-invariant filter W with a spatially-varying adapting filter K. The feature map F picks up local features that reflect the relationships between different regions on the map.

B. Feature map construction

We construct the feature map F as a linear combination of various socio-demographic variables, which is shown in Figure 2. The feature value f_{ij} is calculated as $f_{ij} = \sum_p^P w_p * Z_{ij}^p$ where Z_{ij}^p represents the value of the socio-demographic variable p (e.g. population density, race, income etc.) for region $[i, j]$. To ensure a fair comparison and mitigate any potential bias stemming from variations in measurement scales across variables, each variable p has been standardized using z-scores. By applying a Gaussian kernel function to the feature values of the center pixel and its surrounding pixels, for each pixel value prediction, we emphasize the neighboring pixels that are more similar to this specific pixel in terms of the socio-demographic features. The underlying assumption is that the regions that have similar socio-demographic characteristics with their neighborhoods should have similar level of TNC demand with their neighborhoods as well.

C. LSTM and SAC-LSTM

We use LSTM, a special kind of Recurrent Neural Netowrk (RNN), to process the temporal information. LSTM is designed to avoid the long-term memory problem. The model first passes a sequence of input vectors to the memory cell tensors through the input gate, and then drops the redundant information through the forgot gate, and the cell state will be updated accordingly. Finally, after several iterations, the output gate will output a hidden sequence [58].

When dealing with the travel demand forecasting problem with spatial-temporal data, Ke et al. [3] proposed using the Conv-LSTM, which is a network that combines CNN and LSTM, to capture the spatial dependencies. Unlike LSTM, Conv-LSTM converts all the inputs, memory cell values, hidden states and various gates from 2D matrices to 3D tensors. Besides, Conv-LSTM replaces the Handamard product with the convolutional operator, which is used to explore spatially local correlations. However, Conv-LSTM utilizes the standard convolutional filters which are replicated across the tensors with shared weights, thus failing to account for the heterogeneity of spatial correlations. To address this drawback of standard convolutions, we modify the Conv-LSTM by replacing the standard convoluations with the SAC, and name the new network SAC-LSTM. The formulation of SAC-LSTM is as follows:

$$\begin{aligned} \mathcal{I}_t &= \sigma(W_{xi} * \mathcal{X}_t + W_{hi} * \mathcal{H}_{t-1} + W_{ci} \circ \mathcal{C}_{t-1} + b_i) \\ \mathcal{F}_t &= \sigma(W_{xf} * \mathcal{X}_t + W_{hf} * \mathcal{H}_{t-1} + W_{cf} \circ \mathcal{C}_{t-1} + b_f) \\ \mathcal{C}_t &= \mathcal{F}_t \circ \mathcal{C}_{t-1} + \mathcal{I}_t \circ tanh(W_{xc} \mathcal{X}_t + W_{hc} \mathcal{H}_{t-1} + b_c) \\ \mathcal{O}_t &= \sigma(W_{xo} \mathcal{X}_t + W_{ho} \mathcal{H}_{t-1} + W_{co} \circ \mathcal{C}_t + b_o) \\ \mathcal{H}_t &= \mathcal{O}_t \circ tanh(\mathcal{C}_t) \end{aligned} \quad (3)$$

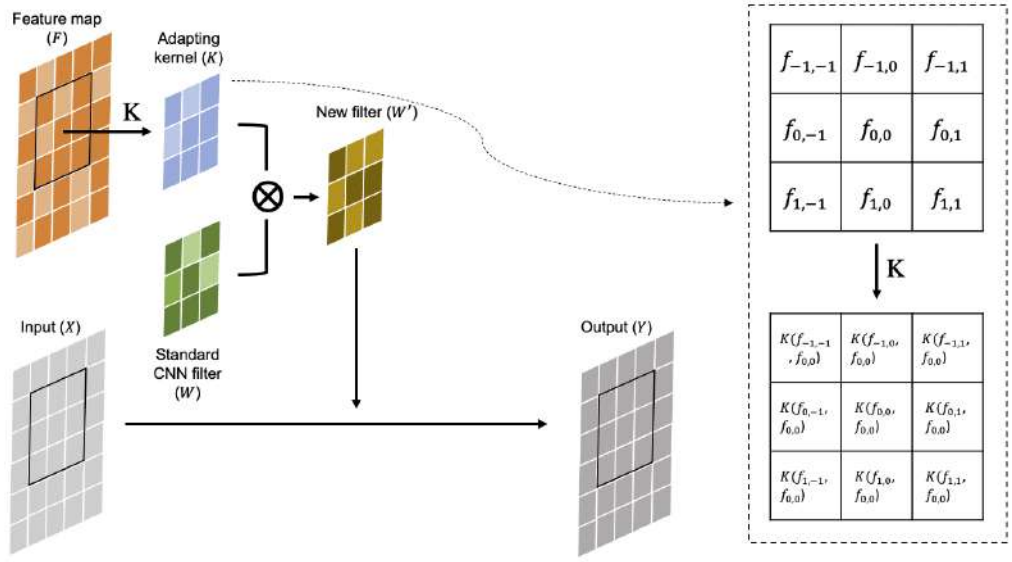


FIGURE 1 : Socially-Aware Convolution [57]

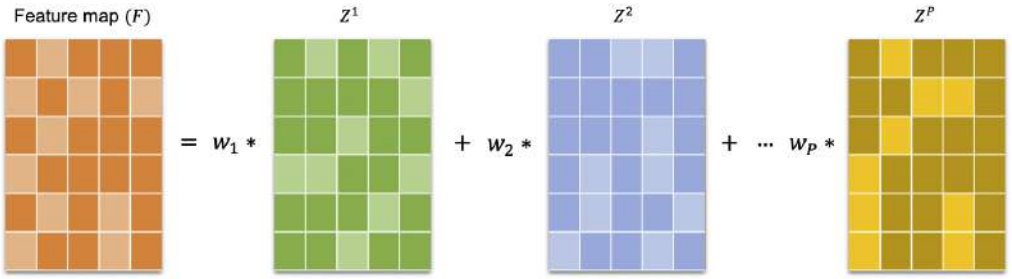


FIGURE 2 : Feature map (F) construction. Z_{ij}^{p} represents the value of the socio-demographic variable Z^p for pixel [i,j]

where the weight matrices $W_{xf}, W_{hf}, W_{xc}, W_{hc}, W_{xo}, W_{ho}$ denote the SAC weights, which are represented by W' in Figure 1. "*" stands for the convolutional operator. $\mathcal{I}_t$, $\mathcal{F}_t$, $\mathcal{C}_t$, $\mathcal{O}_t$, $\mathcal{H}_t$ are improved input gate, forgot gate, cell state, output gate and hidden state that embeds the spatial dependencies. ∘ denotes Hadamard product (i.e. element-wise product). σ and *tanh* are nonlinear activation functions:

$$\sigma(x) = \frac{1}{1+e^{-x}}; \quad tanh(x) = \frac{e^{x}-e^{-x}}{e^{x}+e^{-x}} \tag{4}$$

D. Model description

In this section, we propose a novel *socially aware network* (SA-Net) to forecast the short-term TNC demand. The architecture of the network is illustrated in Figure 3, which is comprised of two parts: the part on the right captures the spatial-temporal variables (i.e. TNC demand) using a stack of SAC-LSTM layers, and the part on the left processes the non-spatial temporal variables using a stack of LSTM layers.

1) Structure for spatial-temporal variables: We use a series of stacked SAC-LSTM layers to capture the spatial dependencies and temporal correlations for the spatial-temporal variable, which is the TNC demand data in our case. $\mathcal{D}_t$ is used to denote the TNC demand for time slot t. Let $\mathcal{C}$ denote a SAC-LSTM cell: $\mathcal{C}: R^{d*M*N*I} \rightarrow R^{d*M*N*O}$, where d denotes the look-back time window, which refers to the number of previous hours taken as predictors for the TNC demand in each time slot. M and N are the dimensions of rows and columns, I and O represent the number of channels for the input and output feature vectors. L denotes the number of stacked SAC-LSTM layers. The formulation of the model architecture that processes the TNC demand data is written as:

$$\begin{aligned} &(\mathcal{U}_{t-d}, \mathcal{U}_{t-d+1}, ..., \mathcal{U}_{t-1}) = \mathcal{C}_L ... \mathcal{C}_1 (\mathcal{D}_{t-d}, \mathcal{D}_{t-d+1}, ..., \mathcal{D}_{t-1}) \\ &\hat{\mathcal{X}}_t^u = W_{ux} * \mathcal{U}_{t-1} + b_u \end{aligned} \tag{5}$$

where $\mathcal{U}_{t-k}$, $k = 1, 2, ..., d$ represent the output tensors at the last layer of the stacked SAC-LSTM layers. W_{ux} represents the convolutional operation with the SAC kernel, which is applied to further capture the spatial dependency at the final layer, and also to reduce the number of output channel to 1.

In a similar manner, we employ a series of stacked SAC-LSTM layers to model the transit service data. Let $\mathcal{G}_t$ denote the level of transit service at time t. $\mathcal{G}_t$ is derived from the concatenation of B_t and R_t: $\mathcal{G}_t = [B_t, R_t]$, where B_t and R_t represent the number of bus and rail visits at time t. The model architecture that processes the transit service data can

