Contents lists available at ScienceDirect

Transportation Research Part C

Improving short-term bike sharing demand forecast through an irregular convolutional neural network

Xinyu Li [a], Yang Xu [a,b,*], Xiaohu Zhang [c], Wenzhong Shi [a,d], Yang Yue [e], Qingquan Li [e,f]

[a] *Department of Land Surveying and Geo-Informatics, The Hong Kong Polytechnic University, Hong Kong, China*
[b] *The Hong Kong Polytechnic University Shenzhen Research Institute, Shenzhen, China*
[c] *Department of Urban Planning and Design, The University of Hong Kong, Hong Kong, China*
[d] *Smart Cities Research Institute, The Hong Kong Polytechnic University, Hong Kong, China*
[e] *Department of Urban Informatics, School of Architecture and Urban Planning, Shenzhen University, Shenzhen 518060, China*
[f] *Guangdong Laboratory of Artificial Intelligence and Digital Economy (SZ), Shenzhen 518107, China*

ARTICLE INFO

ABSTRACT

As an important task for the management of bike sharing systems, accurate forecast of travel demand could facilitate dispatch and relocation of bicycles to improve user satisfaction. In recent years, many deep learning algorithms have been introduced to improve bicycle usage forecast. A typical practice is to integrate convolutional (CNN) and recurrent neural network (RNN) to capture spatial–temporal dependency in historical travel demand. For typical CNN, the convolution operation is conducted through a kernel that moves across a "matrix-format" city to extract features over spatially adjacent urban areas. This practice assumes that areas close to each other could provide useful information that improves prediction accuracy. However, bicycle usage in neighboring areas might not always be similar, given spatial variations in built environment characteristics and travel behavior that affect cycling activities. Yet, areas that are far apart can be relatively more similar in temporal usage patterns. To utilize the hidden linkage among these distant urban areas, the study proposes an irregular convolutional Long-Short Term Memory model (IrConv+LSTM) to improve short-term bike sharing demand forecast. The model modifies traditional CNN with irregular convolutional architecture to leverage the hidden linkage among "semantic neighbors". The proposed model is evaluated with a set of benchmark models in five study sites, which include one dockless bike sharing system in Singapore, and four station-based systems in Chicago, Washington, D.C., New York, and London. We find that IrConv+LSTM outperforms other benchmark models in the five cities. The model also achieves superior performance in areas with varying levels of bicycle usage and during peak periods. The findings suggest that "thinking beyond spatial neighbors" can further improve short-term travel demand prediction of urban bike sharing systems.

1. Introduction

Shared bicycles have received increasing attention in urban transportation during the past few decades (Midgley, 2011; Larsen, 2013). As a green transport option for short-distance travel in cities, bike-sharing services can reduce carbon emissions and enhance last-mile connectivity to public transit (Litman and Burwell, 2006; Steg and Gifford, 2005; Haghshenas and Vaziri, 2012; Gammelli et al., 2022). During COVID-19 pandemic, bike-sharing is found to be a more resilient mode that can mitigate the fear of

---

overcrowding in public transit (Jobe and Griffin, 2021; Hu et al., 2021; Kim, 2021). Given the importance of bike-sharing services in urban transportation, accurate demand forecasting is crucial for effective rebalancing in daily operations. Many studies attempted to develop frameworks that can accurately estimate the bicycle demand throughout the city by applying traditional and machine learning models (Raviv and Kolka, 2013; Dell'Amico et al., 2018, 2014; Singhvi et al., 2015; Kumar and Jain, 1999; Avuglah et al., 2014; Billings and Yang, 2006; Lee and Fambro, 1999)

In recent years, deep learning approaches have been widely used for predicting short-term traffic demand (Ai et al., 2019; Pan et al., 2019; Li et al., 2017; Zhang et al., 2017; Li et al., 2021; Zhang et al., 2020b; Ren et al., 2020; Sathishkumar et al., 2020; Yang et al., 2020). A critical task is to model the spatial–temporal dependency in travel demand. Two mainstream architectures, Convolutional Neural Network (CNN) and Recurrent Neural Network (RNN), are often integrated to capture the spatial and temporal information of traffic demand (Ma et al., 2017; Xiangxue et al., 2019; Fu et al., 2016; Zhao et al., 2017; Du et al., 2019b). Typically, CNN exploits regular convolutional kernels scanning through the input features (e.g., images) to extract spatial characteristics of travel demand (LeCun et al., 2015). RNN leverages the extracted temporal dynamic behavior from the past elements of the sequence to predict the next element (LeCun et al., 2015; Hochreiter and Schmidhuber, 1997). To better capture spatial–temporal information, several hybrid deep learning frameworks incorporating both CNN and RNN architectures are developed and these models achieved good performance in various traffic prediction tasks (Ai et al., 2019; Li et al., 2021; Zhang et al., 2020b; Ren et al., 2020).

