The Journal of Transport and Land Use jtlu.org
Vol. 17 No. 1 [2024] pp. 115–142

Spatial-temporal deep learning model based on Similarity Principle for dock shared bicycles ridership prediction

Jiahui Zhao
Southeast University
jiahuizhao@seu.edu.cn

Zhibin Li
Southeast University
lizhibin@seu.edu.cn

Pan Liu (corresponding author)
Southeast University
liupan@seu.edu.cn

Mingye Zhang
Southeast University
zhangmingye@seu.edu.cn

Abstract: Demand prediction plays a critical role in traffic research. The key challenge of traffic demand prediction lies in modeling the complex spatial dependencies and temporal dynamics. However, there is no mature and widely accepted concept to support the solution of the above challenge. Essentially, a prediction model combined with similar objects in temporal and spatial dimensions could obtain better performance. This paper proposes a concept called the Similarity-based Principle (SP), which is applied to improve the prediction performance of deep learning models in complex traffic scenarios. For the temporal components, the long-term temporal dynamics in contemporaneous historical data for ridership are extracted by the Stacked Autoencoder (SAE) method. For the spatial components, the activity-based spatial geographic information (ABG-information) is used to capture the spatial correlation of the traffic network, which is reflected in the daily activities of humans. Specifically, the SP is applied to a Spatio-temporal Graph Convolutional Neural Network (STGCNN) model. In the case study, the Similarity-based Principle Spatio-temporal Graph Convolutional Neural Network (SP-STGCNN) model predicts demand for bicycle sharing in San Francisco. The results show that the SP effectively improves the model's performance. The prediction accuracy is enhanced by up to 10.34% compared with STGCNN. For spatial relationships, the model using the geographic information attribute performs better than that using the road information attribute and the distance attribute. It is proved that the construction of the Spatio-temporal model-based similarity principle can improve the performance.

Keywords: Traffic demand prediction, Similarity-based Principle, Spatio-temporal Graph Convolutional Neural Network model, activity-based geographic information, prediction of bicycle sharing ridership.

Article history:
Received: February 19, 2023
Received in revised form: September 26, 2023
Accepted: December 24, 2023
Available online: February 27, 2024

https://doi.org/10.5198/jtlu.2024.2348
ISSN: 1938-7849

The *Journal of Transport and Land Use* is the official journal of the World Society for Transport and Land Use (WSTLUR) and is published and sponsored by the University of Minnesota Center for Transportation Studies.

1 Introduction

Traffic prediction is an important research area focused on anticipating traffic demand to mitigate congestion and balance demand and supply (Li et al., 2015; Maleki Vishkaei et al., 2020). Essentially, traffic demand prediction aims to predict a traffic-related value for a location at a timestamp based on historical data. Traffic prediction for a long period gives a detailed predicting of traffic models to evaluate future capacity requirements and therefore permits for more minute planning and better decisions. Short-term traffic prediction can provide decision support for congestion control and for optimal resource management (organization dispatch and path planning) (Chen et al., 2020).

In the past, prediction problems have not been extensively studied, even though the importance of traffic prediction in real traffic management. The primary challenges in traffic prediction lie in temporal variations and spatial variations. Concerning temporal variations, traffic patterns display daily, weekly, and seasonal fluctuations. For instance, morning and evening rush hours may vary, and there might be distinctions between weekday and weekend traffic flows. Without advanced models, accurately predicting these patterns was challenging. Regarding Spatial Variations, traffic dynamics differ across various locations. CBD (Central Business District) may exhibit distinct traffic flows compared to their suburban counterparts. Earlier prediction models struggled to integrate spatial correlations effectively. Two primary reasons underlie these challenges for predict models: traditional traffic prediction models mainly focus on the prediction task of single time series and lack the ability to combined multi-source data (Joshi & Hadi, 2015). In the past decade, with the development of deep learning (DL) prediction models (Long Short Term Memory (LSTM) and Gated Recurrent Unit (GRU)), the prediction accuracy of traffic prediction tasks has been significantly improved (Ramakrishnan & Soni, 2018). Predictions for the single time series have been well addressed. In recent years, how to solve many-to-many traffic prediction have gradually become a hotspot research task (Xu et al., 2018). Due to the spatial-temporal nature of the traffic domain, the critical challenge of traffic demand prediction lies in how to model the complex spatial and temporal dependencies.

In this study, we aim to address certain challenges by proposing a Similarity-based Principle (SP). This principle is designed to construct deep learning models and enhance the performance of the Spatio-temporal Graph Convolutional Neural Network model specifically for traffic demand prediction. For the temporal components, the long-term temporal dynamics in contemporaneous historical data for ridership are extracted by the Stacked Autoencoder (SAE) method. For the spatial components, the activity-based spatial geographic information (ABG-information) is used to capture the spatial correlation of the traffic network. Specifically, the Similarity-based Principle is applied to a Spatio-temporal Graph Convolutional Neural Network (GCNN) model. Our contributions are three-fold:

(1) In this study, an innovative principle, the Similarity-based Principle, is applied to the Spatio-temporal traffic demand prediction model.

(2) This paper uses the Stacked Autoencoder (SAE) method to compress the long-term historical contemporaneous information. The feature extraction method introduces the activity-based data to improve the spatial correlation of the GCN model to improve the prediction accuracy and training efficiency.

(3) This paper verifies a thought that the integration of similar information is helpful to the optimization of the Spatio-temporal demand prediction model.

2 Literature review

Considering the influx of scholars with computer backgrounds into traffic field research in the past two years, their research has been provided a large number of prediction models. 40 highly cited papers published in the last four years are summarized in Table 1.

Table 1. Highly cited papers on traffic prediction

	Prediction objects					Multi-data					Model	
	P1	P2	P3	P4	P5	M1	M2	M3	M4	M5	Temporal model	Spatial model
(J. Feng et al., 2018)					*		*				LSTM	General feature extractor
(S. Feng et al., 2018)		*				*					Spectral Clustering (SC) Algorithm	
(Hulot et al., 2018)		*				*					Distribution hypothesis	
(Chai et al., 2018)		*							*		Convolutional layers	Multiple graphs
(Li et al., 2018)					*					O1	Encoder-decoder architecture	Bidirectional random walks on the graph
(Lin et al., 2018)		*							*		Graph Convolutional Neural Network with data-driven graph filter (GCNN-DDGF) model	
(Yu et al., 2018)					*				*		Gated CNNs	Graph CNNs
(Fang et al., 2019)	*		*	*					*		Multi-resolution temporal module	Global correlated spatial module
(Yao, Wu, et al., 2019)	*					*		*			LSTM	Local CNN
(Liu et al., 2019)	*					*					Temporal evolution context	Local spatial context (LSC) and global correlation context (GCC)
(Pan et al., 2019)	*						*		*		Recurrent neural network	Meta graph attention network
(Saxena & Cao, 2019)	*	*				*	*				ConvLSTM	ST maps
(Yao, Tang, et al., 2019)	*	*								O2	Periodically Shifted Attention Mechanism	Flow Gating Mechanism
(Jiang et al., 2019)		*							*		LSTM	Convolutional neural network
(Li et al., 2019)		*							*		Dynamic attention-based graph embedding model	Dynamic heterogeneous graph
(Wu et al., 2019)					*				*		Gated temporal convolution layer	Graph convolution layer (GCN)
(Do et al., 2019)					*				*		Temporal attention	Spatial attention
(Geng et al., 2019)	*						*				Contextual gated recurrent neural network (CGRNN)	Multi-graph convolution
(Li & Zheng, 2020)		*									Gaussian Process Regressor	Adaptive Transition Constraint (AdaTC) clustering algorithm
(Zhang et al., 2020)	*										Spatiotemporal dynamic time warping (ST-DTW) algorithm	Multi-task learning temporal convolutional neural network (MTL-TCNN)
(Bogaerts et al., 2020)					*				*		LSTM	Graph convolutional neural network

	Prediction objects					Multi-data					Model	
	P1	P2	P3	P4	P5	M1	M2	M3	M4	M5	Temporal model	Spatial model
(Chen et al., 2020)					*						AE-GRU	
(Cui, Henrickson, et al., 2020)					*				*		LSTM	Traffic graph convolution (TGC)
(Cui, Ke, et al., 2020)					*						Gated recurrent structure	Graph wavelet
(Shi et al., 2020)					*				*		Attention-based Periodic-Temporal neural Network (APTN)	Graph Convolution Networks GCN)
(Wang et al., 2020)					*				*		Sequence to sequence (seq2seq) architecture	Graph convolution
(L. Zhao et al., 2020)	*				*				*		GRU	GCN
(Zheng, 2020)					*				*		Temporal attention mechanism models	Spatial attention mechanism
(Bao et al., 2021)	*										LSTM	3D convolution
(L. Chen et al., 2021)	*										CNN	
(Wang et al., 2021)		*					*				Deep Convolutional Neural Network	
(Yao & Qian, 2021)					*	*				O3	Clustered learning structure	Social media data augmentation method
(Zhang et al., 2021)					*			*			Dynamic OD graphs	
(Guo et al., 2021)		*							*		Self-attention mechanism	Dynamic graph convolution module
(Z. Chen et al., 2021)		*							*		Attention-based ST-GCN (AST-GCN)	
(Yin et al., 2021)					*				*		Internal attention mechanism	Dynamic neighborhood-based attention mechanism
(Zi et al., 2021)		*							*		Temporal attention	GCN
(Yuan et al., 2021)	*								*		Temporal graphs	Three spatial graphs
(Du et al., 2021)	*	*					*				Dynamic transition convolution	Clustering algorithm
(Xiao et al., 2021)		*							*		Gated CNNs	Convolutional neural networks

Notes:
P1-Taxi P2-Bike P3-Subway P4-Bus P5-Traffic flow
M1-Weather M2-POI M3-OD M4-Network M5-Other data
O1-Distance O2-Similarity of demand patterns O3-Twitter

Despite the diversity of these models, their common characteristics are 1. The state-of-the-art DL-based methods combine the spatial modules (such as CNN, GCN) with the temporal modules (such as LSTM, GRU) to address the spatial and temporal dependence of traffic prediction problems. 2. Multi-source data are used. 3. Most papers predict traffic flow and used public data sets. However, these methods have three significant limitations in application: 1. There is no clear theoretical guidance on how to select among these modules to build a well-performed model given the diversity of existing deep learning modules. 2. There is no clarity on how to organize these multi-source data to achieve the best performance for models. 3. There is no generalization model applied to traffic flow prediction and demand prediction, which means that traffic flow prediction, cannot be directly transplanted to the demand prediction problems.

First, there is no clear theoretical guidance on how to select among these modules to build a well-performed model given the diversity of existing DL modules. Traffic researchers need to understand how to select and deploy these modules inside a DL network to maximize the effect. Although there are many well-performing models, no study has proposed standardized guidelines for the construction of DL models in different prediction scenarios (Sarker, 2021) (Taye, 2023).