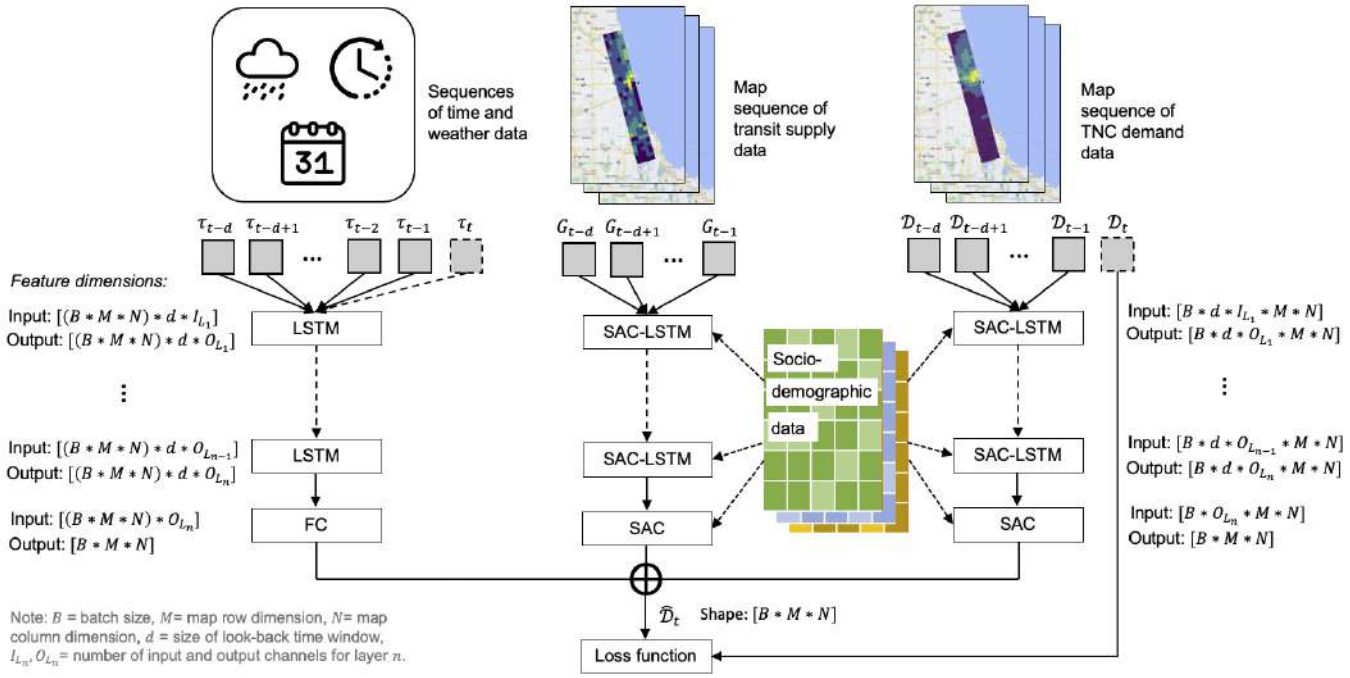


FIGURE 3 : The structure of SA-Net

be expressed as follows:

$$\begin{aligned}&(\mathcal{Q}_{t-d},\mathcal{Q}_{t-d+1},...,\mathcal{Q}_{t-1})=\mathcal{C}_L...\mathcal{C}_1(\mathcal{G}_{t-d},\mathcal{G}_{t-d+1},...,\mathcal{G}_{t-1})\\&\hat{\mathcal{X}}_t^q=W_{qx}*\mathcal{Q}_{t-1}+b_q\end{aligned}\tag{6}$$

2) Structure for temporal variables: The temporal predictors used in this study include the time-related variables and the weather feature. The time-related variables include time-of-day, day-of-week and holiday indicators. The weather feature is represented by the amount of precipitation. We create a new variable $v_t=(dow_t,tod_t,h_t)$ that concatenates dow_t, tod_t and h_t, and use p_t to represent the amount of precipitation at time t. These temporal features are likely to impact the TNC demand across the whole region. Then the network for the time-series variables can be written as follows:

$$\begin{aligned}&(V_{t-d},V_{t-d+1},...,V_{t-1})=\mathcal{L}_L...\mathcal{L}_1(v_{t-d},v_{t-d+1},...,v_{t-1},v_t)\\&\hat{\mathcal{X}}_t^v=\mathcal{F}^R(w_{vx}V_{t-1}+b_v)\\&(P_{t-d},P_{t-d+1},...,P_{t-1})=\mathcal{L}_L...\mathcal{L}_1(p_{t-d},p_{t-d+1},...,p_{t-1})\\&\hat{\mathcal{X}}_t^p=\mathcal{F}^R(w_{px}P_{t-1}+b_p)\end{aligned}\tag{7}$$

where V_{t-d} and P_{t-d}, $k=1,2,...,d$ are the output tensors at the last layer of the stacked LSTM layers for the time variables and the precipitation variable. w_{vx} and w_{px} denote the fully connected layers following the stacked LSTM layers, which reduce the number of output channel to 1. $\mathcal{F}^R$ denotes a reshaping function that repeat a value across the space: $\mathcal{F}^R: R\rightarrow R^{M*N*1}$, where $(\mathcal{F}^R)_{m,n,1}=x$ for any $m\in(1,2,...,M),n\in(1,2,...,N)$. $\mathcal{F}^R$ is deployed to make the dimensions of the LSTM outputs $\hat{\mathcal{X}}_v$ and $\hat{\mathcal{X}}_p$ the same with the SAC-LSTM output $\hat{\mathcal{X}}_u$.

3) Fusion: The final estimated TNC demand at time t is a weighted combination of the estimated outputs from different parts of the network, which is given by:

$$\hat{\mathcal{X}}_t=W_u\circ\hat{\mathcal{X}}_t^u+W_q\circ\hat{\mathcal{X}}_t^q+W_v\circ\hat{\mathcal{X}}_t^v+W_p\circ\hat{\mathcal{X}}_t^p\tag{8}$$

E. Accuracy and fairness metrics

The performance of the various models is evaluated based on two types of metrics: the accuracy metrics and the fairness metrics. Two commonly used accuracy metrics - Mean Absolute Error (MAE) and Mean Absolute Percentage Error (MAPE) - are adopted to evaluate the prediction accuracy of the models in this work. They are defined as below:

$$MAE=\frac{1}{N\times T}\sum_{t=1}^{T}\sum_{i=1}^{N}|y_t^i-\hat{y}_t^i|\tag{9}$$

$$MAPE=\frac{1}{T}\sum_{t=1}^{T}\frac{1}{|\mathcal{N}_t|}\sum_{i\in\mathcal{N}_t}|\frac{y_t^i-\hat{y}_t^i}{y_t^i}|,\ \mathcal{N}_t=\{i:1\le i\le N,\ y_t^i>0.1\}\tag{10}$$

where y_t^i and $\hat{y}_t^i$ are the real and predicted travel demands at time interval t in region i. T represents the total number of time intervals. N represents the total number of regions. $\mathcal{N}_t$ denotes the set of regions with $y_t^i>0.1$, which is defined to guarantee that the denominator of the absolute percentage error for the regions included is not zero.

While MAE and MAPE have been widely utilized to measure the accuracy of the model predictions, one limitation of these two metrics is that they do not consider the directions of the errors. Given that the underestimations and overestimations of the TNC demand predictions have very different practical

implications which should not be ignored, we also examine the Mean Percentage Error (MPE) of the model predictions which is given by:

$$MPE = \frac{1}{T}\sum_{t=1}^{T}\frac{1}{|\mathcal{N}_t|}\sum_{i\in\mathcal{N}_t}\frac{y_t^i-\hat{y}_t^i}{y_t^i},\ \mathcal{N}_t=\{i: 1\le i\le N,\ y_t^i>0.1\} \tag{11}$$

The positive value of MPE indicates the underestimation of the TNC demand (i.e. the real demand is larger than the predicted demand), whereas the negative value of MPE indicates the overestimation of the TNC demand. The magnitude of a positive percentage error in region i at time t can be thought of the chance of an individual in region i at time t who had the TNC demand but failed to receive the service, if the TNC service was exactly allocated based on the TNC demand estimation. Therefore, it is important to make sure that the MPE is not systematically different between the disadvantaged and privileged communities. This concept is connected to one important notion of algorithmic fairness – equality of odds, which states that a predictor $\hat{Y}$ satisfies equalized of odds with respect to protected attribute Z and outcome Y, if $\hat{Y}$ and Z are independent conditional on Y [53].

We propose the MPE gap as a fairness metric, which measures the difference of MPE between two groups (e.g. the black communities and the non-black communities). The metric is defined as:

$$\begin{aligned} MPE\ Gap = &\frac{1}{T}\sum_{t=1}^{T}\frac{1}{|\mathcal{N}_{t,z_0}|}\sum_{i\in\mathcal{N}_{t,z_0}}\frac{y_t^i-\hat{y}_t^i}{y_t^i} \\ &-\frac{1}{T}\sum_{t=1}^{T}\frac{1}{|\mathcal{N}_{t,z_1}|}\sum_{i\in\mathcal{N}_{t,z_1}}\frac{y_t^i-\hat{y}_t^i}{y_t^i} \\ s.t.\ &\mathcal{N}_{t,z_0}=\{i: 1\le i\le N,\ y_t^i>0.1,\ i\in\mathcal{Z}_0\}, \\ &\mathcal{N}_{t,z_1}=\{i: 1\le i\le N,\ y_t^i>0.1,\ i\in\mathcal{Z}_1\} \end{aligned} \tag{12}$$

where Z_0 denotes the minority group and Z_1 denotes the majority group. Therefore, $i\in\mathcal{Z}_0$ represents the set of regions that are within the minority group, and $i\in\mathcal{Z}_1$ represents the set of regions that are within the majority group (i.e. not within the minority group). For example, if the sensitive variable of interest is ethnicity and Z_0 is used to represent the black-dominated communities, then Z_1 represents the non-black communities. In this case, $i\in\mathcal{Z}_0$ and $i\in\mathcal{Z}_1$ denote regions that belong to the black communities and those that belong to the non-black communities, respectively.

To achieve a fair prediction, we want the absolute value of MPE gap to be as close to zero as possible. A positive value of MPE gap indicates that we are underestimating the TNC demand for the minority group compared with the majority group, whereas a negative value of MPE gap suggests a relative underestimation of the demand for the majority group.

F. De-biasing objective function

To jointly train for accuracy and fairness, we use a loss function that is a weighted sum of an accuracy loss and a fairness loss defined as below:

$$L = L_{accuracy} + \gamma L_{fairness} \tag{13}$$

The accuracy loss is aimed at reducing both MAE and MAPE:

$$L_{accuracy} = \sum_{t=1}^{T}\sum_{i=1}^{N}(y_t^i-\hat{y}_t^i)^2 + \lambda\sum_{t=1}^{T}\sum_{i\in\mathcal{N}_t}(\frac{y_t^i-\hat{y}_t^i}{y_t^i})^2, \tag{14}$$

$$s.t.\ \mathcal{N}_t=\{i: 1\le i\le N,\ y_t^i>0.1\} \tag{15}$$

where y_t^i and $\hat{y}_t^i$ are the real and predicted travel demands at time interval t in region i. T represents the total number of time intervals. N represents the total number of regions. $\mathcal{N}_t$ denotes the set of regions with $y_t^i>0.1$. λ is a regularization parameter balancing the MAE and MAPE tradeoff. In this study, we fix λ to be 10 since the magnitude of MAE is roughly ten times that of MAPE.

The fairness loss is proposed as the following:

$$L_{fairness} = |\sum_{t=1}^{T}\sum_{i\in\mathcal{N}_t}\tilde{z}^i * \frac{y_t^i-\hat{y}_t^i}{y_t^i}|,\ s.t.\ \tilde{z}^i=\frac{z^i-\bar{z}}{\sigma_z}, \tag{16}$$

$$s.t.\ \tilde{z}^i=\frac{z^i-\bar{z}}{\sigma_z},\ \mathcal{N}_t=\{i: 1\le i\le N,\ y_t^i>0.1\} \tag{17}$$

where z^i denotes the value of the sensitive attribute (e.g. the proportion of black population) for region i. $\tilde{z}^i$ is the normalized z^i with $\bar{z}$ and σ_z respectively representing the mean and standard deviation of z^i across all regions.