However, CNN has certain shortcomings when it is employed to capture spatial–temporal information of bike sharing demand. CNN achieves desirable performance in object detection for images, because adjacent pixels of the same object are often highly correlated. Unlike images, bicycle usage in neighboring urban areas can be quite different due to spatial variations of travel behavior and built environmental characteristics (Du et al., 2019a; Xu et al., 2019). On the other hand, for certain areas that are far apart, the bicycle usage patterns may exhibit similar temporal rhythms. Given the regular shape of convolutional kernels, deep learning models with typical CNN architecture are not able to capture the similarities of bike usage patterns among distant urban areas. If such similarities can be captured and incorporated into the prediction models, it may further enhance the accuracy and reliability of bike sharing demand forecast.

To bridge the research gap, this paper introduces an irregular convolutional Long Short-Term Memory model (IrConv+LSTM) to improve short-term demand forecast for urban bike-sharing systems. The model employs irregular convolutional architecture to capture the relationship of bicycle usage among distant urban areas. Given the areas being forecast, an irregular convolution operation is performed over their *semantic neighbors*, which refer to places that show similar temporal bicycle usage patterns. Two measures, namely Pearson Correlation Coefficient (IrConv+LSTM:P) and Dynamic Time Warping (IrConv+LSTM:D), are used as similarity metrics to identify semantic neighbors for the areas being forecast. The two variants of the proposed model (IrConv+LSTM:P and IrConv+LSTM:D) and several benchmark models are evaluated and compared over bike-sharing systems in five cities, including one dockless bike-sharing system in Singapore and four station-based systems in Washington D.C., Chicago, New York, and London, respectively.

The remainder of this paper is organized as follows. Section 2 provides a review of relevant literature. Study sites and related terminologies are introduced in Section 3. Sections 4 and 5 illustrate the methodologies and experimental results. In Section 6, we conclude the study and discuss possible future research directions.

2. Literature review

Many studies have been conducted on traffic demand prediction using either parametric or non-parametric models (Nagy and Simon, 2018). In parametric models, data series is modeled as a dynamic variation from a systemic basis. Most parametric models adopt filters to estimate parameters that capture the system characteristics for predicting future status. Autoregressive Integrated Moving Average (ARIMA) model and Kalman Filter are two typical parametric models (Xu et al., 2017). ARIMA model uses Autoregressive (AR) or Moving Average(MA) methods to simulate the temporal autocorrelation of a smoothed sequence. The applications of ARIMA in traffic prediction include forecasting traffic accidents, traffic status from the perspective of speed, volume and travel time (Kumar and Jain, 1999; Avuglah et al., 2014; Billings and Yang, 2006; Lee and Fambro, 1999). Kalman Filtering is an algorithm for optimal system state estimation by using status equations of linear systems. Similar to ARIMA, Kalman Filtering not only predicts traffic status but also assists traffic management and controls (Antoniou et al., 2007; Szeto and Gazis, 1972; van Lint and Djukic, 2012). Such parametric models filter a wealth of information based on the strong assumptions of the data. Also, spatial information is not considered in such models. Therefore, the parametric models cannot fully model the spatial–temporal characteristics from the historical traffic data.