Second, existing studies lack interpretability in selecting and organizing multi-source information. An increasing number of studies are using multi-source data (traffic data, land use, weather, etc.) as input to deep learning models. In Table 1, 35 studies (87.5%) integrate weather, road network, land use, and other external factors into the prediction model. One potential assumption underlying these studies is that the used DL models can understand the dependency and capture beneficial information in the multi-source input dataset (Alzubaidi et al., 2021) (Taye, 2023). To the best of the author's knowledge, no studies have proved this assumption by interpreting the information been captured in the proposed model. It's still not clear how to organize these multi-source data to achieve the best performance.

Third, in Table 1, 13 studies (59.1%) predict traffic flow combined road network between 2020 and 2021. In the task of traffic flow prediction, the adjacent and close monitoring points on the road network have a strong correlation. Taking the intersection as an example, vehicles passing through the detector of the East through lane will probably pass through the detector of the adjacent West through lane. The correlation of detectors is more evident in some highway data sets because vehicles are impossible to leave or turn around easily on highways (Li et al., 2017). This graph network structure constructs the link relationship according to the spatial correlation between detectors or measures it with distance. Such the network relationship can effectively grasp the core of traffic prediction, that is, the strong correlation of adjacent objects. Demand predicting refers to predicting the traffic inflow and outflow of an object (traffic district, station, etc.) (Sun et al., 2022). However, the graph network structure of traffic flow prediction can not be directly transplanted to the demand prediction problems (Z. Chen et al., 2020).

To solve the abovementioned challenges, we propose a novel method named Similarity-based Principle (SP) to construct the deep learning model and organize multi-source data in traffic demand prediction. The SP refers to combining related/similar objects temporally and spatially. In the traffic domain, the similarity feature means that the change of traffic data characteristics has the spatial dependency on locations and the temporal dependency on historical periodicity. The thought of SP is valuable in traffic prediction issues. As the essence of the traffic prediction task is multiple regressions, the more relevant the input information is to the prediction target, the more accurate the prediction result is. SP is consistent with researchers' general consideration of prediction issues. Researchers have pre-processed data for traffic prediction, such as eliminating data outliers (Ma et al., 2021), mining the correlation between variables (Liu et al., 2022), and pre-classifying the sample data. The essence of these operations is to reinforce the similarity between the inputs to the model internally.

To prove the effectiveness and interpretability of this concept, we use SP for training and testing the state-of-the-art deep learning prediction model. Previous research has improved the structure and performance of the model in both temporal and spatial modules. For the temporal components, as demonstrated in Figure 1(a), some researchers assumed the model could capture long-term traffic characteristics (in days or weeks) and utilize them to estimate short-term traffic demand (in hours or minutes). Previous research employed CNN models for the spatial components to extract spatial features in traffic prediction (Pan et al., 2018). They commonly used distance, road network, interaction, and station clustering to show spatial relationships, as shown in Figure 2(a).

However, longer input sequences may contain long-term data features and more irrelevant information, causing the model's performance to decrease rapidly. In addition, the road network topology, station distances, and interworking relationships cannot accurately reflect users' actions in the actual world because daily human activities generate traffic data.

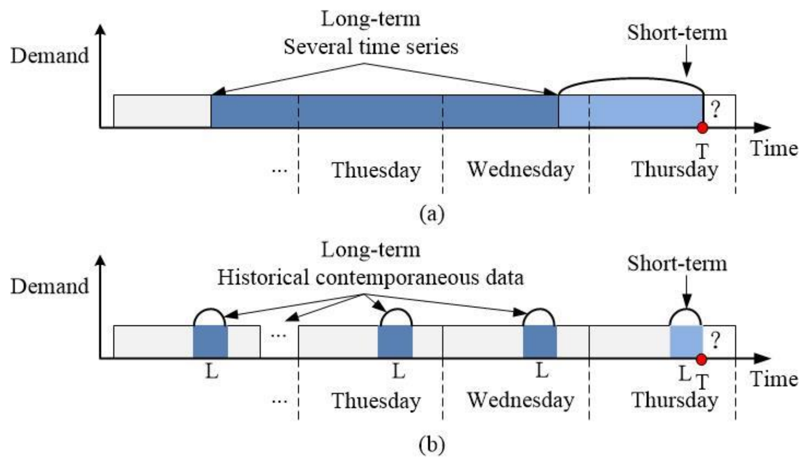

Figure 1. Periodic repetition of traffic demand

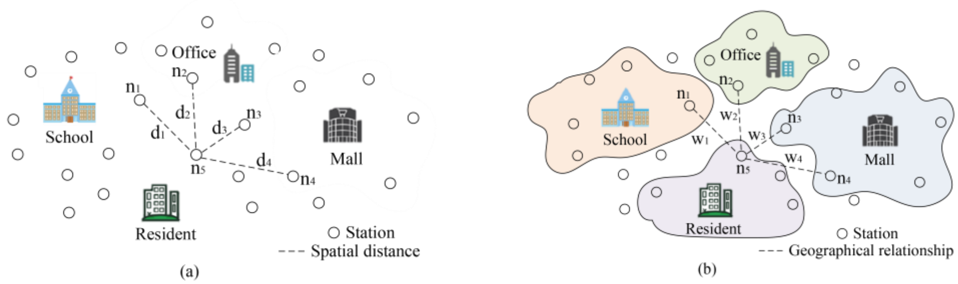

Figure 2. Spatial heterogeneousness among stations

To fill these gaps, SP is applied to the spatio-temporal demand prediction framework. As indicated in Figure 1(b), the similarity-based principle integrates similarities between the features of the period to be predicted and those of historical periods (Yang, 2013). It can successfully reduce the input sequence's dimension and extract long-term historical data features. For spatial relationship, ridership fluctuation has similar characteristics, which is related to geographic information. For example, during the morning rush hour on weekdays, most traffic starts from the residential areas and ends at workplaces. On weekdays, there is an opposite traffic phenomenon during the evening rush hour, as shown in Figure 2(b). In addition, existing research shows that the distribution of shared bicycles is highly coincident with the distribution of hotspot facilities such as urban residences, companies, restaurants, and rail transit stations (J. Zhao et al., 2020). Therefore, the introduction of geographic data to measure spatial relevance may have a better performance in demand prediction. Based above analysis, the Similarity-based Principle is applied to the Spatio-temporal Graph Convolutional Neural Network model in the shared bicycles research scenario.

The rest of this paper is organized as follows. Section 3 describes the proposed prediction model for the traffic system. Section 4 discusses the case study for shared bicycles in San Francisco and results in various scenarios. Finally, the conclusions are provided in Section 5.

3 Problem description and model formulation

3.1 Problem description

The objective of demand prediction is based Similarity Principle to predict the future demand using activity-based spatial geographic information. Our proposed approach consists of two stages: spatio-temporal similarity processing and the ST-GCNN model prediction. At the first stage, the Stacked Autoencoder (SAE) method is applied to extract the similarity of contemporaneous historical data and get long-term traffic data information. The activity-based data are used to measure the spatial similarity among different prediction objects. The spatial geographic information is extracted from the land-use planning data and the points of interest data (POI). At the second stage, the Spatio-temporal Graph Convolutional Neural Network is proposed to reconstruct the graph structure for the demand prediction.

3.2 Bicycle sharing ridership predicting

Both contemporaneous historical data and the geographic information on where the trips are generated and attracted are needed to predict the traffic demand. Making full use of the spatial and temporal dependences is the key to solving traffic demand prediction problems. The architecture of SP-STGCNN is shown in Figure 3.

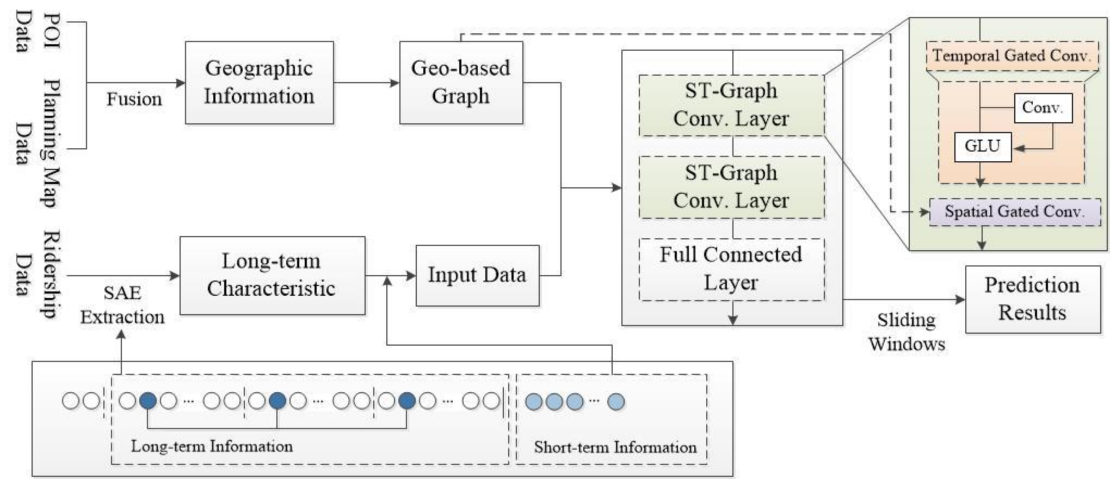

Figure 3. Architecture of SP-STGCNN

The graph structure shows the topological characteristics of the network and its spatial dependence. The network can be defined by a graph as $G = (N, E, W)$, where N is the set of nodes, and E is the set of edges. Each node has its characteristics, and the relationship between each pair of nodes forms a matrix of $N \times N$, known as the weighted adjacency matrix of W. At the t^{th} time step, x_t^n represents the ridership (generation or attraction) of stations n at time step t. E_{n_i,n_j} represents the spatial relationships between any two stations, where $\forall n_i, n_j \in N$. If n_i and n_j are connected ($(n_i, n_j) \in E$), $W_{i,j}$ represents the spatial relationship between n_i and n_j.

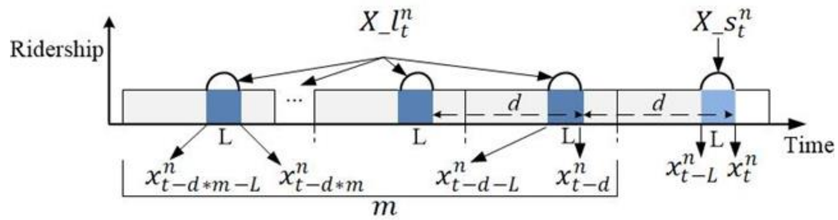

Figure 4. Illustration of ridership data

The ridership data X_t^n contains long-term information and short-term information, which presents as $X_t^n = \left[X_{l_t}^{\ n}, X_{s_t}^{\ n}\right]$. The illustration of ridership data is shown in **Figure 4**. The long-term information $X_{l_t}^{\ n}$ contains several contemporaneous historical data, which can be defined as $X_{l_t}^{\ n} =$ $[x_{t-d*m-l}^n, \dots, x_{t-d*m}^n, \dots, x_{t-d*2-l}^n, \dots, x_{t-d*2}^n, x_{t-d-l}^n, \dots, x_{t-d}^n]$. The SAE model is chosen to reduce the dimension of $Z_{l_t}^{\ n}$ to concentrate on the long-term information $X_{l_t}^{\ n}$. The input data is presented as $X_{in_t}^{\ \ n} = \left[Z_{l_t}^{\ n}, X_{s_t}^{\ n}\right]$. At the same time, POI data and land use planning data are combined to extract the activity-based geographic information, The weighted adjacency matrix of W measures the spatial relationship between nodes based on the geographic information. Finally, the temporal and spatial similarity information is inputted into the deep learning prediction model, and the demand prediction results $\hat{x}_{t+1}^n$ are obtained for the docking station n at time step $t + 1$.