$L_{fairness}$ measures the linear relationship between the sensitive attribute z and MPE across time and space. To be specific, $\tilde{z}^i * \frac{y_t^i-\hat{y}_t^i}{y_t^i}$ measures the joint deviations of $\tilde{z}^i$ and $\frac{y_t^i-\hat{y}_t^i}{y_t^i}$ from zero. Therefore, $L_{fairness}$ indicates the covariance between z and MPE in the prediction, which we want to penalize in our training process.

V. Experiments

A. Data Description

The dataset utilized in this paper is a large-scale TNC trip record dataset collected from Chicago Data Portal [59] during a 14-month period between November 1st, 2018 to December 23rd, 2019. The trip records that started from 6 AM to 10 PM are included. We partition the city of Chicago into $1km\times 1km$ grids, and use totally 35×5 grids for analysis as shown in Figure 4. The hourly TNC demand in a region is represented by the number of trips starting from that region in a 1-hour time interval. The weather data is collected from the website of National Centers for Environmental Information [54]. The socio-demographic variables including the percentages of black population and the percentage of low-income population are extracted from the 2019 American Community Survey(ACS) 5-year estimates [60].

Figure 4 illustrates the distributions of average hourly TNC demand in the study period, the percentage of black population and the percentage of low-income population in the study area. From Figure 4a, we can see that the spatial distribution of the TNC demand is highly uneven, as the downtown area takes up the majority of the TNC demand. In terms of ethnicity, Figure 4b reveals a bimodal distribution of African-American population, with the majority of the northern area having African-American population below 13% and the majority of the southern area having African-American population above 88%. We define population with 2019 household income lower than $25,000 as low-income, and Figure 4c shows that the low-income population is also mainly clustered in the south side of the study area. In this study, we define grids with over 50% of black population as the black communities, and the rest as the non-black communities, which gives us 73 black communities and 102 non-black communities. Regarding income, we defined grids with more than 25% of low-income population as the low-income communities, and the rest as the high-income communities, resulting in totally 90 low-income communities and 85 high-income communities. In both cases, the numbers of disadvantaged and privileged communities are roughly balanced.

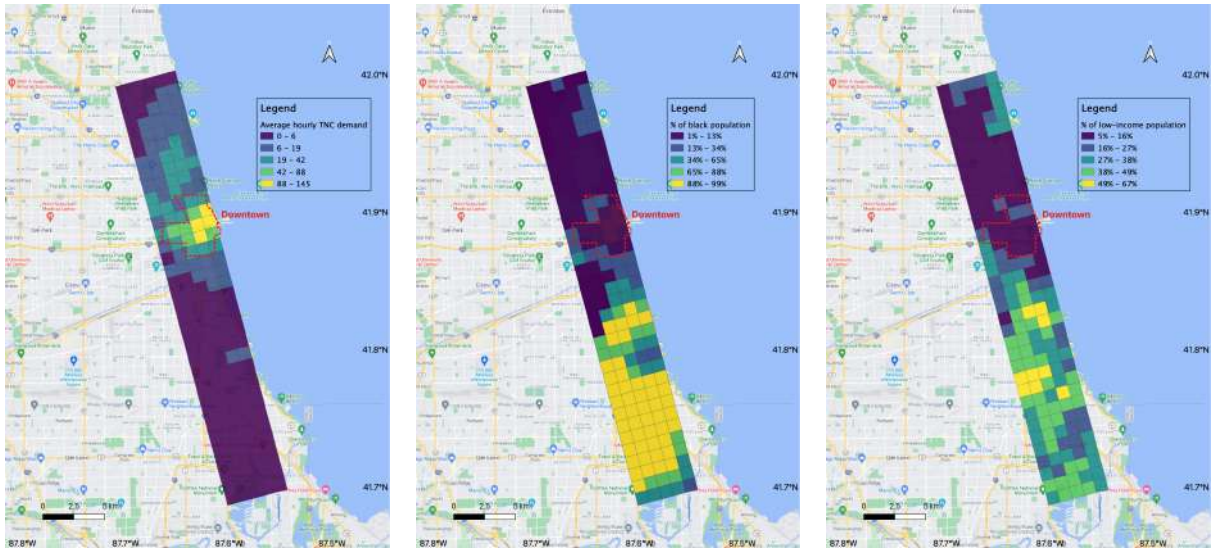
Figure 5 illustrates the average TNC travel demand by time of day, separated by the disadvantaged (black/low-income) and privileged (non-black/high-income) communities. The travel demand in the privileged regions are much larger than that in the disadvantaged regions, therefore the y-axis scales are different in Figure 5(a) and Figure 5(b). The privileged regions and the low-income communities have two peak periods: 7 AM - 10 AM and 5 PM - 8 PM, whereas the black communities only has the morning peak.

In the experiment, we apply a 70-30 training-testing split. The data from November 1st, 2018 to July 23rd, 2019 (265 days) is used for training, the data from July 24th, 2019 to August 21st, 2019 (29 days) is used for validation and the data from August 22nd, 2019 to December 23rd, 2019 (124 days) is used for testing. In the training, validation and testing processes, we use the TNC demand in the previous 6 hours to predict the TNC demand in the next hour (i.e. the look-back window is 6 hours). In addition, we assess the robustness of our modeling results by utilizing data aggregated into 30-minute intervals. In this new prediction task, we leverage the TNC demand observed in the preceding 3 hours to forecast the TNC demand for the subsequent half hour. The findings from this analysis are presented in Appendix A. Before training the models, the collected data is normalized by z-score process to facilitate training. We later denormalize the prediction to get the actual demand values, and reset the negative values to zeros since the demand values cannot be negative.

B. Model Comparison

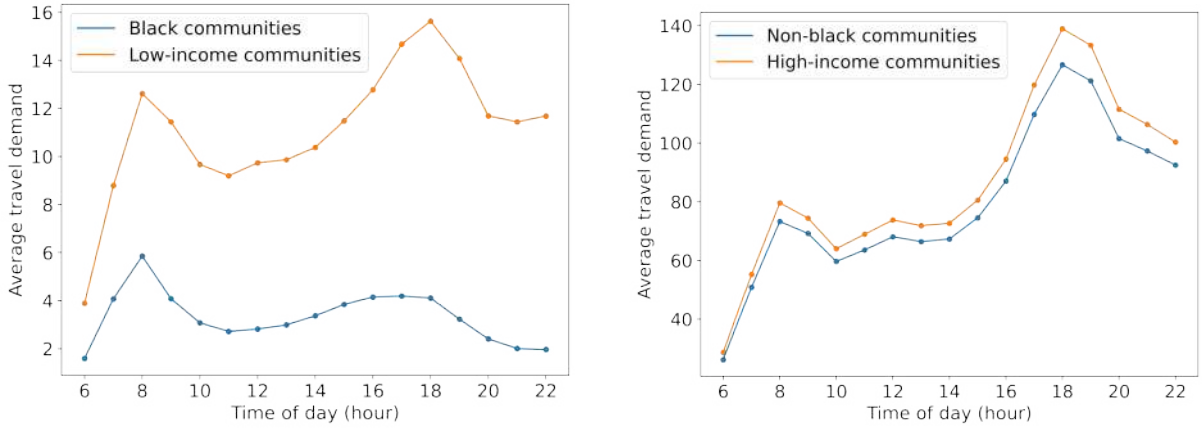
To explore the advantage of our model SA-Net, we compare it against several other benchmark models, which are listed as follows:

- **Historical Average (HA)**: HA predicts the TNC demand by averaging the historical demand which is in the same relative time interval (i.e. the same time of day and the same day of week) in the training set. For instance, the TNC demand in Monday 10 AM -11 AM is predicted as the average TNC demand of all past Monday's at 10 AM -11 AM in the training set.
- **Moving Average (MA)**: MA predicts the TNC demand by averaging the demand in the same relative time interval of several nearest historical values. We use the average of 6 previous TNC demand in grid (i, j) to predict the demand in grid (i, j).
- **Autoregressive Integrated Moving Average Model (ARIMA)**: ARIMA is commonly used for forecasting time-series data [61], and has been widely applied in traffic prediction problems[62, 63]. In this work, to predict the TNC demand in grid (i, j), the inputs to ARIMA were 6 previous demand in the same relative time interval in grid (i, j).
- **LSTM Net**: The LSTM Net processes the TNC demand in each grid separately. The hyperparameters and the structures of the LSTM Net and the SA-Net are the same. The only difference is that while we use a stack of SAC-LSTM to process the spatial-temporal data as shown in Figure 3, the LSTM Net uses the LSTM modules to processes the TNC demand data and does not capture spatial dependencies.
- **LSTM + Social Net**: The LSTM + Social Net adds a socio-demographic feature map to the LSTM Net to facilitate predictions. The feature map is constructed as a linear combination of different socio-demographic variables as shown in Figure 2, and is fused with other parts of the network in the last model layer following Equation 8.
- **Spatiotemporal Graph Convolution Network (STGCN)** [64]: STGCN is a widely compared state-of-the-art model for spatial temporal modeling. It applies a GCN-based method to model spatial correlations with spectral-based GCNs and captures temporal dependencies with temporal convolution layers (TCNs).
- **Graph Wavenet [65]**: Graph Wavenet follows a similar GCN and gated TCNs frameworks as STGCN but with adaptive dependency matrix learning. The model performs well in the traffic forecasting task with high accuracy and fast convergence speed.
- **Conv-LSTM Net**: The Conv-LSTM Net is a fusion convolutional LSTM specified in [3]. The hyperparamters and the structure of the Conv-LSTM Net are the same with the SA-Net and the LSTM Net. The difference is that the Conv-LSTM Net uses the traditional Conv-LSTM modules instead of the SAC-LSTM modules in Figure 3 to process the spatial-temporal TNC demand data.
- **Conv-LSTM + Social Net**: Similar to the LSTM + Social Net, the Conv-LSTM + Social Net adds a socio-demographic feature map to the Conv-LSTM Net to facilitate predictions.



(a) Average hourly TNC demand (b) Percentage of black population (c) Percentage of low-income population

FIGURE 4 : Distributions of TNC demand, black population and low-income population in the study area



(a) Disadvantaged regions (b) Privileged regions

FIGURE 5 : Average TNC travel demand by time of day

C. Experiment setup

When training Conv-LSTM Net and SA-Net, we use kernels with size of 3×3. Each Conv-LSTM cell and each SAC-LSTM cell consists of 32 filters/channels to capture the spatial information. The experiments are implemented in Pytorch using the mini-batch stochastic gradient descent method with a batch size of 64 and a step size of 0.001 in each training. The model that produces the lowest prediction loss on the validation set among the 300 epochs is chosen. To determine the best performing model, both Conv-LSTM Net and SA-Net are trained with 1, 2, and 3 layers, selecting the model that yields the lowest prediction loss. For STGCN, the model is trained with graph convolution kernel sizes of 2 and 3, as well as temporal kernel sizes of 2 and 3. For Graph Wavenet, we utilize the open-source codes provided by the original authors, employing their default parameter settings. The model with the lowest prediction loss is chosen as the optimal model, which is then used for prediction on the test data. The optimal model later performs prediction over the test data. We run the training procedure 3 times and report the average prediction results on the test set.