With the advancement of computing power, machine learning and deep learning models have been increasingly adopted for travel demand forecast. These models include Support Vector Machine (SVM), Random Forest, Bayesian Network, Markov Model, Neural Network, and hybrid deep learning models (Zhang and Liu, 2009; Hou et al., 2017; Sun et al., 2006, 2005). In particular, deep learning models have been attracting much attention in the last decade. Many models are applied to predict the traffic status, such as Convolutional Neural Network (CNN), Recurrent Neural Network (RNN) with its variants, encoder–decoder and attention mechanisms for sequential prediction, and Graph Convolutional Network (GCN) for graphic knowledge learning (Ma et al., 2017; Zhao et al., 2017; Wang et al., 2020; Zi et al., 2021; Liu et al., 2018; Zhang et al., 2018). For example, Ma et al. establishes a space–time matrix representing traffic sensors ordered by road directions and time, then adopts CNN to capture spatial and temporal information for predicting (Ma et al., 2017). In Fu et al. (2016), Zhao et al. (2017) and Shin et al. (2020), RNN and LSTM are employed to predict the traffic status, including traffic speed and congestion. To incorporate spatial and temporal characteristics, many scholars also propose several hybrid models to forecast the traffic demands or flows. Zhang et al. employs three residual

Table 1
Description of research areas.

Para.	City				
	Singapore	Chicago	Washington D.C.	New York	London
Spatial resolution			1 Kilometer		
Temporal resolution			1 Hour		
Training period	16/06–02/08,2017	01/06–30/09,2019	01/06–30/09,2019	01/06–30/09,2019	01/06–30/09,2019
Validation period	03/08–31/08,2017	01/10–25/10,2019	01/10–31/10,2019	01/10–25/10,2019	01/10–31/10,2019

CNNs to capture spatial–temporal information of historical pedestrian flows for predicting citywide crowd flows (Zhang et al., 2017). In Zhang et al. (2020a), CNN and attention LSTM are adopted to forecast passenger flow in urban rail transit. Ren et al. adopt residual CNN and LSTM blocks for extracting high-level spatial–temporal information of pedestrian flow volumes to forecast citywide pedestrian volumes (Ren et al., 2020). In general, deep learning models has achieved better performance in predicting traffic status than the parametric models.

Deep learning models are also widely used in short-term bike-sharing demand prediction. In general, CNN and GCN are commonly used deep learning models to capture the relationship of bicycle usage in prediction tasks with different data aggregation strategies (grid-based and graph-based), respectively (Jiang, 2022). For studies with CNN, the study areas are firstly divided into regular grid maps based on specific spatial resolutions. Then, CNN is adopted to capture the spatial correlation of bicycle usage from spatially adjacent cells (urban areas) (Li et al., 2021; Miao et al., 2022; Tian et al., 2021). Unlike CNN, GCN captures the usage relationship in topological space (spectral domain). The topological space is defined based on the relationship between nodes (stations or station clusters), such as spatial distance or usage correlation matrix (Yang et al., 2020; Lin et al., 2018; Chai et al., 2018). Generally, the prediction models with CNN can be used to predict bicycle usage across dockless and station-based systems (Li et al., 2021; Ai et al., 2019; Tian et al., 2021). However, the models with GCN are commonly used to forecast the usage in station-based systems (Yang et al., 2020; Lin et al., 2018; Chai et al., 2018; Li et al., 2019). The topological structure of dockless systems is denser than that of station-based systems, leading to time and space complexity increment for model training. Therefore, the prediction models with CNN are more appropriate for predicting bicycle usage over different bike-sharing systems.

3. Study areas and data preprocessing

In this study, we use bike-sharing datasets in five different cities to assess the performance of our proposed model. These datasets include one dockless bike-sharing system (DBSS) in Singapore, and four station-based systems (SBSS) in Chicago, Washington D.C., New York, and London, respectively. The datasets of SBSS are collected from the bicycle trip records that document the departure and arrival time and stations. The dataset of DBSS in Singapore is collected from raw GPS coordinates of starting and ending locations of a trip. The GPS trajectories are preprocessed to remove GPS drifts and outliers using the approach from a prior study in Singapore (Xu et al., 2019).