3.3 Similarity principle in temporal module

Predicting future traffic demand needs historical demand series data. Yao et al. (2018) focused on the impact of long-term and short-term time series on prediction, and he pointed out that periodically shifted attention mechanism captures the long-term dependency and temporal shifting, and the LSTM captures the short-term temporal dependence. For the prediction model, the length of the data collection time frame can affect the accuracy of the results significantly as the long-term effects on the demand could not be captured if the data are collected during a short period. Meanwhile, the short-term effects could be missed if data collected over a long period are applied to the model (Wu et al., 2019; Wang et al., 2020). Therefore, this study uses Stacked Autoencoder mode to integrate the historical contemporaneous information based on the Similarity Principle.

The Stacked Autoencoder (SAE) is connected to several Autoencoders (AE) (Hinton, 2006). AE is used for dimension reduction and abnormal detection. Specifically, AE encodes the input data X to obtain a new low-dimensional feature Z, and then decodes Z into Y, and makes Y as close to the input X as possible.

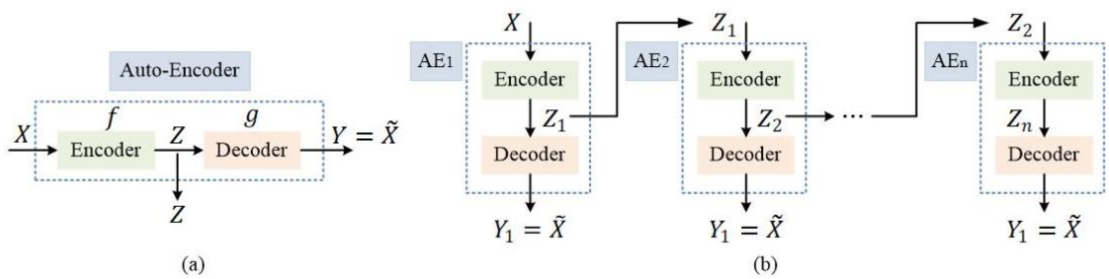

Figure 5. Structure of Autoencoder and Stacked Autoencoder

The AE network can be divided into two parts: the encoder represented by the function $Z = f(X)$ and the decoder $Y = g(Z)$ to generate the reconstruction using the following formula.

$$z_i = f(w_t \cdot x_i + b_t) \tag{1}$$

$$y_i = g\left(w_y \cdot t_i + b_y\right) \tag{2}$$

$$Y = g(Z) = g\left(f(x)\right) \approx X \tag{3}$$

Using the model parameters, reconstruction errors $L(X, Z)$ can be minimized and denoted as θ.

$$\theta = \arg\min_{\theta} L(X, Z) = \arg\min_{\theta} \frac{1}{2} \sum_{i=1}^{N} \|x_i - z_i\|^2 \tag{4}$$

After encoding, the intermediate state Z has a low-dimensional feature that contains long-term historical information. When repeating this process, the intermediate state Z is used as the new input X for the new AE. A new Auto-Encoder is trained and stacked layer by layer to form a Stacked Auto-Encoder.

In this paper, the sliding window algorithm is used to make predictions for multiple time steps. The algorithm divides the time window into smaller periods. Every period, the time window slides a grid to the right. The algorithm is shown in Figure 6. The counters in all periods are accumulated when calculating the total number of requests over the entire time window. The finer the time window is, the smoother the scrolling of the sliding window is.

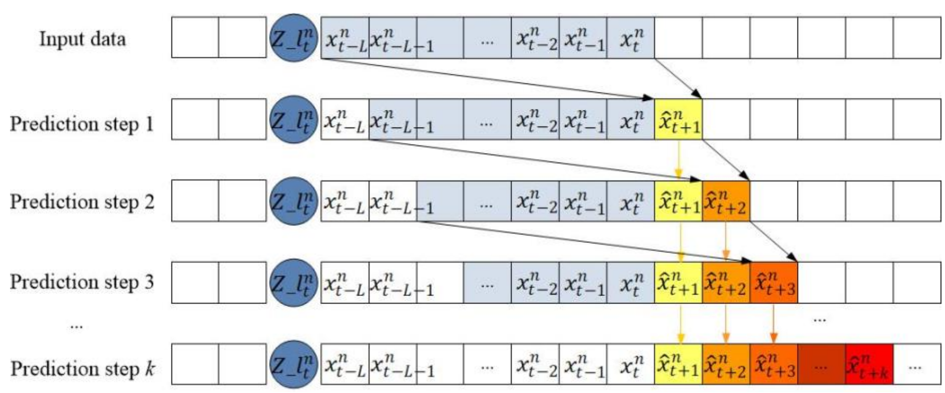

Figure 6. Sliding windows algorithm

3.4 Similarity principle in spatial module

Given that the traffic state is spatio-temporal related in nature, the traffic prediction needs to consider the impact of the spatial correlations, which can be achieved by characterizing similar traffic measures from one road to another (Triguero et al., 2017). Many studies concluded that spatial correlations are determined by distance attributes, for example, Bogaerts et al. (2020) found that the adjacent sections exhibit the same phenomena as distance attributes. However, compared with the distance attribute, the spatial correlation of ridership is more suitable to be measured by the built environment or land use which can reflect human beings' daily activities. In this study, graph structures use the activity-based geographic information as the spatial similarity matrix in the ridership prediction.

The activity-based geographic information (ABG-information) is estimated to describe the land-use characteristics and intensity using existing land-use planning and GIS data. The first ABG-information dataset (D-1) is based on the method introduced by Zhao et al. (2020). Four types of land-use categories are included: work, consumption, transit, residence. The work level is evaluated based on factors provided by land-use planning reports and GIS datasets, such as the level of commerce, the density of industries, and the density of companies. The consumption level is estimated using the level of commerce, the density of shopping centers, and the density of restaurants. The transit level is evaluated using the level of transport, the number of transit lines, and the density of transit stations. The level of residence refers to the density of residence. The weighted adjacency matrix of ABG-information is presented as ${W_1}_{{abg}_{n_i,n_j}}$,

$$ {W_1}_{{abg}_{n_i,n_j}} = \sum_{k=1}^{K} \left| g_{n_i}^{k} - g_{n_j}^{k} \right| \tag{5} $$

where k is one of the ABG-information categories, K=4, and g is the value of activity-based data for work, consumption, transit, or residence.

The second ABG-information dataset (D-2) is from (Zhao et al., 2022). There are seven categories: CIE (cultural, institutional, and educational), MED (medical), MIPS (office (management, information, and professional services)), PDR (industrial (production, distribution, and repair)), RES (residential), RETAIL (retail, and entertainment), VISITOR (retail, entertainment, hotels and visitor services). The weighted adjacency matrix of ABG-information is presented as as ${\mathrm{W}_2}_{{\mathrm{abg}}_{\mathrm{n_i,n_j}}}$,

$$ {W_2}_{{abg}_{n_i,n_j}} = \sum_{k=1}^{K} \left| g_{n_i}^{k} - g_{n_j}^{k} \right| \tag{6} $$

where k is one of the ABG-information categories, K=7, and g is the value of activity-based data for seven categories.

The road network dataset is used in the comparison experiments and obtained from the DataSF government website. This dataset includes bikeway network, street intersection, traffic signal, traffic stop and the number, type and length of road segment. This weigh adjacency matrix of road information ${R_{info}}_{n_i,n_j}$ is defined as,

$$ {R_{info}}_{n_i,n_j} = \sum_{k=1}^{K} \left| r_{n_i}^{k} - r_{n_j}^{k} \right| \tag{7} $$

where k is one of the road network information categories, K=8, and r is the value of road information for eight categories.

The weighted adjacency matrix of distance ${W_{dis}}_{n_i,n_j}$ is defined and used in the comparison experiments, and is presented as,

$$W_{dis_{n_i,n_j}} = \sqrt{\left(dx_{n_i} - dx_{n_j}\right)^2 + \left(dy_{n_i} - dy_{n_j}\right)^2} \tag{8}$$

where dx is the latitude of the node n, and dy is the longitude of the node n.

3.5 Prediction methodology of STGCNN

The Similarity-based Principle is applied to the Spatio-temporal Graph Convolutional Neural Network model which combines both temporal and spatial dimensions. In the temporal blocks, the Gated Liner United model is chosen to capture temporal dynamic behaviors in traffic demand. In the spatial blocks, Graph Convolution Network is applied to extract features in the space domain.

The temporal module of GCNN can extract temporal correlations from the historical data. According to Dauphin et al. (2017), the Gated Liner United (GLU) model requires fewer computational resources than the LSTM model with simple operation. The GLU model cancels the forget gate and uses the activation function to control the transmission of input information, thereby forming long-term memory. The temporal gated convolution. The temporal gated convolution $H(y)$ can be defined as,

$$H(y) = (y * W + b) \otimes \sigma(y * V + c) \tag{9}$$

where y is the input information of the layer, W, V, b, c are learnable parameters and $\otimes$ is the element-wise product between matrices. In addition, σ is the Sigmoid activation function. The expressions can be indicated by the following formula.

$$sigmiod(y) = \frac{1}{1+e^{-y}} \tag{10}$$

The spatial module of GCNN can extract spatial correlations between any two nodes in the traffic network. The spectral convolution of the graph $G(x)$ is defined as the operation of the signal $x \in R^N$ and the filter $g_\theta = diag(\theta)$,

$$G(x) = g_\theta * x = U g_\theta(L) U^T x \tag{11}$$

where θ is the learnable parameters, and U is the eigenvector of the normalized Laplacian. However, the Laplace eigenvalue decomposition results in high complexity in the convolution computation. Traditionally, the convolution kernel is replaced by the Chebyshev polynomial. The convolution formula in the spectrum domain can be obtained as follows.

$$G(x) \approx \sum_{k=0}^{K-1} \theta_k' \, T_k(\tilde{\mathrm{L}}) x \tag{12}$$

According to Yu et al. (2018), the GCN model integrates both spatial and temporal domain features. For the input value x^l of layer l, the prediction value x^{l+1} is denoted by

$$x^{l+1} = H_1^l\left(ReLU\left(G^l H_0^l(x^l)\right)\right) \tag{13}$$

In the process, the L2 loss function is adopted to train this model,

$$L(n, \hat{n}) = \omega(\theta)(\hat{n} - n)^2 \tag{14}$$

where θ is trainable parameters, n is the real data and $\hat{n}$ is the prediction data.