D. Results

We compare our proposed algorithms (SA-Net with bias-mitigation regularization) with baseline models along two dimensions: accuracy and fairness, and show that our algorithm achieves better results regarding both accuracy and fairness. The better prediction accuracy is demonstrated by lower MAE and MAPE compared with baseline models. Our proposed bias mitigation strategy demonstrates improved prediction fairness by effectively reducing the MPE gap between disadvantaged and privileged groups. Importantly, this reduction in bias does not compromise the overall prediction accuracy, allowing us to achieve fairness in our predictions. In the following sections, we present the results obtained from data aggregated to one-hour intervals. The corresponding results obtained from data aggregated to

30-minute intervals are included in the Appendix A.

1) Prediction accuracy: The spatial-temporal deep learning algorithms (i.e. STGCN, Wavenet, Conv-LSTM Net, Conv-LSTM + Social Net and SA-Net) outperform the classical statistical models (i.e. HA, MA, ARIMA), and our proposed SA-Net model produces the smallest overall MAE, RMSE, and MAPE among all models on the test set. Table I presents the overall MAE, RMSE, and MAPE, along with the MAE specifically for black, non-black, low-income, and high-income communities when the data is aggregated to one-hour intervals. Regarding the overall MAE, RMSE, and MAPE, Conv-LSTM Nets and SA-Net significantly outperform other models. Conv-LSTM + Social Net shows a slight improvement over Conv-LSTM Net in terms of MAE and RMSE, but performs worse in terms of MAPE. This finding suggests that incorporating socio-demographic variables as predictors does not significantly enhance the model's performance. Next, we compare the results of Conv-LSTM Net with SA-Net.

When comparing Conv-LSTM Net and SA-Net, we observe that SA-Net effectively decreases MAE for both the black and non-black communities. Furthermore, it also reduces MAE for both the low-income and high-income communities. These findings suggest that the inclusion of socio-demographic information in SA-Net yields benefits for both disadvantaged and privileged groups.

SA-Net improves prediction accuracy for the black communities at all times of day compared with Conv-LSTM Net. We examine the model performance for Conv-LSTM Net and SA-Net across different times of day in Figure 6. The upper row of Figure 6 shows the predictive results for the black communities, whereas the bottom row of Figure 6 shows the results for the non-black communities. Figures 6a, 6b, 6d, and 6e illustrate that SA-Net yields lower MAE during the morning peak hours (6 PM - 8 PM), as well as reduced MAPE across all times of the day, compared to Conv-LSTM Net for the black and non-black communities.

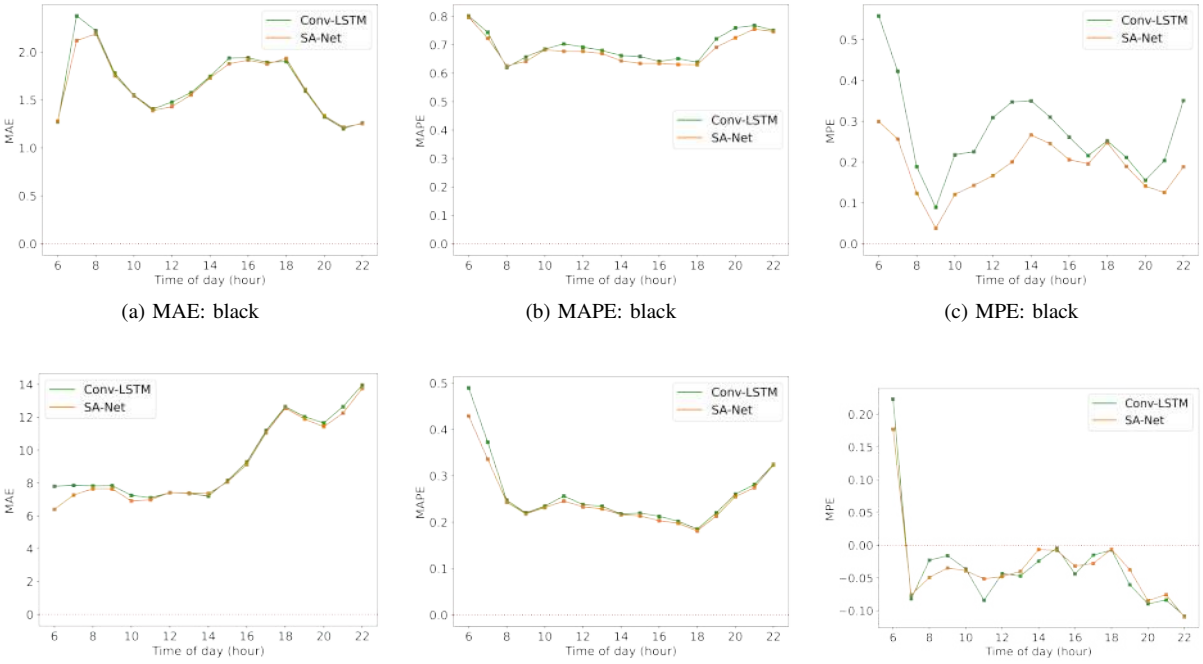
Figure 6c shows the MPE for the black communities at various times of day. The MPEs are consistently positive, indicating that both Conv-LSTM Net and SA-Net underpredict the travel demand for the black communities at different times of day. However, SA-Net consistently gives smaller MPEs than Conv-LSTM Net, showing that the former model reduces the magnitude of the underprediction of the black communities' travel demand. The MPEs for the non-black communities with Conv-LSTM Net and SA-Net are more similar at different times of day as shown in Figure 6f.

2) Prediction fairness: Having demonstrated the superiority of our proposed SA-Net over the benchmark models in terms of prediction accuracy, we now test the effectiveness of our bias mitigation strategy stated in Section IV-F for fairness improvement. First, we test the results when the sensitive attribute is race, namely when z denotes the proportion of black population in Equation 16. Table II presents the results, and Figures 7a and 7b plot MPE for black and non-black groups as well as the overall MAE. Table II shows that when the de-biasing regularizer is not applied ($\gamma = 0$), both Conv-LSTM Net and SA-Net produce large MPE gaps between black and non-black groups. Specifically, the MPE gap (race) with $\gamma = 0$ is 0.306 for Conv-LSTM Net, whereas the MPE gap (race) with $\gamma = 0$ is 0.224 for SA-Net. For both models, the large MPE gap comes from a large, positive MPE for the black group and a small, negative MPE for the non-black group. Note that the magnitude of a positive MPE indicates the degree of underestimation of the demand, since MPE represents the average gap of the actual and predicted demand weighted by the actual demand. Larger the MPE, higher the underestimation. Therefore, the large MPE gaps between black and non-black groups indicate that training models using the traditional objective function without bias mitigation leads to systematic underestimation for the black group compared with the non-black group.

Recognizing the prediction bias using only $L_{accuracy}$ in training, we adopt bias mitigation by increasing the bias mitigation weight γ from 0 to 5 and 10. The results for "MPE gap (race)" in Table II show that for both models, as γ increases, the MPE gap between black and non-black groups decreases, and this reduction in MPE gap mainly stems from the reduction in MPE for the black group. Specifically, when increasing γ from 0 to 10, the MPE gap between the black and non-black groups drops from 0.306 to 0.040 for Conv-LSTM Net, and drops from 0.224 to 0.074 for SA-Net. It is also found that by mitigating the racial bias, the MPE gap between the low-income and high-income groups has also been reduced, probably because most low-income and black communities are clustered in the south side of Chicago (Figure 4), thus by mitigating bias for race, the prediction bias (MPE gap) for income has been reduced simultaneously.

Figures 7a and 7b plot MPE for black and non-black groups as well as the overall MAE corresponding to Table II. As we increase the bias mitigation weight (γ), the prediction MPEs for the black population (denoted by the green bars) decrease considerably, indicating that with the bias mitigation loss function, the underestimation of TNC demand for the black population has been mitigated. Additionally, the MAE also decreases as the bias mitigation weight increases.

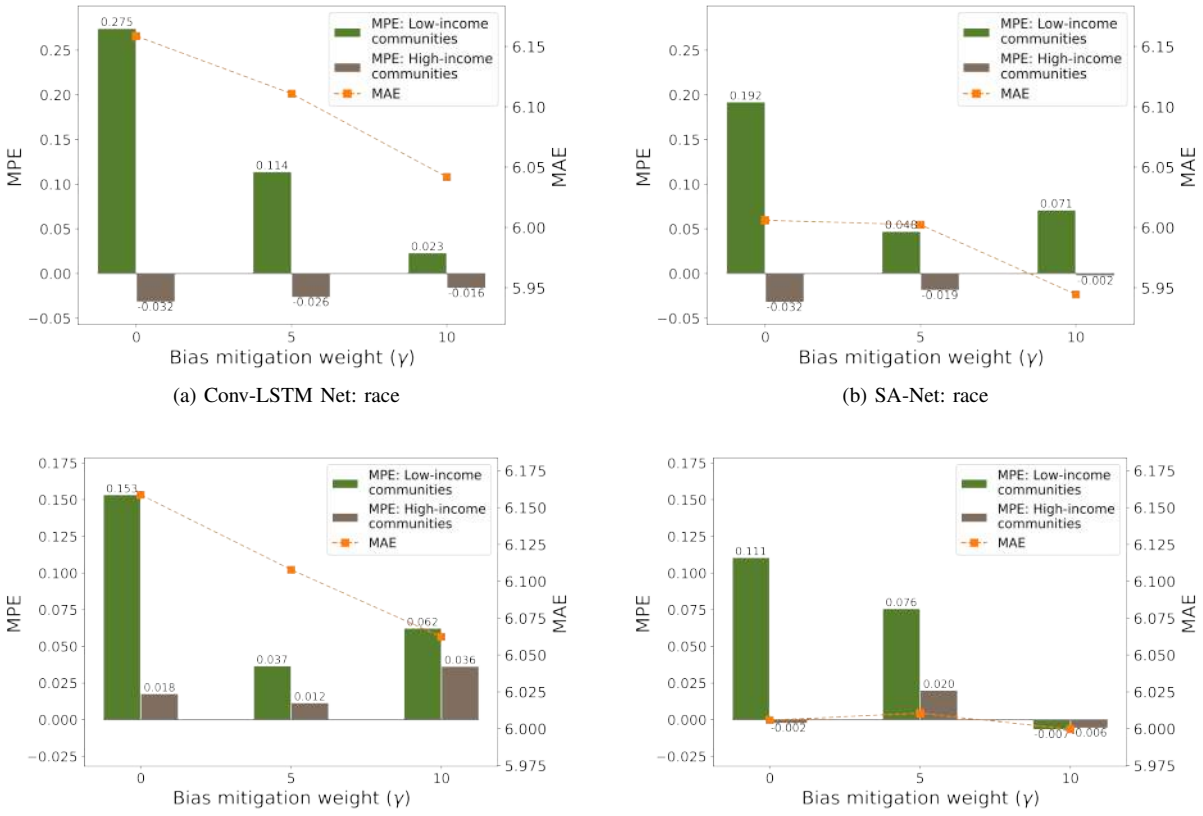
Then, we apply the same bias mitigation strategy to mitigate the MPE gap between the low-income and high-income groups. In this case, z denotes the proportion of low-income population in Equation 16. The results for "MPE gap (income)" in Table III show that similar to the bias mitigation results for race, the MPE gaps generally decline as γ increases when mitigating the income bias for both Conv-LSTM Net and SA-Net. Specifically, when γ increases from 0 to 10, the MPE gap between the low-income and high-income groups decreases from 0.136 to 0.026 for Conv-LSTM Net, and decreases from 0.113 to -0.001 for SA-Net. The MPE gaps between two income groups are also plotted in Figure 7c and