As shown in Fig. 1, we use regular grids to summarize the usage pattern of shared bicycles in the five cities for both DBSS and SBSS. A city area is divided into a regular $w * h$ grid map under a specific spatial resolution. During a time interval, the grid values represent the number of bicycle pick-ups distributed in the city. As shown in Table 1, we set 1 km as the spatial resolution and 1 h as the temporal resolution for this study. At the kth time interval, the definition of a grid map ($X_k(w, h)$) is shown in Eq. (1). $x_k(i, j)$ denotes the number of pick-ups in the cell located in the ith row and the jth column at the kth time interval, defined in Eq. (2).

$$X_k(w, h) = \begin{bmatrix} x_k(1,1) & x_k(1,2) & \cdots & x_k(1, h-1) & x_k(1, h) \\ x_k(2,1) & x_k(2,2) & \cdots & x_k(2, h-1) & x_k(2, h) \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ x_k(w,1) & x_k(w,2) & \cdots & x_k(w, h-1) & x_k(w, h) \end{bmatrix} \tag{1}$$

$$x_k(i, j) = | \{ T \in \mathbb{T}_k \mid T(O) \in (i, j) \wedge T(D) \notin (i, j) \} | \tag{2}$$

Here $\mathbb{T}_k$ denotes the trajectories occurred in kth time slot; $T(O)$ and $T(D)$ represent the departure and arrival locations of a trajectory T, respectively; $T(O) \in (i, j) \wedge T(D) \notin (i, j)$ denotes a trajectory starts from the cell (i, j) but ends in another cell except the cell (i, j); $|\cdot|$ denotes the cardinality of a set. In this study, trips that start and end in the same cell are excluded because they do not affect the overall balance of bicycle supply and demand within a cell.

4. Methodology

4.1. Overall architecture of the proposed model

Fig. 2 illustrates the overall architecture of the proposed model. The model consists of three separate modules with the same structure. Each module takes a specific set of historical observations (bike-sharing demand) as input. Instead of using all the historical observations to train the model, we identify key periods with different levels of recency to the target period (for which the prediction is made) and feed them into the three modules. The approach effectively reduces the time complexity of the training model by also
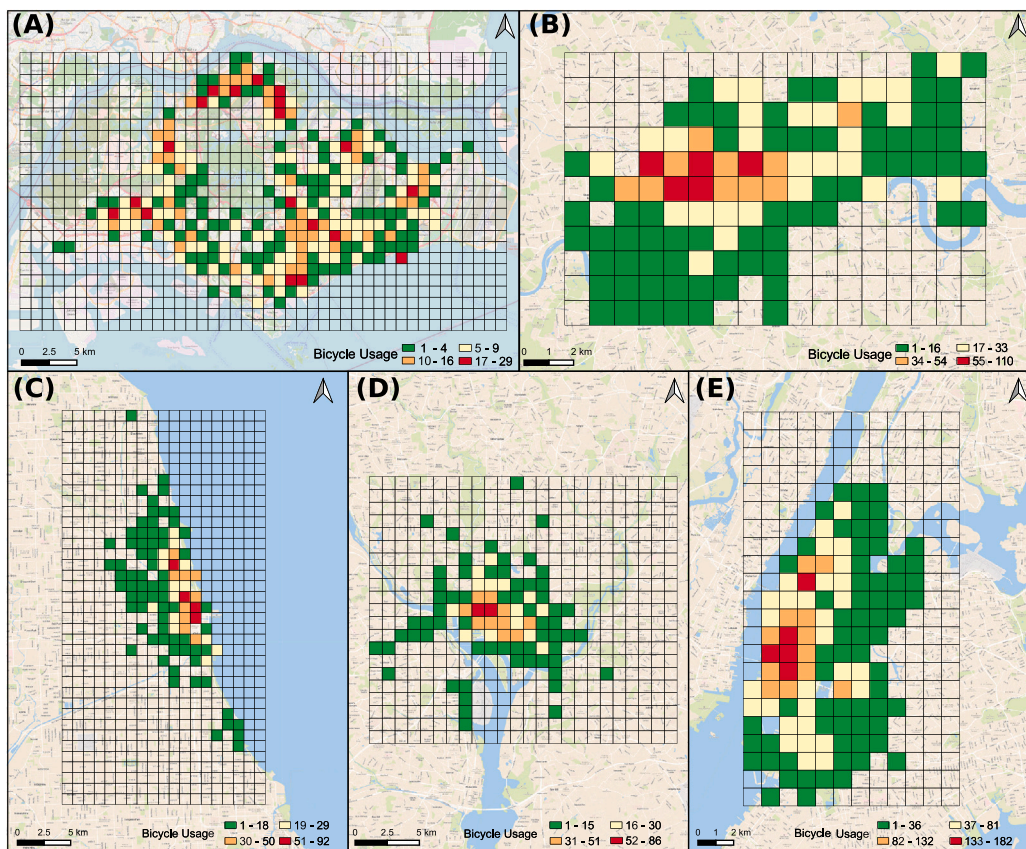
Fig. 1. Spatial distribution of bicycle usage in five cities. **(A)** Singapore during 5–6 PM on August 25th, 2017; **(B)** London during 5–6 PM on August 24th, 2019; **(C)** Chicago during 1–2 PM on August 25th, 2019; **(D)** Washington, D.C. during 5–6 PM on August 25th, 2019; **(E)** New York during 5–6 PM on August 25th, 2019.