4 Modeling results

4.1 Data source

In the case study, the Similarity-based Principle Spatio-temporal Graph Convolutional Neural Network (SP-STGCNN) model predicts the demand for bicycle sharing. The ridership data are provided by Ford Go Bike, a regional public bicycle sharing program in the San Francisco. The data are collected hourly from April 1 to July 31, 2018, and contain 87 weekdays and 35 weekend days. The area with 122 docking stations is selected for the research. The descriptive statistics for ridership data for each day in docking stations are presented in Table 2. Bicycle docking stations and geographic elements are used to demarcate the land-use zones, which are similar to traffic analysis zones (TAZ). Each zone has independent road network information and land-use characteristics and intensity based on the map. Three rules are applied in the delimitation. First, each zone has only one docking station for shared bicycles. Second, the zones are separated by geographical elements such as roads, railways, rivers. Finally, the study area (Figure 7) is completely covered by the zones, no space is left out. In addition, part of the road network information is depicted in Figure 8.

Table 2. Descriptive statistics of ridership data

	Weekday		Weekend	
	Attraction	Generation	Attraction	Generation
Total	410008	277370	24689	84897
Mean	3360.72	2273.53	202.37	695.88
Std. Dev.	2772.74	2132.19	159.40	497.01
Minimum	243	526	28	132
25% Percentile	1637	1011	110	426
Median	2587.5	1739	167.5	577.5
75% Percentile	3813	2574	250	818
Maximum	12681	11423	1133	3223
Range	12438	10897	1105	3091

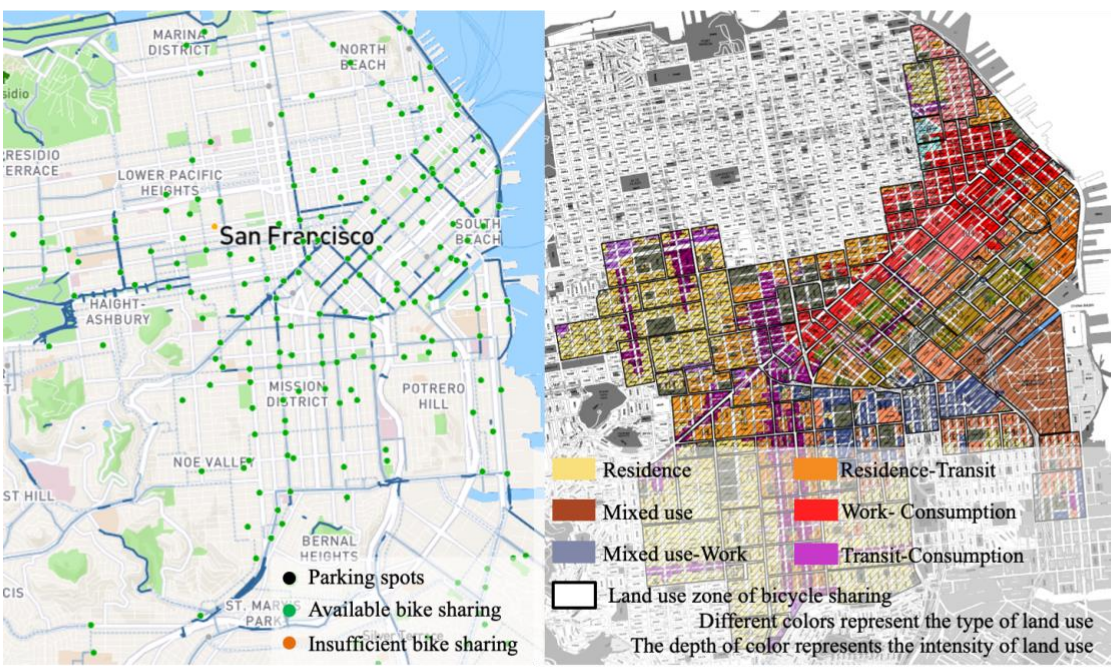

Figure 7. Docking stations and land-use map (D-1) of bicycle sharing in the San Francisco

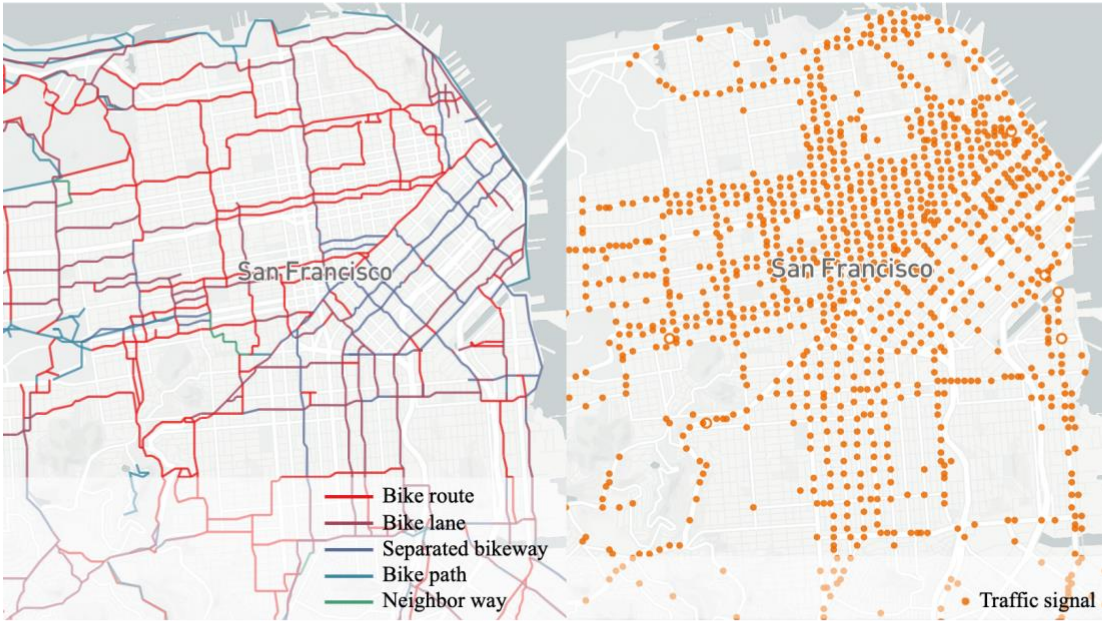

Figure 8. The bikeway network and traffic signal in the study area

4.2 Experimental results

4.2.1 Geographic information results

A general statistic summary of spatial geographic attributes-based activity data for two datasets and one road network information dataset is given in Table 3. The geographic attributes in dataset 1 are work, consumption, resident, and transit. For example, the work attribute indicates the work-related human activities within a zone. Land-use categories in dataset 2 are as follows: Cultural Institutional Educational (CIE), Medical (MED), Management Information Professional Services (MIP), Retail Entertainment (RETAIL), Production Distribution Repair Industrial (PDR), Residential (RES), and Hotels Visitor Services (VISITOR). The road network information in dataset 3 is including bikeway network, street intersection, traffic signal, traffic stop and the number, type and length of road segment.

Table 3. Descriptive statistics of variables

Variable	Description	Min	Max	Mean	S.D.
Dataset 1 – related variables					
Work	The type of work-related human activities in each TAZ	0	100	30.66	25.65
Consumption	The type of consumption-related human activities in each TAZ	1.09	67.83	32.15	18.07
Resident	The type of resident-related human activities in each TAZ	0.44	74.41	29.56	17.35
Transit	The type of transit-related human activities in each TAZ	0	87.41	25.68	19.71
Dataset 2 – related variables					
CIE index	The type of cultural, institutional, and educational in each TAZ	0	434782	48641.32	82964.95
MED index	The type of medicine in each TAZ	0	838225	20259.38	50785.60
MIPS index	The type of management, information, and professional services in each TAZ	0	1979905	141318.83	200079.01
PDR index	The type of production, distribution, and repair in each TAZ	0	515686	52926.50	85792.67
RES index	The type of residence in each TAZ	0	2371	371.59	322.71
RETAIL index	The type of retail, and entertainment in each TAZ	0	656973	48562.81	54346.64
VISITOR index	The type of hotels and visitor services in each TAZ	0	182184	4445.26	13475.95
Dataset 3 – related variables					
Total roads	Road segments count in each TAZ	0	92	20.27	15.85
Total length	Road segments length in each TAZ (ft)	0	60759.21	14268.84	8378.91
Bikeway percentage	Bikeway segments /total road segments in 0 each TAZ (%)	0	0.26	0.17	0.21
Street percentage	Street segments /total road segments in 0 each TAZ (%)	0	0.92	0.48	0.53
Avenue percentage	Avenue segments /total road segments in 0 each TAZ (%)	0	0.76	0.26	0.36
Intersection	The number of intersections in each TAZ	47	86	16.38	14.69
Stop	Total traffic stop/intersection in each 0 TAZ (%)	0	0.76	0.57	0.49
Signal	Total traffic signal/intersection in each 0 TAZ (%)	0	1	0.12	1.21

4.2.2 Parameters settings

Pytorch is used to build the SAE model and the ST-GCN model. For the temporal module, contemporaneous historical data and short-term data are compressed by the SAE model. For the spatial module, three datasets are used to measure spatial correlation. These data are processed based on Python 3.7, ArcGIS 10.2, and MATLAB 2020b. All the experiments are conducted on a Windows server (IntelI CoITM) i5-7200U CPU @2.50GHz 2.71 GHz). Of the data, 70% are selected to train the model, 20% to validate the set, and the remaining 10% to test the set. In terms of the SAE model, the epoch is 120, the batch size is 64, and the learning rate is 0.005. The encoder structure is 30-64-32-6-32-64-30, and the activate function is RuLU. To find an optimal structure for the neural network, training algorithm ADAM and several key hyper-parameters are determined based on comparative experiments, where epoch is 160, the batch size is 64 in the hidden layers, the learning rate is 10^{-3}.

4.2.3 Temporal performance of SP-STGCNN

This paper uses the dataset 1 to measure spatial correlation and shows the result of the temporal performance of the proposed model. Here, four docking stations, i.e., Mechanics Monument Plaza (ID 12), Valencia St at 21st St (ID 65), Townsend St at 7th St (ID 94), and S Park St at 3rd St (ID 118) are chosen to compare the prediction results. Figure 9 shows the predicted ridership on 24 July 2008 using different methods.

Compared with baseline models (ARIMA, SVM, LSTM, and STGCNN), the proposed model has a better performance. For the conventional statistical, the average differences in MAE/RMSE between SP-STGCNN and ARIMA for the generation and attraction data are 39.59%/44.36% and 37.57%/34.21%, indicating that the proposed model increased the prediction accuracy by 38.58% when compared with the ARIMA model. For weekdays, the proposed model improved the prediction accuracy by 12.73%/14.53%, 15.83%/15.86%, and 17.13%/17.56% when compared with the STGCNN, LSTM, and SVM. For weekends, the proposed model improved the prediction accuracy by 12.93%/14.65%, 15.26%/16.13%, and 16.48%/16.81% when compared with the STGCNN, LSTM, and SVM. The SP-STGCNN model captured the peak values with less delay and better accuracy when compared with the deep learning models due to the inclusion of long-term information and historical periodic data. Meanwhile, it can denoise and extract the key features and help traffic demand prediction in the real world.