(a) MAE: black (b) MAPE: black (c) MPE: black

(d) MAE: non-black (e) MAPE: non-black (f) MPE: non-black

FIGURE 6 : Performance measures by model and time of day



(a) Conv-LSTM Net: race (b) SA-Net: race

(c) Conv-LSTM Net: income (d) SA-Net: income

FIGURE 7 : Performance measures by model and sensitive variable, corresponding to Table II and Table III

TABLE I : Accuracy comparisons among different models

	MAE	RMSE	MAPE	Black (MAE)	Non-black (MAE)	Low-income (MAE)	High-income (MAE)
HA	8.00	21.46	0.64	1.55	12.61	6.53	12.11
MA	7.93	22.04	0.55	1.41	12.60	6.49	11.96
ARIMA	7.16	19.81	0.54	**1.36**	11.31	5.93	10.60
LSTM Net	8.97	18.51	0.54	2.23	13.80	7.14	14.13
LSTM + Social Net	8.83	18.53	0.49	2.05	13.69	7.00	13.97
STGCN	6.59	14.26	0.59	1.58	10.17	5.27	10.29
Wavenet	6.30	14.53	0.50	1.63	9.64	5.09	9.69
Conv-LSTM Net	6.16	13.16	0.42	1.67	9.37	4.99	9.43
Conv-LSTM + Social Net	6.15	13.11	0.44	1.72	9.31	5.02	9.31
SA-Net	**6.01**	**13.00**	**0.41**	1.65	**9.13**	**4.92**	**9.06**

Note: for the deep learning models, we report the results when the models are trained with $\gamma = 0$

TABLE II : Fairness and accuracy comparisons with bias mitigation for race

	MAE	RMSE	MAPE	MPE gap (race)	Black (MPE)	Non-black (MPE)	MPE gap (income)	Low-income (MPE)	High-income (MPE)
Conv-LSTM Net:									
$\gamma = 0$	6.159	**13.163**	0.420	0.306	0.275	-0.032	0.136	0.153	0.018
$\gamma = 5$	6.111	13.278	0.409	0.141	0.114	-0.026	0.041	0.048	0.006
$\gamma = 10$	**6.042**	13.265	**0.408**	**0.040**	**0.023**	**-0.016**	**-0.016**	**-0.010**	**0.006**
SA-Net:									
$\gamma = 0$	6.006	**12.998**	0.408	0.224	0.192	-0.032	0.113	0.111	-0.002
$\gamma = 5$	6.002	13.190	**0.406**	**0.066**	**0.048**	-0.019	0.008	**0.010**	**0.002**
$\gamma = 10$	**5.944**	13.114	0.411	0.074	0.071	**-0.002**	**0.005**	0.028	0.023

Note: γ represents the bias mitigation weight

Figure 7d, where we can see that the de-biasing regularization works well to reduce the MPE gap for both Conv-LSTM Net and SA-Net.

In addition, we find that the improving prediction fairness does not necessarily sacrifice prediction accuracy. The orange dots in Figure 7 denote the MAEs produced by different models, which show that the application of bias mitigation actually also brings down MAE. Notably, when increasing the mitigation weight γ for income from 0 to 10 for SA-Net, the prediction accuracy has been greatly improved (MAE=6.006) compared with the case when no bias mitigation is adopted (MAE=5.944).

We also examine the change of average MPE in different times of day with different bias mitigation strategies in Figure 8. Figure 8a and 8d show that by increasing the bias mitigation weight γ from 0 to 5 and 10, the MPE for the black communities decreases in all times of day for both the Conv-LSTM Net and the SA-Net, and the bias mitigation effect is slightly stronger in the Conv-LSTM Net case. For the Conv-LSTM Net, we observe that when $\gamma = 10$, the morning peak MPE decreases to around zero, and the MPE in the morning and evening peak periods becomes negative. On the contrary, Figure 8b and 8e show that the effects of the bias mitigation method on the MPE for the non-black communities are relatively small. The increase of γ is associated with a small drop of MPE in the Conv-LSTM Net case and a small rise of MPE in the SA-Net case. Figure 8c and 8f plot the gaps in MPE between the black and non-black communities given by Conv-LSTM Net and SA-Net, which show that for all times of day, increasing the bias mitigation weight from zero reduces the MPE gap between the black and non-black communities. All in all, our results suggest that our proposed bias mitigation strategy can significantly mitigate the travel demand underprediction issue for the black communities in all times of day with both the Conv-LSTM Net and the SA-Net, and can effectively reduce the prediction bias between the black and non-black groups.

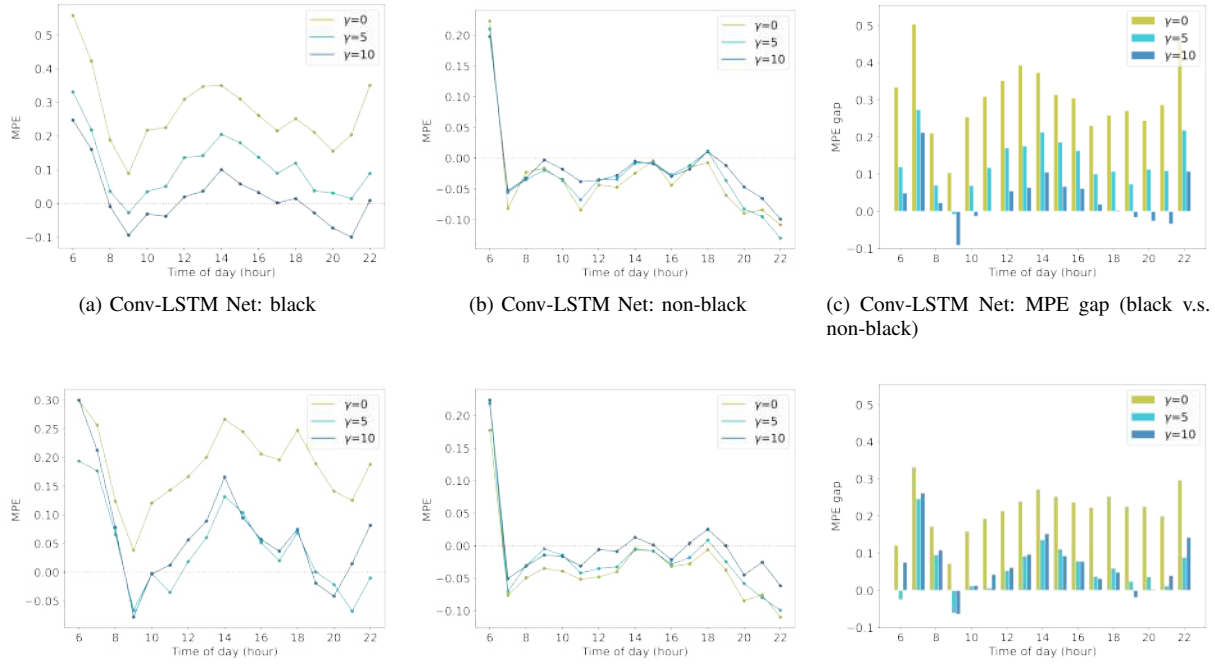
In summary, our proposed de-biasing regularization method can considerably reduce the prediction bias measured by the MPE gap between the disadvantaged and the privileged groups for both Conv-LSTM Net and SA-Net. This gain in prediction fairness can be achieved while keeping the prediction accuracy high. For SA-Net, adopting bias mitigation can even increase prediction accuracy.

3) Spatial patterns of errors: To better understand the spatial heterogeneity of the prediction errors, we show in Figure 9 the spatial distributions of MPE using SA-Net for three prediction strategies: prediction with no bias-mitigation, with race bias mitigation ($\gamma = 5$) and with income bias mitigation ($\gamma = 5$). Areas with positive MPE (indicating that

TABLE III : Fairness and accuracy comparisons with bias mitigation for income

	MAE	RMSE	MAPE	MPE gap (race)	Black (MPE)	Non-black (MPE)	MPE gap (income)	Low-income (MPE)	High-income (MPE)
Conv-LSTM Net:									
$\gamma = 0$	6.159	**13.163**	0.420	0.306	0.275	-0.032	0.136	0.153	0.018
$\gamma = 5$	6.108	13.423	0.411	**0.142**	**0.112**	-0.030	**0.025**	**0.037**	**0.012**
$\gamma = 10$	**6.062**	13.352	**0.401**	0.174	0.158	**-0.016**	0.026	0.062	0.036
SA-Net:									
$\gamma = 0$	6.006	**12.998**	0.408	0.224	0.192	-0.032	0.113	0.111	**-0.002**
$\gamma = 5$	6.010	13.150	**0.407**	0.192	0.167	**-0.025**	0.056	0.076	0.020
$\gamma = 10$	**6.000**	13.112	0.413	**0.108**	**0.061**	-0.047	**-0.001**	**-0.007**	-0.006

Note: γ represents the bias mitigation weight



(a) Conv-LSTM Net: black (b) Conv-LSTM Net: non-black (c) Conv-LSTM Net: MPE gap (black v.s. non-black)

(d) SA-Net: black (e) SA-Net: non-black (f) SA-Net: MPE gap (black v.s. non-black)