mitigating the negative effect of redundant information in historical data. This practice has proved to achieve better performance than models trained using full historical observations (Zhang et al., 2017; Ren et al., 2020; Li et al., 2021; Wu et al., 2018). The definitions of the three key periods (*trend*, *period* and *closeness*) will be elaborated in Section 4.3.

As shown in Fig. 2, each module adopts three layers of irregular convolutional architecture to capture the characteristics of bicycle demand among urban areas. The vector sequence formed by flattening the output of the irregular convolution is used as the input to the LSTM model[1] to extract the temporal information in the sequence. The outputs of three hybrid modules are fed into a feature fusion layer. The output of the feature fusion layer is activated by a non-linear function generating the predicted value. The predicted value with its corresponding actual usage value participates loss estimation and backpropagation to update parameters in the model. In the following section, we formally introduce the architecture of irregular convolutional network.

4.2. Irregular convolutional neural network

The major difference between irregular convolution and traditional convolution lies in the cells involved in the convolutional operation. Generally, the number of cells involved in the convolution is known as the convolutional kernel size. Among cells corresponding to each kernel, the cell being forecast is called the central cell, and the other cells involved in convolution are called neighbors. Taking the convolutional kernel size of nine as an example, the neighbors involved in traditional convolution are spatially adjacent to the central cell, as shown in Fig. 3A. In contrast, the neighbors can be located anywhere in the study area for irregular convolution (Fig. 3B). In this study, for each central cell, we identify the top eight cells which show similar temporal bicycle usage patterns observed from the historical observation data. We call these cells as *semantic neighbors*. Compared to the traditional convolution, irregular convolution is more flexible to exploit cells with similar temporal usage patterns to the central cell.

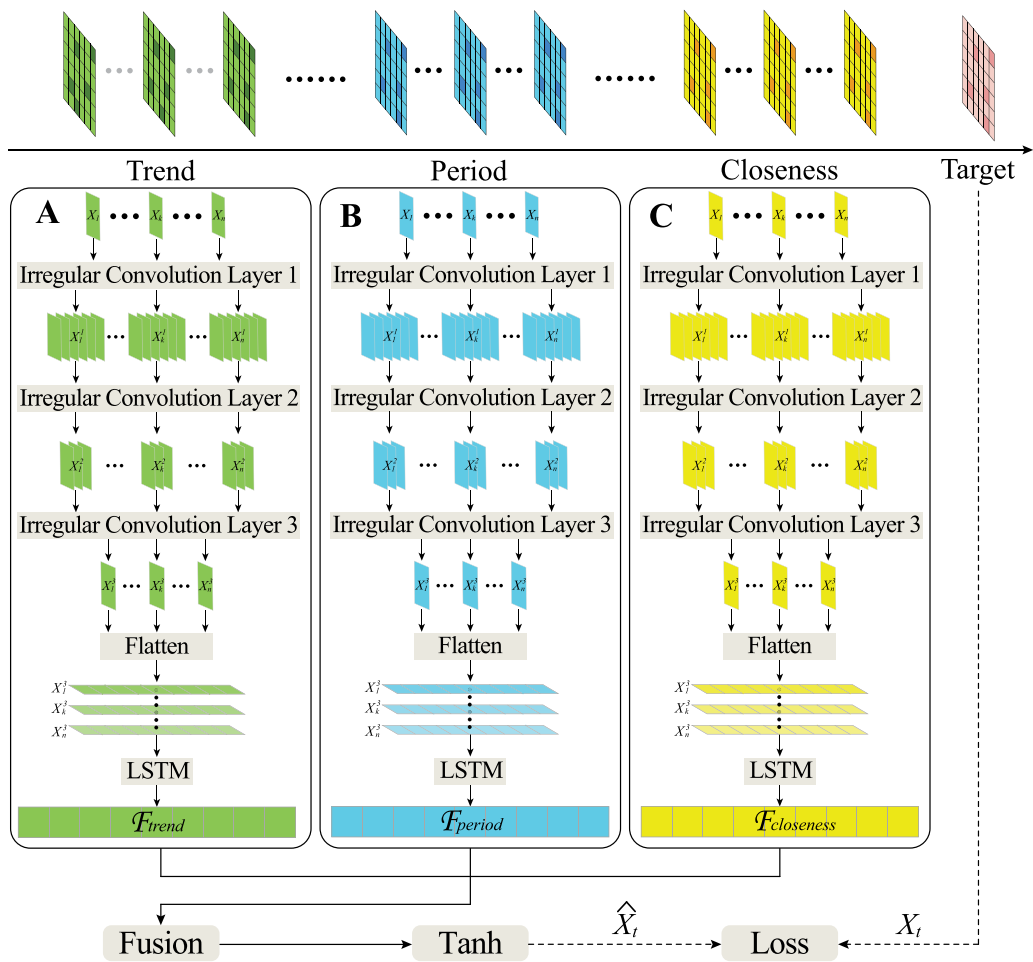[1] The description of the LSTM model is provided in Appendix.