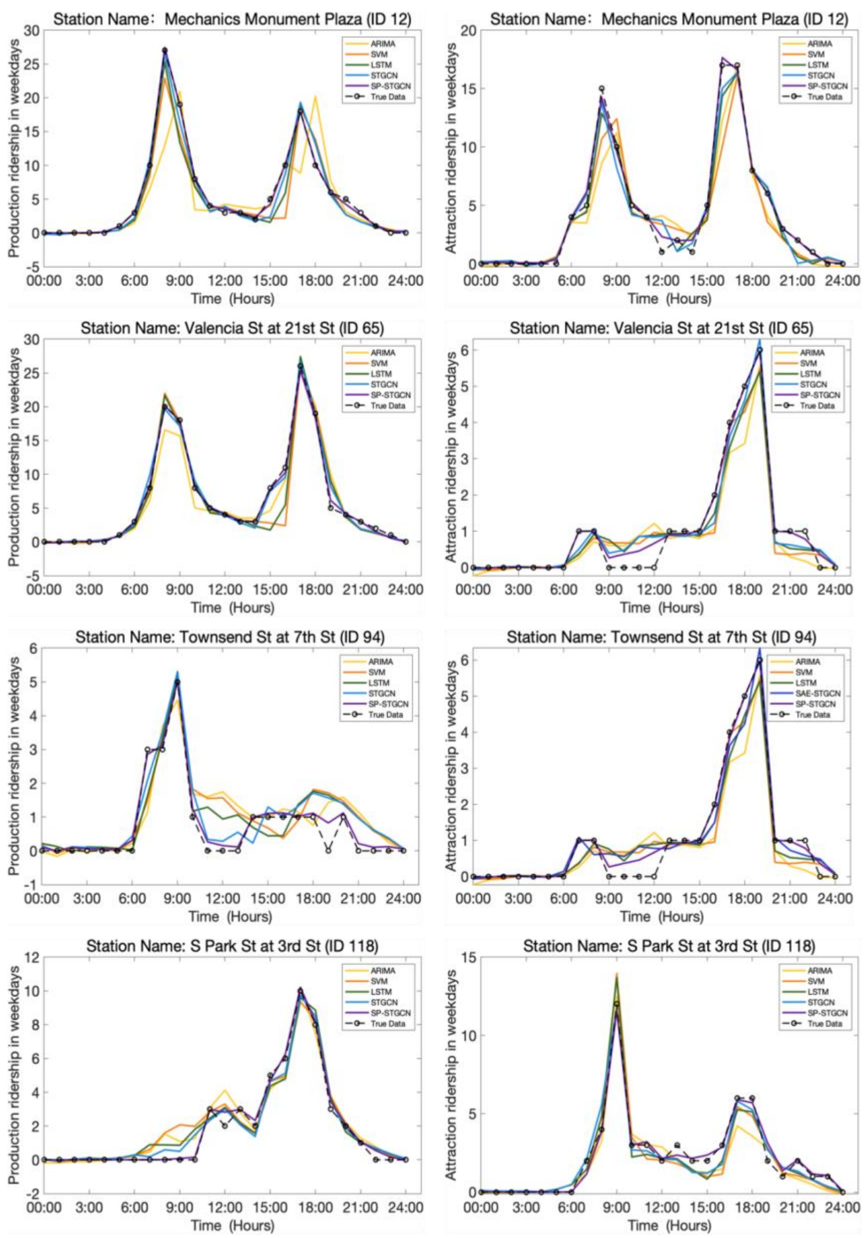

Figure 9. Prediction of ridership on weekdays

When different time intervals are applied in the SP-STGCNN model (Table 4), the model performance increased with the length of the time interval. As we have known, LSTM produces only one value of time granularity hence is suitable for short-term prediction. In this study, the sliding time window method is used to show the short-, medium-, and long-term prediction. The prediction accuracy improves when the interval shortens, with a good performance in the medium- and long-term predictions.

Table 4. The prediction errors for the different time interval

Time Interval	Weekdays				Weekends			
	Attraction		Generation		Attraction		Generation	
	MAE	RMSE	MAE	RMSE	MAE	RMSE	MAE	RMSE
1 Hour	1.110	1.766	1.104	1.782	1.047	1.767	0.965	1.784
2 Hour	1.112	1.837	1.108	1.804	1.013	1.783	1.089	1.798
3 Hour	1.118	1.884	1.103	1.802	1.110	1.811	1.133	1.831
4 Hour	1.127	1.881	1.141	1.836	1.141	1.854	1.158	1.836
5 Hour	1.154	1.926	1.156	1.924	1.169	1.890	1.160	1.853
6 Hour	1.169	1.984	1.167	1.905	1.188	1.907	1.167	1.871

4.2.4 Spatial performance of SP-STGCNN

The study area is divided into 6 blocks according to the administrative region to analyze the effect of the spatial attributes on the proposed model and demonstrate the impacts of the different spatial correlations on the predict based dataset 1. The deeper the color (orange) is, the greater the prediction errors are. As shown in Figure 10, districts 1 and 6 are strongly affected by the work attribute and the transit attribute. Districts 2, 3, and 4 are strongly influenced by the consumption attribute and the resident attribute. District 5 is deeply affected by the work attribute and the resident attribute.

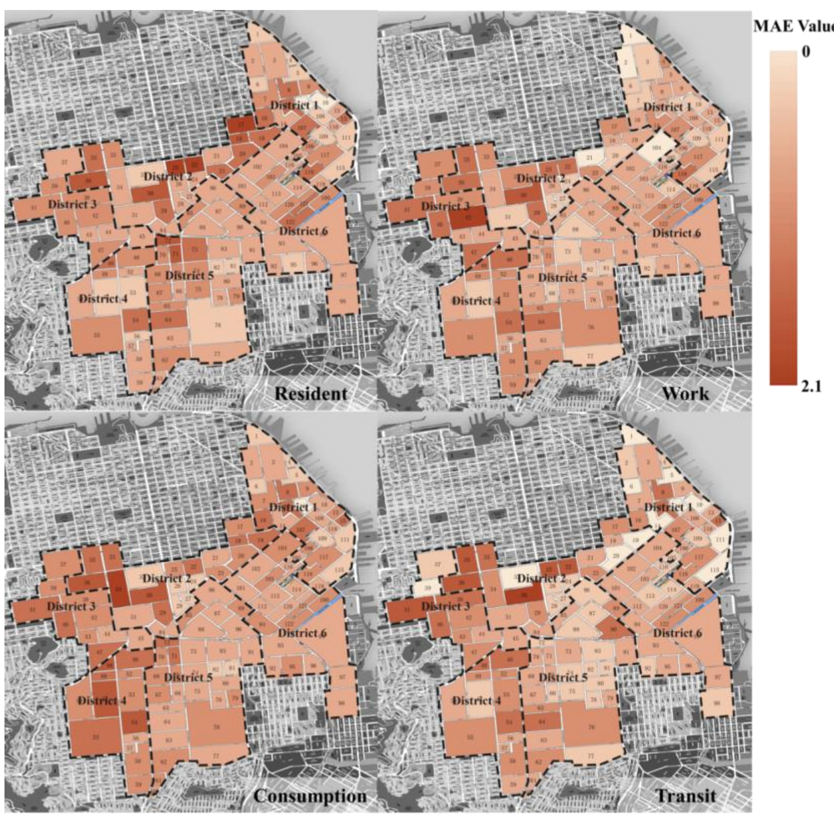

Figure 10. Prediction errors (MAE) of generation on weekdays

The impact of spatial attributes on the prediction errors is shown in Figure 11. The darker the color (blue) is, the greater the prediction errors are. Compared with the distance attribute, the ABG-information synthesis attribute as the weighted adjacency matrix of the model produced better performance. When predicting trip generation and attraction on weekdays, the accuracy of the model using the ABG-information synthesis attribute has increased the model accuracy by 10.627% and 13.058% than using the distance attribute, respectively. When predicting trip generation and attraction on weekends, the increase is 10.179% and 12.962%, respectively. The key element of spatial association in the network is the human activities in the region, which reflect the lifestyle of the residents. Therefore, using the activity-based geographic information synthesis attribute as the adjacency matrix is more suitable for the demand predicting model.

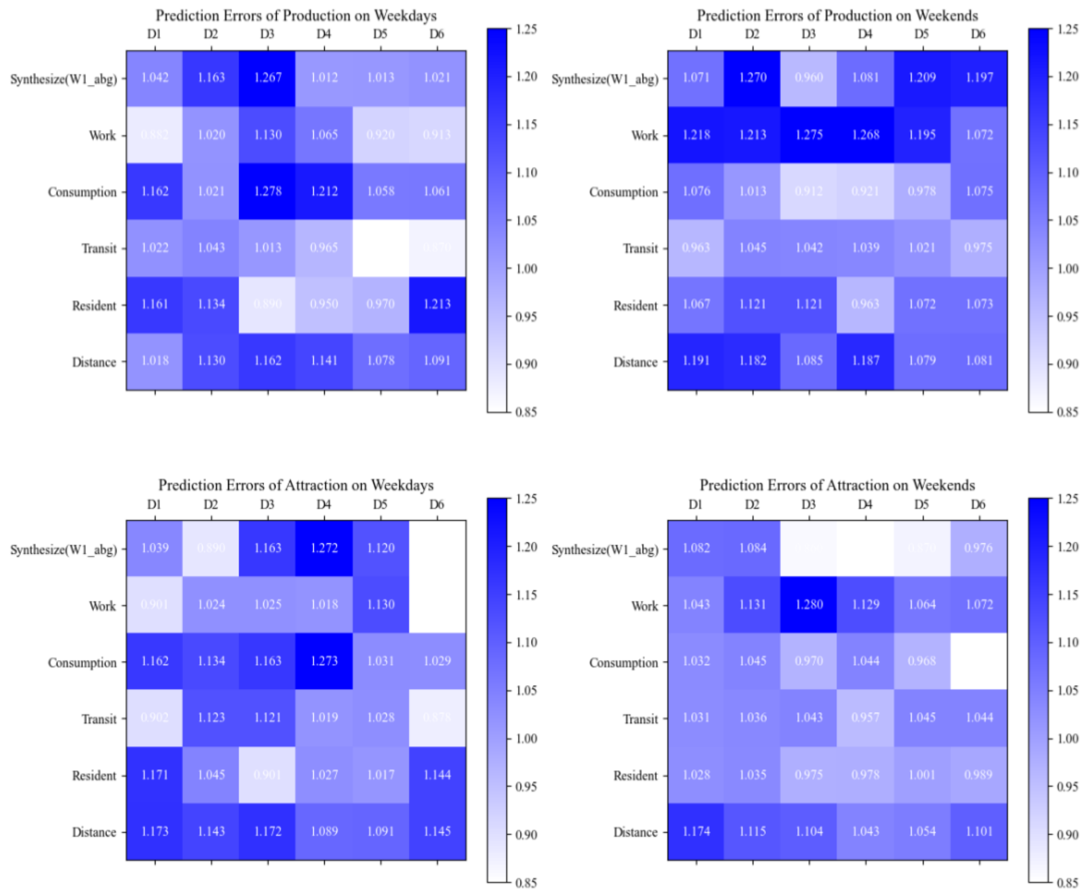

Figure 11. Different spatial correlations affect prediction results

The best prediction results are produced when the transit attribute is applied because the bicycle-sharing program is mainly designed to serve the Central Business District (CBD) in San Francisco as a convenient travel choice. The city has many metro lines and bus routes, and a bicycle sharing program plays a vital role in connecting various transportation services.