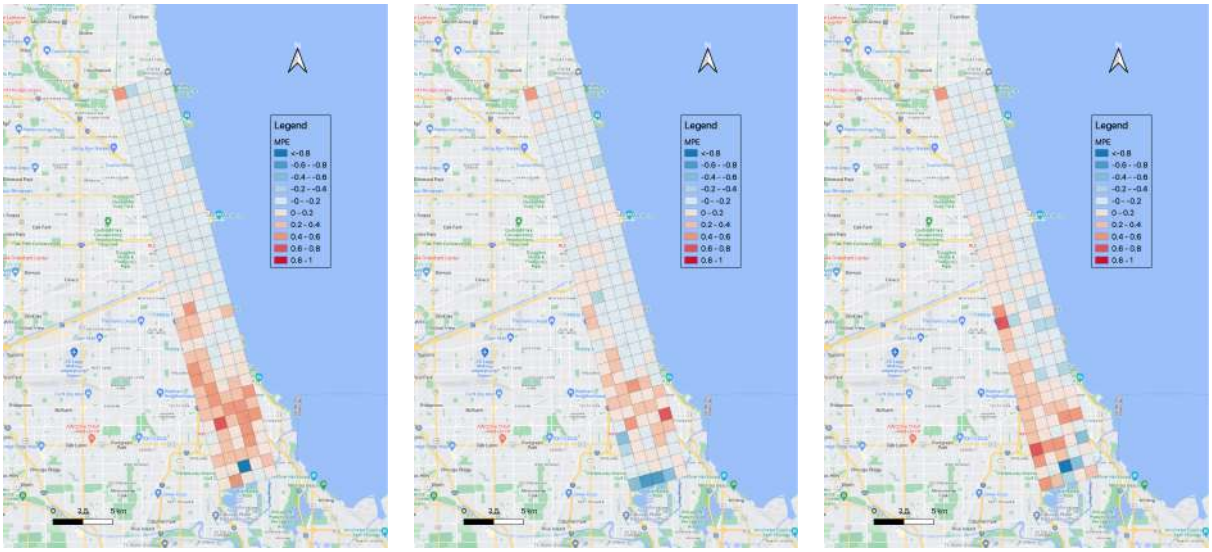
FIGURE 8 : MPE for different racial groups with different mitigation weights (γ) by time of day

the TNC demand has been underestimated) are denoted by the red color, whereas areas with negative MPE (indicating demand overestimation) are denoted by the blue color. Figure 9a shows that when no bias mitigation is adopted, the south side of the study area, which has greater populations of low-income and African-American people, suffers from severe demand underestimation. When we add the bias mitigation for race and income, the results in Figure 9b and Figure 9c show that the grid colors in the southern areas have become much lighter, and the colors of several areas in the south switch from red to blue, suggesting that the underestimation issue has been remarkably alleviated.

VI. Ablation analysis

To gain a deeper insight into the performance of SA-Net concerning the different employed features, we conduct an ablation analysis on each feature group. In this series of experiments, we systematically remove individual feature groups from the complete set and evaluate the resulting performance. Common intuition suggests that removing a crucial feature group would result in a notable decline in performance. Figure IV shows that transit service supply is the most important feature regarding MAE and RMSE, as removing this feature will lead to the largest increase in MAE and RMSE. Employment is the second important feature in terms of contribution to MAE. Removing employment will also lead to the largest increase in MAPE.

It can be observed that each feature contributes to the prediction, as none of them individually outperforms the model with the complete feature set. Furthermore, when different ablation strategies are applied, MAEs for black, non-black, low-income, and high-income groups are similar in scale.



(a) No bias mitigation (b) Bias mitigation for race (c) Bias mitigation for income

FIGURE 9 : Spatial distributions of mean percentage errors using SA-Net for different bias mitigation strategies

TABLE IV : Ablation analysis of feature set

Feature group	MAE	RMSE	MAPE	Black (MAE)	Non-black (MAE)	Low-income (MAE)	High-income (MAE)
Transit service supply	6.28	13.56	0.41	1.71	9.55	5.14	9.49
Employment	6.13	13.09	0.43	1.74	9.27	5.00	9.30
Ethnicity	6.08	13.49	0.39	1.63	9.26	4.96	9.21
Gender	6.07	13.04	0.42	1.64	9.24	4.92	9.30
Income	6.04	12.82	0.41	1.67	9.17	4.93	9.16
Population	6.04	12.90	0.41	1.71	9.14	4.93	9.15
Precipitation	6.04	13.16	0.41	1.68	9.16	4.91	9.19
Travel modes	6.03	13.06	0.41	1.63	9.18	4.91	9.18
Age	6.03	13.17	0.40	1.65	9.17	4.92	9.15
Time indicators	6.01	12.86	0.41	1.68	9.11	4.92	9.06
Foreign	6.01	13.15	0.40	1.63	9.15	4.91	9.11
Education	6.01	13.14	0.40	1.63	9.15	4.90	9.12

Note: "Population" includes total population and population per squared kilometers; "Foreign" includes the percentage of spanish speakers and the percentage of foreign-born population; "Travel modes" include the percentage of transit commuters and the percentage of population with no household vehicles. "Time indicators" include time-of-day, day-of-week, and holiday indicators.

VII. Sensitivity analysis

In this section, we perform a sensitivity analysis and parameter tuning on SA-Net. We investigate three types of parameters: the bias mitigation weight (λ), the kernel size, and the number of output channels (O_{L_n}). Our benchmark model is set with λ = 10, kernel size = 3, and 32 filters. For each trial, we vary the value of one of these three parameters while keeping the other two parameters the same as the benchmark model. The results on the validation set are reported in Table V for reference.

When examining the parameter λ, we observe that as λ increases, the RMSE also increases, while MAPE generally decreases. After careful consideration, we choose λ = 10 as it strikes a good balance between MAE and MAPE. For the kernel size, we evaluate both 3 and 5 and find that a kernel size of 3 yields better results in terms of MAE and MAPE. Regarding the number of filters in the convolutional units, we discover that using 32 filters produces the best outcome in terms of MAE and RMSE. Overall, the results demonstrate that our benchmark model provides the optimal performance based on the metrics evaluated.

VIII. Discussion and conclusions

Fairness has long been a critical concern in transportation studies. However, the prediction fairness issue in spatial-temporal travel demand forecasting has been neglected in previous literature. In this paper, we propose a two-pronged approach to enhance fairness in TNC demand forecasting, and test the effectiveness of our innovative method on Chicago's TNC data.

First, though previous studies have shown that there is significant difference in travel characteristics across demographic groups [66, 67, 68], most spatial-temporal research failed to account for the socio-demographic heterogeneity in travel demand predictions. We introduce a novel model structure, SA-Net, designed to effectively

TABLE V : Sensitivity analysis

	MAE	RMSE	MAPE		MAE	RMSE	MAPE
λ:				Kernel size= 5	5.75	**13.43**	0.416
λ = 0	**5.646**	**13.333**	0.612	*Number of output channels* O_{L_n}:			
λ = 5	5.726	13.464	0.424	O_{L_n} = 8	5.73	13.70	0.396
λ = 10	5.669	13.507	0.399	O_{L_n} = 16	5.79	13.87	0.395
λ = 100	5.945	13.996	**0.368**	O_{L_n} = 32	**5.67**	**13.51**	0.399
λ = 1000	6.413	14.135	0.399	O_{L_n} = 64	5.82	13.59	0.416
Kernel size:				O_{L_n} = 128	5.68	13.53	**0.389**
Kernel size= 3	**5.669**	13.507	**0.399**				

capture spatial correlations variations across various socio-demographic groups. Experimental results affirm that our proposed SA-Net substantially improves overall prediction accuracy by effectively integrating socio-demographic and contextual information, such as transit service information, weather data, and time indicators. To achieve this, the SA-Net incorporates the Socially-Aware Convolution (SAC) module, which adapts the standard invariant kernel based on the socio-demographic composition of each area within the study region.

Specifically, the SA-Net demonstrates notable improvements in performance compared to the best benchmark model, Conv-LSTM Net, across various evaluation metrics. During the morning peak hours (6 AM - 8 AM), SA-Net achieves lower MAE, and it consistently exhibits reduced MAPE throughout the day. Furthermore, when examining the predictions for both black and non-black communities, SA-Net yields MPEs that are closer to zero than Conv-LSTM Net, indicating that it effectively mitigates underestimation for the black communities. Additionally, our findings underscore the significance of transit service level and employment information as the most influential features contributing to the enhanced prediction accuracy achieved by SA-Net.

Second, in our analysis, we uncover a prevalent issue in previous solutions for spatial-temporal travel demand prediction: the tendency to underestimate demand in low-demand regions. This arises due to the significant impact of high errors in these areas on MAPE, which results in a bias towards optimizing MAPE at the expense of accurate predictions in low-demand regions. To address this challenge, we propose a novel approach that employs the mean percentage error gap as a measure of prediction fairness. Additionally, we introduce a regularization method designed to mitigate bias between disadvantaged and privileged groups. This is accomplished by disentangling the correlation between the sensitive attribute and the mean percentage error.

Our experimental results strongly confirm the efficacy of the new algorithm in effectively mitigating prediction bias for both the traditional Conv-LSTM Net and the newly proposed SA-Net. Moreover, our findings demonstrate that the method excels at protecting disadvantaged regions against systematic underestimation. Specifically, our proposed bias mitigation strategy significantly mitigates the issue of travel demand underprediction for the black communities across all times of day, thus ensuring fairer predictions.

Overall, we argue that the prediction bias issue revealed in this work should attract the attention of the researchers and policy makers, because if the travel demand in the disadvantaged neighborhoods is systematically underpredicted, we may fail to provide enough TNC services to these communities, and the limited services will in turn lead to further decrease of the ridership, which will eventually lead to a negative feedback loop. The method proposed in this study has been proven to be capable of tackling this prediction bias issue and promoting both accuracy and fairness. This has practical implications for policymakers and transportation planners, as it helps ensure equitable access to TNC services for all communities, including those historically underserved or underpredicted. This can contribute to reducing transportation inequities and enhancing social inclusion.

We identify several future research directions worth investigating. First, this paper evaluates fairness in travel demand prediction and demonstrates the utility of the de-biasing mitigation method on Conv-LSTM Net and the SA-Net. However, the proposed fairness evaluation metrics and the bias mitigation method are widely applicable. They can also be applied to other spatial-temporal deep learning networks such as the spatial-temporal residual networks ST-ResNet [69] and RSTN [10]. Second, this study aims to implement fair predictions for on-demand ride service. However, our proposed fairness-enhancing method should also work well for other spatial-temporal settings, such as bikeshare demand prediction, public transport demand prediction and crime incidents prediction. Future research can test the performance of the proposed method on various downstream applications. Third, we test our method on Chicago's TNC data as the real-world application. Future work can study the transferability of our method to other applications or cities.

APPENDIX A
EXPERIMENTS WITH DATA AGGREGATED TO 30-MINUTE INTERVALS

In this section, we test the robustness of our results by conducting experiments on data aggregated into 30-minute intervals. Table VI presents the overall MAE, RMSE, and MAPE, along with the MAE specifically for different social groups. Consistent with the outcomes obtained from data aggregated into one-hour intervals, our findings indicate that SA-Net consistently achieves the lowest overall MAE and MAPE. This underscores the enhanced prediction accuracy

achieved by employing SA-Net.