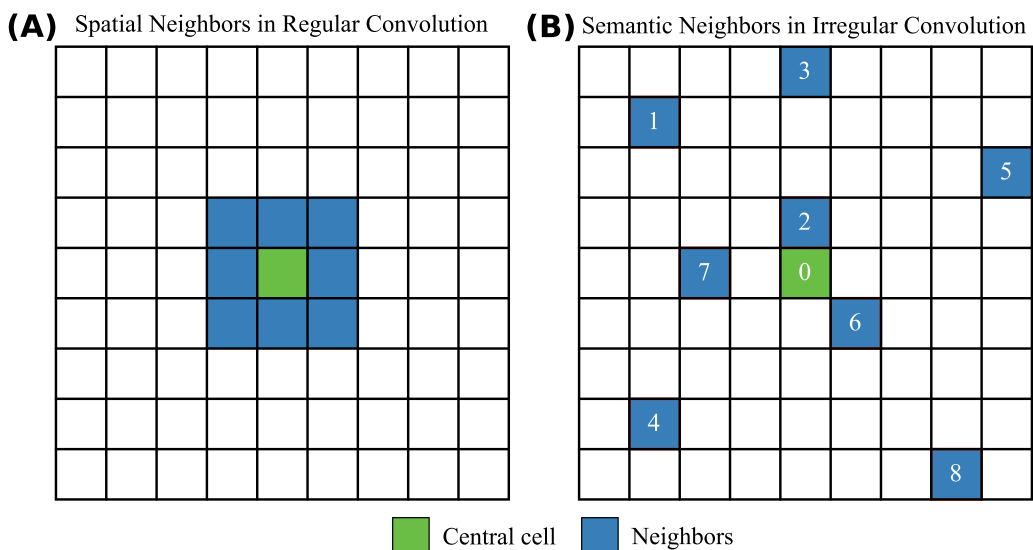**Fig. 2.** Overall architecture of the proposed model.