The prediction results for districts 5 and 6 are better on weekdays than on weekends (Figure 6). This fits with the general lifestyle because people travel between home and workplace on weekdays and carry out consumption (e.g., leisure and shopping) activities on weekends. Therefore, a more accurate prediction on weekdays is obtained when work and residence attributes are applied, and the application of consumption attributes would improve the accuracy of predictions on weekends.

Districts 3 and 4 are resident areas. For the same reason, the proposed model produced better prediction results when the residence attribute is considered instead of other ABG-information attributes. Hence, the application of the dominant activity data as the spatial relationship matrix in the proposed model can improve performance and reduce errors.

On the other hand, the prediction errors of trip generation are higher than that of trip attractions on weekends. This indicates that people tend to choose faster and less effort means of transportation for outbound trips and a healthier, greener, and cheaper means for inbound trips. However, when work attributes are applied, no significant differences between trip generation/attraction on weekdays are revealed. It suggests that these trips are more likely to be commuting journeys. Meanwhile, the application of the transit attribute does not produce a different prediction for trip generation/attraction on weekdays from that on weekends. Such results indicate that this proposed model has strong interpretability and reliability.

4.2.5 Temporal similarity analysis of SP-STGCNN

To reflect the performance of temporal similarity to the model, this paper designs the Stacked Autoencoder-Spatiotemporal Graph Convolution Neural Network (SAE-STGCNN) model for comparison. The SAE-STGCN model compresses a period of historical data by SAE to demonstrate the effect of the similarity principle proposed in this paper. Specifically, the shared bicycle ridership data X_t^n contains long-term information and short-term information, which presents as $X_t^n = \left[{X_{l}}_t^n, {X_{s}}_t^n\right]$. The long-term information ${X_{l}}_t^n$ contains a period of historical data, which can be defined as ${X_{l}}_t^n = [x_{t-L-d}^n, \dots, x_{t-L-3}^n, x_{t-L-2}^n, x_{t-L-1}^n]$, where d means the historical time step. The SAE model is chosen to reduce the dimension of ${Z_{l}}_t^n$ to concentrate on the long-term information ${X_{l}}_t^n$. The input data is presented as ${X_{in}}_t^n = \left[{Z_{l}}_t^n, {X_{s}}_t^n\right]$. The rest of the process is similar to Section 3.

Table 5. Different temporal similarities affect prediction results

Spatial correlation	Weekdays				Weekends			
	Attraction		Generation		Attraction		Generation	
	MAE	RMSE	MAE	RMSE	MAE	RMSE	MAE	RMSE
STGCNN	1.243	1.951	1.231	1.958	1.159	1.975	1.091	1.943
SAE-STGCNN	1.126	1.812	1.116	1.834	1.054	1.811	0.982	1.809
SP-STGCNN	1.110	1.766	1.104	1.782	1.047	1.767	0.965	1.784

The prediction results are as follows. For weekdays, the proposed model improved the prediction accuracy by 8.29%/8.74% (MAE/RMSE) and 10.73%/11.53% when compared with the SAE-STGCNN and STGCNN. For weekends, the proposed model improved the prediction accuracy by 9.32%/10.98% and 12.93%/14.65% when compared with the SAE-STGCNN and STGCNN. The ranking results of the prediction errors of the three models are as follows: SP-STGCNN < SAE-STGCNN < STGCNN. The prediction accuracy of the two models (SP-STGCNN and SAE-STGCNN) is better than that of STGCNN. Processing temporal information based on the principle of similarity can effectively improve the performance of the model and the higher the similarity of temporal information, the better the prediction accuracy.

4.2.6 Spatial similarity analysis of SP-STGCNN

The four datasets are used to describe the spatial correlations between different docking stations. The prediction errors of the SP-STGCNN model using different spatial correlations are as follows: SP-STGCNN based $W_{1_{abg}}$ < SP-STGCNN based $W_{2_{abg}}$ < SP-STGCNN based W_{dis} < SP-STGCNN based R_{info}. As the travel mode of the last kilometer, the demand for shared bicycles will be affected by the distance factor and road network characteristics (Guo et al., 2022; Jiao et al., 2021; Bai et al, 2021, El-Assi et al, 2017). However, when using geographic information to describe the spatial similarity of different stations, the prediction accuracy is higher. Therefore, the main spatial factor affecting the demand for shared bicycles is the geographic information characteristics. On the other hand, Dataset 1 containing transit-related information such as transport level, transit lines, and station density outperforms Dataset 2, which lacks such specifics. The results emphasizing factors like road design and infrastructure could be crucial for accurately predicting micro-mobility demand.

Table 6. Different spatial similarity correlations affect prediction results

Spatial correlation	Weekdays				Weekends			
	Attraction		Generation		Attraction		Generation	
	MAE	RMSE	MAE	RMSE	MAE	RMSE	MAE	RMSE
Road network	1.321	1.938	1.323	1.901	1.421	1.985	1.231	1.987
Distance	1.313	1.932	1.321	1.898	1.342	1.963	1.202	1.906
$W_{2_{abg}}$	1.223	1.828	1.218	1.791	1.108	1.811	1.181	1.897
$W_{1_{abg}}$	1.110	1.766	1.104	1.782	1.047	1.767	0.965	1.784

4.2.7 Interpretable analysis of SP-STGCNN

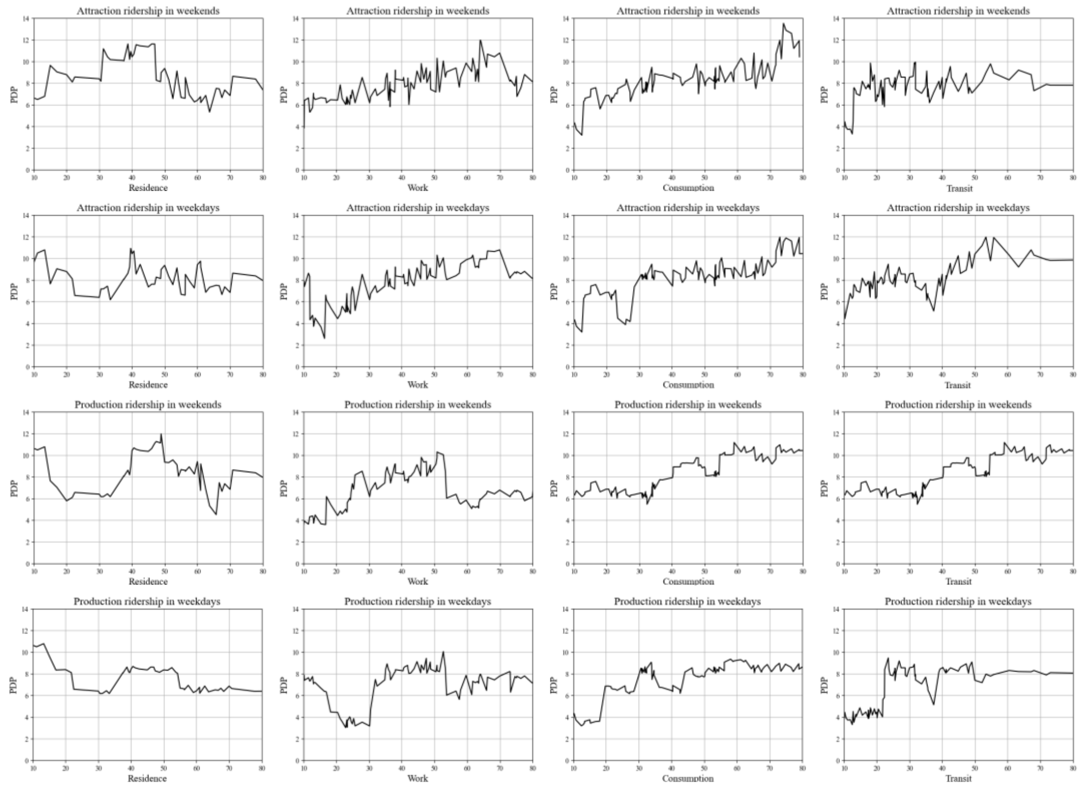

Figure 12. PDPs of the land-use intensity

Partial Dependence Plots (PDPs) are valuable tools for visualizing the overall marginal impact of features on predictions. This paper utilizes PDP analysis to indicate the influence of land-use variables on trips. Figure 12 in the paper illustrates PDPs depicting the correlation between land-use intensity and trips. In the figure, the x-axis represents land-use intensity distribution, with tick marks denoting varying levels. For brevity, we present results only for the top-performing model. The PDPs for land use reveal crucial insights into the relationships.

Consumption and work with attraction showed positive associations, although most of their trends were not strictly monotonic but exhibited threshold effects. Specifically, they exhibited positive relationships within specific intervals but remained steady in other ranges. For instance, consumption intensity exhibited a positive connection only when surpassing 40, and work intensity displayed a positive association only between 30 and 50.

Resident and work with production presented negative relations with ridership. Similarly, these relationships were not strictly monotonic. For instance, resident intensity showed a negative relation only when below 30, and work intensity exhibited a negative relation between 50 and 70.

Other variables, mainly transit with weekdays, presented more complex nonlinear relations with population inflow. For instance, transit intensity exhibited a downward parabolic pattern when situated between 20 and 30.

Several noteworthy observations have been made. Firstly, there exists an intermediate phase of stagnation before reaching a plateau in the relationship between production ridership and resident. These peculiar patterns are likely attributable to the interdependence of features, a phenomenon well-documented in previous studies employing PDP to delineate correlations (Molnar, 2020). Secondly, notable irregularities, such as abrupt drops or spikes, are discernible in specific relationships, such as that between attraction ridership and consumption. These deviations are likely influenced by outliers. Moreover, protracted plateaus, both preceding and following these irregularities, are consistently observed across most of these relationships. The origin of these plateaus, whether they stem from the presence of outliers or signify genuine threshold effects, remains a subject of uncertainty.

5 Conclusions

This study developed a Similarity-based Principle - Spatio-temporal Graph Convolutional Neural Network (SP-STGCNN) model to predict the ridership of shared bicycles using the deep learning approach. Data are provided by Ford Go Bike in the San Francisco. The Similarity-based feature extraction method is integrated with the activity-based spatial geographic information to reflect human beings' daily lives and improve the spatial correlation of the model. The SAE method is employed to improve prediction accuracy and train efficiency by reducing the dimension of contemporaneous historical data and getting long-term traffic data information.