We proceed to present the results with bias mitigation for race in Table VII, and the results with bias mitigation for income in Table VIII. The findings demonstrate the effectiveness of the two bias mitigation strategies in addressing bias related to the corresponding sensitive attributes. Consistent with the results obtained from data aggregated to one-hour intervals, we observe that the MAE is further improved in SA-Net when the bias mitigation approach is implemented.

TABLE VI : Accuracy comparisons among different models (data aggregated to 30-minute intervals)

	MAE	RMSE	MAPE	Black (MAE)	Non-black (MAE)	Low-income (MAE)	High-income (MAE)
HA	4.48	11.35	0.70	0.99	6.99	3.65	6.82
MA	4.49	11.68	0.62	0.92	7.05	3.66	6.83
ARIMA	4.03	10.42	0.61	**0.88**	6.28	3.32	6.02
LSTM Net	4.39	8.63	0.47	1.18	6.68	3.48	6.94
LSTM + Social Net	4.39	8.60	0.47	1.16	6.70	3.47	6.96
STGCN	3.57	7.29	0.67	1.02	5.40	2.86	5.57
Wavenet	3.47	7.45	0.48	0.98	5.26	2.78	5.40
Conv-LSTM	3.44	6.84	0.41	1.06	5.14	**2.78**	5.29
Conv-LSTM + Social Net	3.46	**6.79**	0.43	1.07	5.17	2.80	5.30
SA-Net	**3.43**	6.85	**0.41**	1.06	**5.13**	2.79	**5.23**

Note: for the deep learning models, we report the results when the models are trained with $\gamma = 0$

TABLE VII : Fairness and accuracy comparisons with bias mitigation for race (data aggregated to 30-minute intervals)

	MAE	RMSE	MAPE	MPE gap (race)	Black (MPE)	Non-black (MPE)	MPE gap (income)	Low-income (MPE)	High-income (MPE)
Conv-LSTM Net:									
$\gamma = 0$	**3.442**	**6.842**	**0.413**	0.373	0.422	0.049	0.215	0.292	0.077
$\gamma = 5$	3.473	6.865	0.425	0.226	0.284	0.058	0.104	0.191	0.086
$\gamma = 10$	3.456	6.896	0.446	**0.064**	**0.069**	**0.006**	**0.037**	**0.047**	**0.010**
SA-Net:									
$\gamma = 0$	3.429	**6.852**	**0.408**	0.341	0.391	0.051	0.193	0.270	0.078
$\gamma = 5$	3.454	6.862	0.417	0.305	0.365	0.060	0.162	0.251	0.089
$\gamma = 10$	**3.415**	6.896	0.415	**0.031**	**0.063**	**0.032**	**0.004**	**0.045**	**0.041**

Note: γ represents the bias mitigation weight

TABLE VIII : Fairness and accuracy comparisons with bias mitigation for income (data aggregated to 30-minute intervals)

	MAE	RMSE	MAPE	MPE gap (race)	Black (MPE)	Non-black (MPE)	MPE gap (income)	Low-income (MPE)	High-income (MPE)
Conv-LSTM Net:									
$\gamma = 0$	**3.442**	6.842	**0.413**	0.373	0.422	**0.049**	0.215	0.292	**0.077**
$\gamma = 5$	3.462	**6.833**	0.434	0.256	0.307	0.051	0.104	0.194	0.090
$\gamma = 10$	3.499	6.924	0.434	**0.168**	**0.238**	0.070	**0.026**	**0.141**	0.115
SA-Net:									
$\gamma = 0$	3.429	6.852	0.408	0.341	0.391	0.051	0.193	0.270	0.078
$\gamma = 5$	3.427	6.813	0.427	0.383	0.448	0.065	0.188	0.297	0.109
$\gamma = 10$	**3.422**	**6.813**	**0.412**	**0.235**	**0.299**	**0.064**	**0.082**	**0.188**	**0.106**

Note: γ represents the bias mitigation weight

REFERENCES

[1] M. Diao, H. Kong, and J. Zhao, "Impacts of transportation network companies on urban mobility," *Nature Sustainability*, vol. 4, no. 6, pp. 494–500, 2021.

[2] Y. Zheng, P. Meredith-Karam, A. Stewart, H. Kong, and J. Zhao, "Impacts of congestion pricing on ride-hailing ridership: Evidence from chicago," *Transportation Research Part A: Policy and Practice*, vol. 170, p. 103639, 2023.

[3] J. Ke, H. Zheng, H. Yang, and X. M. Chen, "Short-term forecasting of passenger demand under on-demand ride services: A spatio-temporal deep learning approach," *Transportation Research Part C: Emerging Technologies*, vol. 85, pp. 591–608, 2017.

[4] X. Guo, Q. Wang, and J. Zhao, "Data-driven vehicle rebalancing with predictive prescriptions in the ride-hailing system," *IEEE Open Journal of Intelligent Transportation Systems*, vol. 3, pp. 251–266, 2022.

[5] X. Guo, N. S. Caros, and J. Zhao, "Robust matching-integrated vehicle rebalancing in ride-hailing system with uncertain demand," *Transportation Research Part B: Methodological*, vol. 150, pp. 161–189, 2021.

[6] N. Zhang, Y. Zhang, and H. Lu, "Seasonal autoregressive integrated moving average and support vector machine models: prediction of short-term traffic flow on free-

ways," *Transportation Research Record*, vol. 2215, no. 1, pp. 85–92, 2011.

[7] X. Li, G. Pan, Z. Wu, G. Qi, S. Li, D. Zhang, W. Zhang, and Z. Wang, "Prediction of urban human mobility using large-scale taxi traces and its applications," *Frontiers of Computer Science*, vol. 6, no. 1, pp. 111–121, 2012.

[8] H. Xu, J. Ying, H. Wu, and F. Lin, "Public bicycle traffic flow prediction based on a hybrid model," *Applied Mathematics & Information Sciences*, vol. 7, no. 2, pp. 667–674, 2013.

[9] Y. Li, Y. Zheng, H. Zhang, and L. Chen, "Traffic prediction in a bike-sharing system," in *Proceedings of the 23rd SIGSPATIAL International Conference on Advances in Geographic Information Systems*, 2015, pp. 1–10.

[10] G. Guo and T. Zhang, "A residual spatio-temporal architecture for travel demand forecasting," *Transportation Research Part C: Emerging Technologies*, vol. 115, p. 102639, 2020.

[11] C. Li, L. Bai, W. Liu, L. Yao, and S. T. Waller, "A multi-task memory network with knowledge adaptation for multimodal demand forecasting," *Transportation Research Part C: Emerging Technologies*, vol. 131, p. 103352, 2021.

[12] Y. Yang, A. Heppenstall, A. Turner, and A. Comber, "Using graph structural information about flows to enhance short-term demand prediction in bike-sharing systems," *Computers, Environment and Urban Systems*, vol. 83, p. 101521, 2020.

[13] F. Liu, J. Wang, J. Tian, D. Zhuang, L. Miranda-Moreno, and L. Sun, "A universal framework of spatiotemporal bias block for long-term traffic forecasting," *IEEE Transactions on Intelligent Transportation Systems*, 2022.

[14] M. Lippi, M. Bertini, and P. Frasconi, "Short-term traffic flow forecasting: An experimental comparison of time-series analysis and supervised learning," *IEEE Transactions on Intelligent Transportation Systems*, vol. 14, no. 2, pp. 871–882, 2013.

[15] E. I. Vlahogianni, M. G. Karlaftis, and J. C. Golias, "Short-term traffic forecasting: Where we are and where we're going," *Transportation Research Part C: Emerging Technologies*, vol. 43, pp. 3–19, 2014.

[16] H. Wang, H. Chen, Q. Wu, C. Ma, and Y. Li, "Multi-intersection traffic optimisation: A benchmark dataset and a strong baseline," *IEEE Open Journal of Intelligent Transportation Systems*, vol. 3, pp. 126–136, 2021.

[17] M. Zhong, J. Kim, and Z. Zheng, "Estimating link flows in road networks with synthetic trajectory data generation: Inverse reinforcement learning approach," *IEEE Open Journal of Intelligent Transportation Systems*, 2023.

[18] J. Hoppe, F. Schwinger, H. Haeger, J. Wernz, and M. Jarke, "Improving the prediction of passenger numbers in public transit networks by combining short-term forecasts with real-time occupancy data," *IEEE Open Journal of Intelligent Transportation Systems*, vol. 4, pp. 153–174, 2023.

[19] S. Wang, Q. Wang, and J. Zhao, "Deep neural networks for choice analysis: Extracting complete economic information for interpretation," *Transportation Research Part C: Emerging Technologies*, vol. 118, p. 102701, 2020.

[20] S. Wang, B. Mo, and J. Zhao, "Deep neural networks for choice analysis: Architecture design with alternative-specific utility functions," *Transportation Research Part C: Emerging Technologies*, vol. 112, pp. 234–251, 2020.

[21] Y. Li and B. Shuai, "Origin and destination forecasting on dockless shared bicycle in a hybrid deep-learning algorithms," *Multimedia Tools and Applications*, vol. 79, no. 7, pp. 5269–5280, 2020.

[22] R. Fu, Z. Zhang, and L. Li, "Using lstm and gru neural network methods for traffic flow prediction," in *2016 31st Youth Academic Annual Conference of Chinese Association of Automation (YAC)*. IEEE, 2016, pp. 324–328.

[23] C. Xu, J. Ji, and P. Liu, "The station-free sharing bike demand forecasting with a deep learning approach and large-scale datasets," *Transportation research part C: emerging technologies*, vol. 95, pp. 47–60, 2018.

[24] R. Yu, Y. Li, C. Shahabi, U. Demiryurek, and Y. Liu, "Deep learning: A generic approach for extreme condition traffic forecasting," in *Proceedings of the 2017 SIAM international Conference on Data Mining*. SIAM, 2017, pp. 777–785.

[25] L. Lin, Z. He, and S. Peeta, "Predicting station-level hourly demand in a large-scale bike-sharing network: A graph convolutional neural network approach," *Transportation Research Part C: Emerging Technologies*, vol. 97, pp. 258–276, 2018.

[26] Z. Cui, K. Henrickson, R. Ke, and Y. Wang, "Traffic graph convolutional recurrent neural network: A deep learning framework for network-scale traffic learning and forecasting," *IEEE Transactions on Intelligent Transportation Systems*, vol. 21, no. 11, pp. 4883–4894, 2019.

[27] Z. Zhang, M. Li, X. Lin, Y. Wang, and F. He, "Multistep speed prediction on traffic networks: A deep learning approach considering spatio-temporal dependencies," *Transportation research part C: emerging technologies*, vol. 105, pp. 297–322, 2019.