Experimental results show that the proposed model can predict the bicycle sharing ridership generation and attraction citywide in the future. The distribution of the error terms can be used to analyze zones where the supply of shared bicycles may be exceeded or lacking. The comparison results suggest that the SP-STGCNN model provides higher prediction accuracy than commonly used statistical models and machine learning algorithms. Specifically, it is 12.83%, 15.53%, 16.31%, and 38.58% higher than that of the STGCNN, LSTM, SVM, and ARIMA, respectively. The spatial analyses are carried out using the activity-based dataset 1 for four categories: work, transportation, residence, and consumption. For the spatial weighted adjacency matrix, when the dominant human activities are selected accordingly (i.e., work or residence attributes on weekdays, and consumption attributes on weekends), the prediction error is reduced. Importantly, for the similarity of the temporal module, the performance of the proposed model is 10.31% and 11.43% higher than that of the SAE-STGCNN, STGCNN. Periodic historical series, rather than a period, can better capture the long-term characteristics of data volatility. For the similarity of the spatial module, the model with the application of the ABG-information attribute improved the performance by 13.26% and 11.63% when compared with the application of the road information attribute and the distance attribute.

Moreover, this paper supplements the analysis of the results obtained by applying SP-STGCNN to address the three gaps highlighted in the literature review. Firstly, by incorporating the SP method into the well-performing STGCN model, we observe a significant improvement of 12.83%. This enhancement highlights the effectiveness of the SP approach in optimizing model performance. Secondly, we utilize the integration of multi-source data with the temporal-spatial modules. To capture both short-term and long-term traffic data information, we employ the SAE method, resulting in a notable improvement of 10.31%. Furthermore, we apply activity-based data to measure spatial similarity, leading to a 6.7% improvement. Lastly, in the context of demand prediction, we introduce an innovative approach by constructing activity-based data, replacing the conventional distance-based data in flow prediction for graph network structure construction. This change also contributes to a 6.7% improvement, highlighting the advantages of this novel approach in demand prediction. Such results indicate that the similarity-based principle is helpful in improving the performance of the Spatio-temporal model.

While our research has primarily focused on addressing gaps in existing methodologies, it is crucial to emphasize the practical implications and benefits that SP-STGCNN brings to the domain of urban planning and decision-making. By improving the accuracy of traffic prediction and modeling, our model equips urban planners and policymakers with a more robust understanding of traffic patterns, congestion dynamics, and demand fluctuations. This, in turn, enables better-informed decisions regarding infrastructure development, traffic management strategies, and resource allocation. Moreover, multi-source data integration, including activity-based data and the SAE method, contributes to a more comprehensive assessment of short-term and long-term traffic trends. This holistic view empowers urban planners to devise strategies that consider not only immediate traffic issues but also the long-range implications of their decisions. The innovative use of activity-based data for constructing graph network structures in demand prediction further enhances the accuracy of future traffic forecasts. This innovation is precious for urban planning, as it facilitates more precise estimations of transportation needs, optimizing resource allocation and investment in transportation infrastructure. In summary, our SP-STGCNN model addresses existing research gaps and offers a practical and impactful tool for urban planning decision-makers. By improving the accuracy and comprehensiveness of traffic analysis and prediction, this model contributes to more effective and informed planning strategies, ultimately leading to more efficient and sustainable urban development.

In conclusion, the development and implementation of the SP-STGCNN model represent an improvement in addressing gaps within urban transportation prediction methodologies. While the model showcases promising predictive capabilities in forecasting bike-sharing ridership, several limitations and ethical considerations warrant attention for future research. The reliance on localized data from Ford Go Bike in San Francisco, although instrumental, limits the model's generalizability. Diversifying datasets encompassing various cities and bike-sharing systems would enhance their applicability and robustness. Additionally, the model's capacity to adapt to sudden shifts in user behavior or unforeseen events requires further exploration to bolster its real-time adaptability. Ethical implications surrounding biases in data and algorithmic fairness underscore the need for rigorous assessment and mitigation strategies. Future research avenues should focus on expanding datasets, enhancing real-time adaptability, addressing biases, and fostering interpretability. By incorporating these directions, the SP-STGCNN model can evolve into a more versatile, ethical, and reliable tool for urban planners and decision-makers, facilitating more informed and equitable urban development strategies.

Acknowledgments

This work was supported by the National Natural Science Foundation of China (51925801, 52232012) and Postgraduate Research & Practice Innovation Program of Jiangsu Province (KYCX22_0274).

References

Alzubaidi, L., Zhang, J., Humaidi, A. J., Al-Dujaili, A., Duan, Y., Al-Shamma, O., … & Farhan, L. (2021). Review of deep learning: Concepts, CNN architectures, challenges, applications, future directions. *Journal of Big Data*, *8*(1), 53. https://doi.org/10.1186/s40537-021-00444-8

Bai, S., Jiao, J., Chen, Y., & Guo, J. (2021). The relationship between E-scooter travels and daily leisure activities in Austin, Texas. *Transportation Research Part D: Transport and Environment*, 95, 102844. https://doi.org/10.1016/j.trd.2021.102844

Bao, Y.-X., Shi, Q., Shen, Q.-Q., & Cao, Y. (2021). Spatial-temporal 3D residual correlation network for urban traffic status prediction. *Symmetry*, *14*(1), 33. https://doi.org/10.3390/sym14010033

Bogaerts, T., Masegosa, A. D., Angarita-Zapata, J. S., Onieva, E., & Hellinckx, P. (2020). A graph CNN-LSTM neural network for short and long-term traffic forecasting based on trajectory data. *Transportation Research Part C: Emerging Technologies*, *112*, 62–77. https://doi.org/10.1016/j.trc.2020.01.010

Chai, D., Wang, L., & Yang, Q. (2018). Bike flow prediction with multi-graph convolutional networks. *Proceedings of the 26th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems*, 397–400. https://doi.org/10.1145/3274895.3274896

Chen, D., Wang, H., & Zhong, M. (2020). *A short-term traffic flow prediction model based on AutoEncoder and GRU*. Paper presented at the12th International Conference on Advanced Computational Intelligence, March 14–16, Dali, Yunnan, China.

Chen, L., Thakuriah, P., & Ampountolas, K. (2021). Short-term prediction of demand for ride-hailing services: A deep learning approach. *Journal of Big Data Analytics in Transportation*, *3*(2), 175–195. https://doi.org/10.1007/s42421-021-00041-4

Chen, Z., Wu, H., O'Connor, N. E., & Liu, M. (2021). A comparative study of using spatial-temporal graph convolutional networks for predicting availability in bike sharing schemes. *ArXiv*: 2104.10644 [Cs, Eess]. http://arxiv.org/abs/2104.10644

Chen, Z., Zhao, B., Wang, Y., Duan, Z., & Zhao, X. (2020). Multitask learning and GCN-based taxi demand prediction for a traffic road network. *Sensors, 20*(13), 3776. https://doi.org/10.3390/s20133776

Cui, Z., Henrickson, K., Ke, R., & Wang, Y. (2020). Traffic graph convolutional recurrent neural network: A deep learning framework for network-scale traffic learning and forecasting. *IEEE Transactions on Intelligent Transportation Systems*, *21*(11), 4883–4894. https://doi.org/10.1109/TITS.2019.2950416

Cui, Z., Ke, R., Pu, Z., Ma, X., & Wang, Y. (2020). Learning traffic as a graph: A gated graph wavelet recurrent neural network for network-scale traffic prediction. *Transportation Research Part C: Emerging Technologies*, *115*, 102620. https://doi.org/10.1016/j.trc.2020.102620

Dauphin, Y. N., Fan, A., Auli, M., & Grangier, D. (2017). Language modeling with gated convolutional networks. *Proceedings of the 34th International Conference on Machine Learning, 70*, 933–941.

Do, L. N. N., Vu, H. L., Vo, B. Q., Liu, Z., & Phung, D. (2019). An effective spatial-temporal attention based neural network for traffic flow prediction. *Transportation Research Part C: Emerging Technologies*, *108*, 12–28. https://doi.org/10.1016/j.trc.2019.09.008

Du, B., Hu, X., Sun, L., Liu, J., Qiao, Y., & Lv, W. (2021). Traffic demand prediction based on dynamic transition convolutional neural network. *IEEE Transactions on Intelligent Transportation Systems*, *22*(2), 1237–1247. https://doi.org/10.1109/TITS.2020.2966498

El-Assi, W., Salah Mahmoud, M. & Nurul Habib, K. (2017). Effects of built environment and weather on bike sharing demand: A station level analysis of commercial bike sharing in Toronto. *Transportation* 44, 589–613. https://doi.org/10.1007/s11116-015-9669-z

Fang, S., Zhang, Q., Meng, G., Xiang, S., & Pan, C. (2019). GSTNet: Global spatial-temporal network for traffic flow prediction. *Proceedings of the 28th International Joint Conference on Artificial Intelligence*, 2286–2293. https://doi.org/10.24963/ijcai.2019/317

Feng, J., Chen, X., Gao, R., Zeng, M., & Li, Y. (2018). DeepTP: An end-to-end neural network for mobile cellular traffic prediction. *IEEE Network, 32*(6), 108–115. https://doi.org/10.1109/MNET.2018.1800127

Feng, S., Chen, H., Du, C., Li, J., & Jing, N. (2018). A hierarchical demand prediction method with station clustering for bike sharing system. *2018 IEEE Third International Conference on Data Science in Cyberspace (DSC)*, 829–836. https://doi.org/10.1109/DSC.2018.00133

Geng, X., Li, Y., Wang, L., Zhang, L., Yang, Q., Ye, J., & Liu, Y. (2019). Spatiotemporal multi-graph convolution network for ride-hailing demand forecasting. *Proceedings of the AAAI Conference on Artificial Intelligence*, *33*, 3656–3663. https://doi.org/10.1609/aaai.v33i01.33013656

Guo, S., Lin, Y., Wan, H., Li, X., & Cong, G. (2021). Learning dynamics and heterogeneity of spatial-temporal graph data for traffic forecasting. *IEEE Transactions on Knowledge and Data Engineering*, *34,* 5415–5428. https://doi.org/10.1109/TKDE.2021.3056502

Guo, Y., Yang, L., & Chen, Y. (2022). Bike share usage and the built environment: A review. *Frontiers in Public Health*, 10, 848169. https://doi.org/10.3389/fpubh.2022.848169

Hinton, G. E. (2006). Reducing the dimensionality of data with neural networks. *Science*, *313*(5786), 504–507. https://doi.org/10.1126/science.1127647

Hulot, P., Aloise, D., & Jena, S. D. (2018). Towards station-level demand prediction for effective rebalancing in bike-sharing systems. *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, 378–386. https://doi.org/10.1145/3219819.3219873

Jiang, J., Lin, F., Fan, J., Lv, H., & Wu, J. (2019). A destination prediction network based on spatiotemporal data for bike-sharing. *Complexity*, *2019*, 1–14. https://doi.org/10.1155/2019/7643905

Jiao, J., Bai, S., & Choi, S. J. (2021). Understanding e-scooter incidents patterns in street network perspective: A case study of Travis County, Texas. *Sustainability*. *13*(19), 10583. https://doi.org/10.3390/su131910583

Joshi, M., & Hadi, T. H. (2015). A review of network traffic analysis and prediction techniques. *ArXiv:1507.05722 [Cs]*. http://arxiv.org/abs/1507.05722

Li, Y., Yu, R., Shahabi, C., & Liu, Y. (2017). Diffusion convolutional recurrent neural network: Data-driven traffic forecasting. *ArXiv Preprint arXiv:1707.01926*.