[28] Y. Wu, D. Zhuang, A. Labbe, and L. Sun, "Inductive graph neural networks for spatiotemporal kriging," *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 35, no. 5, pp. 4478–4485, May 2021. [Online]. Available: https://ojs.aaai.org/index.php/AAAI/article/view/16575

[29] G. Jin, Y. Cui, L. Zeng, H. Tang, Y. Feng, and J. Huang, "Urban ride-hailing demand prediction with multiple spatio-temporal information fusion network," *Transportation Research Part C: Emerging Technologies*, vol. 117, p. 102665, 2020.

[30] X. Wan, H. Ghazzai, and Y. Massoud, "A generic data-driven recommendation system for large-scale regular and ride-hailing taxi services," *Electronics*, vol. 9, no. 4, p. 648, 2020.

[31] F. Rodrigues, I. Markou, and F. C. Pereira, "Combining time-series and textual data for taxi demand prediction in event areas: A deep learning approach," *Information Fusion*, vol. 49, pp. 120–129, 2019.

[32] Z. Huang, W. Zhang, D. Wang, and Y. Yin, "A gan framework-based dynamic multi-graph convolutional network for origin–destination-based ride-hailing demand prediction," *Information Sciences*, vol. 601, pp. 129–146, 2022.

[33] M. H. Rahman and S. M. Rifaat, "Using spatio-temporal deep learning for forecasting demand and supply-demand gap in ride-hailing system with anonymised spatial adjacency information," *IET Intelligent Transport Systems*, vol. 15, no. 7, pp. 941–957, 2021.

[34] L. Chen, P. Thakuriah, and K. Ampountolas, "Short-term prediction of demand for ride-hailing services: A deep learning approach," *Journal of Big Data Analytics in Transportation*, vol. 3, pp. 175–195, 2021.

[35] J. Angwin, J. Larson, S. Mattu, and L. Kirchner, "Machine bias," *ProPublica, May*, vol. 23, p. 2016, 2016.

[36] A. Chouldechova, "Fair prediction with disparate impact: A study of bias in recidivism prediction instruments," *Big data*, vol. 5, no. 2, pp. 153–163, 2017.

[37] A. Rajkomar, M. Hardt, M. D. Howell, G. Corrado, and M. H. Chin, "Ensuring fairness in machine learning to advance health equity," *Annals of Internal Medicine*, vol. 169, p. 866, 2018.

[38] S. N. Goodman, S. Goel, and M. R. Cullen, "Machine learning, health disparities, and causal reasoning," *Annals of internal medicine*, vol. 169, no. 12, pp. 883–884, 2018.

[39] S. Y. Deku, A. Kara, and P. Molyneux, "Access to consumer credit in the uk," *The European Journal of Finance*, vol. 22, no. 10, pp. 941–964, 2016.

[40] M. S. A. Lee, "Context-conscious fairness in using machine learning to make decisions," *AI Matters*, vol. 5, no. 2, pp. 23–29, 2019.

[41] S. Corbett-Davies and S. Goel, "The measure and mismeasure of fairness: A critical review of fair machine learning," *CoRR*, vol. abs/1808.00023, 2018. [Online]. Available: http://arxiv.org/abs/1808.00023

[42] P. Gajane, "On formalizing fairness in prediction with machine learning," *CoRR*, vol. abs/1710.03184, 2017. [Online]. Available: http://arxiv.org/abs/1710.03184

[43] S. A. Friedler, C. Scheidegger, S. Venkatasubramanian, S. Choudhary, E. P. Hamilton, and D. Roth, "A comparative study of fairness-enhancing interventions in machine learning," in *Proceedings of the Conference on Fairness, Accountability, and Transparency - FAT* '19*. ACM Press, 2019, pp. 329–338. [Online]. Available: http://dl.acm.org/citation.cfm?doid=3287560.3287589

[44] T. Kamishima, S. Akaho, and J. Sakuma, "Fairness-aware learning through regularization approach," in *2011 IEEE 11th International Conference on Data Mining Workshops*, 2011, pp. 643–650.

[45] B. H. Zhang, B. Lemoine, and M. Mitchell, "Mitigating unwanted biases with adversarial learning," in *Proceedings of the 2018 AAAI/ACM Conference on AI, Ethics, and Society*. ACM, 2018, pp. 335–340. [Online]. Available: https://dl.acm.org/doi/10.1145/3278721.3278779

[46] Kenneth Holstein, Jennifer Wortman Vaughan, Hal Daumé, Miro Dudik, and Hanna Wallach, "Improving fairness in machine learning systems: What do industry practitioners need?" in *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems - CHI '19*. ACM Press, 2019, pp. 1–16. [Online]. Available: http://dl.acm.org/citation.cfm?doid=3290605.3300830

[47] D. Gunduz, P. de Kerret, N. D. Sidiropoulos, D. Gesbert, C. R. Murthy, and M. van der Schaar, "Machine learning in the air," *IEEE Journal on Selected Areas in Communications*, vol. 37, no. 10, pp. 2184–2199, 2019. [Online]. Available: https://ieeexplore.ieee.org/document/8839651/

[48] Y. Zheng, S. Wang, and J. Zhao, "Equality of opportunity in travel behavior prediction with deep neural networks and discrete choice models," *Transportation Research Part C: Emerging Technologies*, vol. 132, p. 103410, 2021.

[49] A. Yan and B. Howe, "Fairness-aware demand prediction for new mobility," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 34, no. 01, 2020, pp. 1079–1087.

[50] ——, "Equitensors: Learning fair integrations of heterogeneous urban data," in *Proceedings of the 2021 International Conference on Management of Data*, 2021, pp. 2338–2347.

[51] S. Barocas and A. D. Selbst, "Big data's disparate impact," *California law review*, pp. 671–732, 2016.

[52] N. Mehrabi, F. Morstatter, N. Saxena, K. Lerman, and A. Galstyan, "A survey on bias and fairness in machine learning," *ACM Computing Surveys (CSUR)*, vol. 54, no. 6, pp. 1–35, 2021.

[53] M. Hardt, E. Price, and N. Srebro, "Equality of opportunity in supervised learning," *Advances in neural information processing systems*, vol. 29, 2016.

[54] N. C. for Environmental Information, "Climate data online," *National Oceanic and Atmospheric Administration*, 2021.

[55] United States Census Bureau, "Longitudinal Employer-Household Dynamics," Online, https://lehd.ces.census.gov/data/, 2023.

[56] Chicago Transit Authority, "CTA GTFS," Online, https://transitfeeds.com/p/chicago-transit-authority/165, 2023.

[57] H. Su, V. Jampani, D. Sun, O. Gallo, E. Learned-Miller, and J. Kautz, "Pixel-adaptive convolutional neural networks," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2019, pp. 11 166–11 175.

[58] H. Yao, F. Wu, J. Ke, X. Tang, Y. Jia, S. Lu, P. Gong, J. Ye, and Z. Li, "Deep multi-view spatial-temporal network for taxi demand prediction," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 32, no. 1, 2018.

[59] C. D. Portal, "Transportation network providers–trips," 2020.

[60] U. C. Bureau, "2015–2019 american community survey 5-year estimates," 2019.

[61] G. E. Box and D. A. Pierce, "Distribution of residual autocorrelations in autoregressive-integrated moving

average time series models," *Journal of the American statistical Association*, vol. 65, no. 332, pp. 1509–1526, 1970.

[62] M. Van Der Voort, M. Dougherty, and S. Watson, "Combining kohonen maps with arima time series models to forecast traffic flow," *Transportation Research Part C: Emerging Technologies*, vol. 4, no. 5, pp. 307–318, 1996.

[63] B. M. Williams and L. A. Hoel, "Modeling and forecasting vehicular traffic flow as a seasonal arima process: Theoretical basis and empirical results," *Journal of transportation engineering*, vol. 129, no. 6, pp. 664–672, 2003.

[64] B. Yu, H. Yin, and Z. Zhu, "Spatio-temporal graph convolutional networks: A deep learning framework for traffic forecasting," *arXiv preprint arXiv:1709.04875*, 2017.

[65] Z. Wu, S. Pan, G. Long, J. Jiang, and C. Zhang, "Graph wavenet for deep spatial-temporal graph modeling," *arXiv preprint arXiv:1906.00121*, 2019.

[66] G. Giuliano, "Travel, location and race/ethnicity," *Transportation Research Part A: Policy and Practice*, vol. 37, no. 4, pp. 351–372, 2003.

[67] Q.-L. Gao, Y. Yue, W. Tu, J. Cao, and Q.-Q. Li, "Segregation or integration? exploring activity disparities between migrants and settled urban residents using human mobility data," *Transactions in GIS*, vol. 25, no. 6, pp. 2791–2820, 2021.

[68] Y. Zheng, H. Kong, G. Petzhold, M. M. Barcelos, C. P. Zegras, and J. Zhao, "Gender differences in the user satisfaction and service quality improvement priority of public transit bus system in porto alegre and fortaleza, brazil," *Travel Behaviour and Society*, vol. 28, pp. 22–37, 2022.

[69] J. Zhang, Y. Zheng, and D. Qi, "Deep spatio-temporal residual networks for citywide crowd flows prediction," in *Thirty-first AAAI conference on artificial intelligence*, 2017.

Qingyi Wang is a PhD student in transportation at MIT. She graduated with a Bachelor's of Applied Science in Engineering Science from the University of Toronto in 2018, and with a Master of Science in Transportation from MIT in 2020. Her research focuses on bringing machine learning into transportation demand modelling, and making use of domain knowledge to improve and better interpret machine learning models.

Dingyi Zhuang is a master student at MIT Urban Mobility Lab. He received a B.S. degree in Mechanical Engineering from Shanghai Jiao Tong University in 2019, and an M.Eng. degree in Transportation Engineering from McGill University, in 2021. His research interests lie in deep learning, urban computing, and spatiotemporal data modeling.

Shenhao Wang is an assistant professor at the University of Florida and research affiliate at MIT Urban Mobility Lab and Human Dynamics Group in Media Lab. His research focuses on developing interpretable, generalizable, and ethical deep learning models to analyze individual decision-making with applications to urban mobility. He synergizes discrete choice models and deep neural networks in travel demand modeling by interpreting the "black box" deep neural networks with economic theory. Shenhao Wang completed his interdisciplinary Ph.D. in Computer and Urban Science from MIT in 2020.

Yunhan Zheng is a PhD student in transportation. She received the bachelor's degrees in urban management and economics from Peking University and the master's degrees in city planning and transportation from MIT. Her research interests lie in designing fair transport algorithms and policies for sustainable mobility.

Jinhua Zhao is currently the Edward H. and Joyce Linde Associate Professor of city and transportation planning at MIT. He brings behavioral science and transportation technology together to shape travel behavior, design mobility systems, and reform urban policies. He directs the MIT Urban Mobility Laboratory and Public Transit Laboratory.