Li, Y., Yu, R., Shahabi, C., & Liu, Y. (2018). Diffusion convolutional recurrent neural network: Data-driven traffic forecasting. *ArXiv:1707.01926 [Cs, Stat]*. http://arxiv.org/abs/1707.01926

Li, Y., & Zheng, Y. (2020). Citywide bike usage prediction in a bike-sharing system. *IEEE Transactions on Knowledge and Data Engineering*, *32*(6), 1079–1091. https://doi.org/10.1109/TKDE.2019.2898831

Li, Y., Zheng, Y., Zhang, H., & Chen, L. (2015). Traffic prediction in a bike-sharing system. *Proceedings of the 23rd SIGSPATIAL International Conference on Advances in Geographic Information Systems*, 1–10. https://doi.org/10.1145/2820783.2820837

Li, Y., Zhu, Z., Kong, D., Xu, M., & Zhao, Y. (2019). Learning heterogeneous spatial-temporal representation for bike-sharing demand prediction. *Proceedings of the AAAI Conference on Artificial Intelligence*, *33*, 1004–1011. https://doi.org/10.1609/aaai.v33i01.33011004

Lin, L., He, Z., & Peeta, S. (2018). Predicting station-level hourly demand in a large-scale bike-sharing network: A graph convolutional neural network approach. *Transportation Research Part C: Emerging Technologies*, *97*, 258–276. https://doi.org/10.1016/j.trc.2018.10.011

Liu, J., Si, Y., Niu, Y., & Zhang, R. (2022). Projection quantile correlation and its use in high-dimensional grouped variable screening. *Computational Statistics & Data Analysis*, *167*, 107369. https://doi.org/10.1016/j.csda.2021.107369

Liu, L., Qiu, Z., Li, G., Wang, Q., Ouyang, W., & Lin, L. (2019). Contextualized spatial-temporal network for taxi origin-destination demand prediction. *ArXiv:1905.06335 [Cs, Stat]*. http://arxiv.org/abs/1905.06335

Ma, Y., Zhao, X., Zhang, C., Zhang, J., & Qin, X. (2021). Outlier detection from multiple data sources. *Information Sciences*, *580*, 819–837. https://doi.org/10.1016/j.ins.2021.09.053

Maleki Vishkaei, B., Mahdavi, I., Mahdavi-Amiri, N., & Khorram, E. (2020). Balancing public bicycle sharing system using inventory critical levels in queuing network. *Computers & Industrial Engineering*, *141*, 106277. https://doi.org/10.1016/j.cie.2020.106277

Molnar, C. (2020). *Interpretable machine learning*. Morrisville, NC: Lulu.com.

Pan, X., Shi, J., Luo, P., Wang, X., & Tang, X. (2018). *Spatial as deep: Spatial CNN for traffic scene understanding*. Washington, DC: Association for the Advancement of Artificial Intelligence.

Pan, Z., Liang, Y., Wang, W., Yu, Y., Zheng, Y., & Zhang, J. (2019). Urban traffic prediction from spatio-temporal data using deep meta learning. *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, 1720–1730. https://doi.org/10.1145/3292500.3330884

Ramakrishnan, N., & Soni, T. (2018). Network traffic prediction using recurrent neural networks. *17th IEEE International Conference on Machine Learning and Applications (ICMLA)*, 187–193. https://doi.org/10.1109/ICMLA.2018.00035

Sarker, I. H. (2021). Deep learning: A comprehensive overview on techniques, taxonomy, applications and research directions. *SN Computer Science*, *2*(6), 420. https://doi.org/10.1007/s42979-021-00815-1

Saxena, D., & Cao, J. (2019). D-GAN: Deep generative adversarial nets for spatio-temporal prediction. *ArXiv*. Retrieved from https://arxiv.org/pdf/1907.08556.pdf

Shi, X., Qi, H., Shen, Y., Wu, G., & Yin, B. (2020). A spatial-temporal attention approach for traffic prediction. *IEEE Transactions on Intelligent Transportation Systems*, *22*(8), 4909–4918. https://doi.org/10.1109/TITS.2020.2983651

Sun, J., Zhang, J., Li, Q., Yi, X., Liang, Y., & Zheng, Y. (2022). Predicting citywide crowd flows in irregular regions using multi-view graph convolutional networks. *IEEE Transactions on Knowledge and Data Engineering*, *34*(5), 2348–2359. https://doi.org/10.1109/TKDE.2020.3008774

Taye, M. M. (2023). Understanding of machine learning with deep learning: Architectures, workflow, applications and future directions. *Computers*, *12*(5), 91. https://doi.org/10.3390/computers12050091

Triguero, I., Figueredo, G. P., Mesgarpour, M., Garibaldi, J. M., & John, R. I. (2017). Vehicle incident hot spots identification: An approach for big data. *2017 IEEE Trustcom/BigDataSE/ICESS*, 901–908. https://doi.org/10.1109/Trustcom/BigDataSE/ICESS.2017.329

Wang, W., Zhao, X., Gong, Z., Chen, Z., Zhang, N., & Wei, W. (2021). An attention-based deep learning framework for trip destination prediction of sharing bike. *IEEE Transactions on Intelligent Transportation Systems*, *22*(7), 4601–4610. https://doi.org/10.1109/TITS.2020.3008935

Wang, X., Guan, X., Cao, J., Zhang, N., & Wu, H. (2020). Forecast network-wide traffic states for multiple steps ahead: A deep learning approach considering dynamic non-local spatial correlation and non-stationary temporal dependency. *Transportation Research Part C: Emerging Technologies*, *119*, 102763. https://doi.org/10.1016/j.trc.2020.102763

Wu, Z., Pan, S., Long, G., Jiang, J., & Zhang, C. (2019). Graph WaveNet for deep spatial-temporal graph modeling. *Proceedings of the 28th International Joint Conference on Artificial Intelligence*, 1907–1913.

Xiao, G., Wang, R., Zhang, C., & Ni, A. (2021). Demand prediction for a public bike sharing program based on spatio-temporal graph convolutional networks. *Multimedia Tools and Applications*, *80*(15), 22907–22925. https://doi.org/10.1007/s11042-020-08803-y

Xu, C., Ji, J., & Liu, P. (2018). The station-free sharing bike demand forecasting with a deep learning approach and large-scale datasets. *Transportation Research Part C: Emerging Technologies*, *95*, 47–60. https://doi.org/10.1016/j.trc.2018.07.013

Yang, S. (2013). On feature selection for traffic congestion prediction. *Transportation Research Part C: Emerging Technologies*, *26*, 160–169. https://doi.org/10.1016/j.trc.2012.08.005

Yao, H., Tang, X., Wei, H., Zheng, G., & Li, Z. (2018). Revisiting spatial-temporal similarity: A deep learning framework for traffic prediction. *ArXiv:1803.01254 [Cs]*. http://arxiv.org/abs/1803.01254

Yao, H., Tang, X., Wei, H., Zheng, G., & Li, Z. (2019). Revisiting spatial-temporal similarity: A deep learning framework for traffic prediction. *Proceedings of the AAAI Conference on Artificial Intelligence*, *33*, 5668–5675. https://doi.org/10.1609/aaai.v33i01.33015668

Yao, H., Wu, F., Ke, J., Tang, X., Jia, Y., Lu, S., … & Li, Z. (2019). Deep multi-view spatial-temporal network for taxi demand prediction. *Proceedings of the AAAI Conference on Artificial Intelligence*, *32*(1), 2588–2595.

Yao, W., & Qian, S. (2021). From Twitter to traffic predictor: Next-day morning traffic prediction using social media data. *Transportation Research Part C: Emerging Technologies*, *124*, 102938. https://doi.org/10.1016/j.trc.2020.102938

Yin, X., Wu, G., Wei, J., Shen, Y., Qi, H., & Yin, B. (2021). Multi-stage attention spatial-temporal graph networks for traffic prediction. *Neurocomputing*, *428*, 42–53. https://doi.org/10.1016/j.neucom.2020.11.038

Yu, B., Yin, H., & Zhu, Z. (2018). Spatio-temporal graph convolutional networks: A deep learning framework for traffic forecasting. *Proceedings of the 27th International Joint Conference on Artificial Intelligence*, 3634–3640. https://doi.org/10.24963/ijcai.2018/505

Yuan, H., Li, G., Bao, Z., & Feng, L. (2021). An effective joint prediction model for travel demands and traffic flows. *2021 IEEE 37th International Conference on Data Engineering (ICDE)*, 348–359. https://doi.org/10.1109/ICDE51399.2021.00037

Zhang, D., Xiao, F., Shen, M., & Zhong, S. (2021). DNEAT: A novel dynamic node-edge attention network for origin-destination demand prediction. *Transportation Research Part C: Emerging Technologies*, *122*, 102851. https://doi.org/10.1016/j.trc.2020.102851

Zhang, K., Liu, Z., & Zheng, L. (2020). Short-term prediction of passenger demand in multi-zone level: Temporal convolutional neural network with multi-task learning.

IEEE Transactions on Intelligent Transportation Systems, *21*(4), 1480–1490. https://doi.org/10.1109/TITS.2019.2909571

Zhao, J., Bi, H., & Chen, K. (2022). Identifying land-use types and intensity using traffic volume and traffic event data based on ensemble learning approaches. Paper presented at the Transportation Research Board 101st Annual Meeting, January 9–13, Washington DC.

Zhao, J., Fan, W., & Zhai, X. (2020). Identification of land-use characteristics using bicycle sharing data: A deep learning approach. *Journal of Transport Geography*, *82*, 102562. https://doi.org/10.1016/j.jtrangeo.2019.102562

Zhao, L., Song, Y., Zhang, C., Liu, Y., Wang, P., Lin, T., … & Li, H. (2020). T-GCN: A temporal graph convolutional network for traffic prediction. *IEEE Transactions on Intelligent Transportation Systems*, *21*(9), 3848–3858. https://doi.org/10.1109/TITS.2019.2935152

Zheng, C., Fan, X., Wang, C, & Qi, J. (2020). GMAN: A graph multi-attention network for traffic prediction. *Proceedings of the AAAI Conference on Artificial Intelligence, 34*(1), 1234-1241.

Zi, W., Xiong, W., Chen, H., & Chen, L. (2021). TAGCN: Station-level demand prediction for bike-sharing system via a temporal attention graph convolution network. *Information Sciences, 561*, 274–285. https://doi.org/10.1016/j.ins.2021.01.065