Fig. 3. Illustration of traditional and irregular convolution with a kernel size of 9: **(A)** spatial neighbors are adjacent to the central cell in the regular convolution; **(B)** in the irregular convolution, semantic neighbors are identified based on the top 8 cells with the highest similarity of temporal bicycle usage patterns to the central cell.
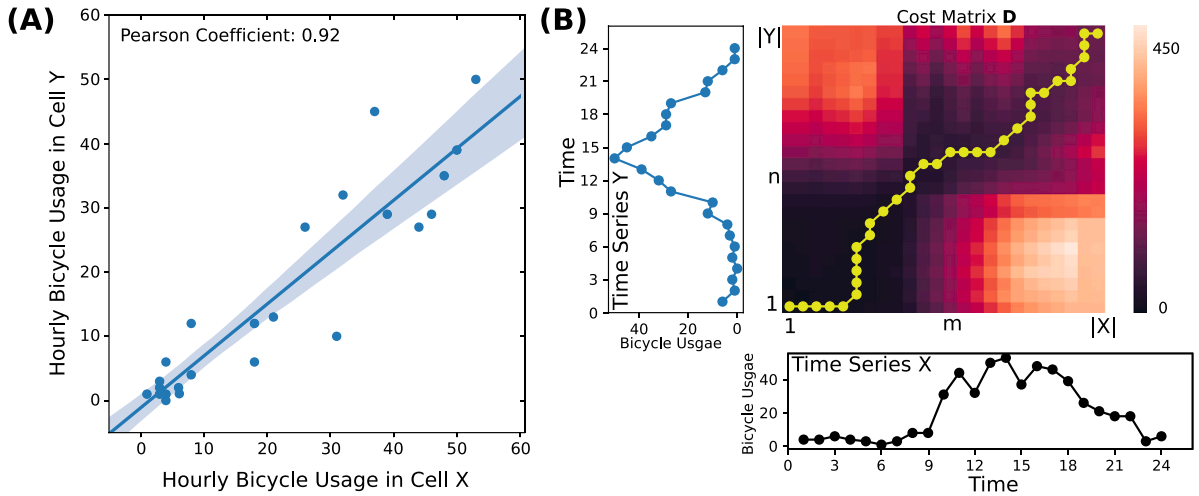
Fig. 4. Two similarity metrics adopted in this study: **(A)** Pearson correlation coefficient; **(B)** Dynamic time warping (DTW).

This study uses two metrics, namely, the Pearson Correlation Coefficient (Benesty et al., 2009; Kristoufek, 2014) and Dynamic Time Warping (DTW) (Müller, 2007; Taylor et al., 2015), to quantify the similarity of temporal bicycle usage patterns between the cells. Because these two metrics measure temporal similarity from different perspectives, we aim to assess which metrics tends to result into a better prediction accuracy.

The Pearson correlation coefficient measures the strength of a linear association between two sequences. The coefficient is the ratio between covariance of two variables and the product of their standard deviations. The two sequences are more positively correlated when the ratio is closer to 1. Eq. (3) shows the calculation of the Pearson correlation coefficient. Fig. 4A shows an example of the Pearson correlation coefficient of two bicycle usages sequences within 24 h.

$$r_{xy} = \frac{\sum_{i=1}^{n} (X_i - \overline{X})(Y_i - \overline{Y})}{\sqrt{\sum_{i=1}^{n} (X_i - \overline{X})^2 \sum_{i=1}^{n} (Y_i - \overline{Y})^2}} \tag{3}$$

where $\overline{X}$ and $\overline{Y}$ denote the means of sequence X and Y, respectively. n denotes the length of sequence.

For DTW, given two bicycle usage sequence $X = x_1, x_2, \ldots, x_i, \ldots, x_{|X|}$ and $Y = y_1, y_2, \ldots, y_j, \ldots, y_{|Y|}$, the optimization objective is to find a shortest warp distance between two sequences $dist(W)$ (typically Euclidean distance), as shown from Eq. (4) to Eq. (6):

$$dist(W) = minimum(\sum_{k=1}^{k=K} distance(w_{ki}, w_{kj})) \tag{4}$$

$subject\ to:$

$$W = w_1, w_2, \ldots, w_K, \ max(|X|, |Y|) \leq K < |X| + |Y| \tag{5}$$

$$w_k = (i, j), w_{k+1} = (i', j'), \ i \leq i' \leq i + 1, j \leq j' \leq j + 1 \tag{6}$$

where K is the length of warp path, w_k denotes the kth element of the warp path, i and j represent the index of a certain record in two sequences X and Y, respectively. $distance(w_{ki}, w_{kj})$ is the distance between two records (the former from X and the latter from Y) in the kth element of the warp path. Thus, a cost matrix D is constructed based on the distance from any record in one of sequences (X/Y) to any record in another sequence (Y/X). DTW adopts a greedy search from $D(|X|, |Y|)$ to $D(1, 1)$ to find the minimum distance between two sequences. In this study, we use historical observations of each cell during the whole training period to estimate the similarity. The length of the evaluated sequences are the same for two mentioned metrics ($n = |X| = |Y|$). Fig. 4B shows an example of finding the shortest-distance warp path based on the cost matrix D for two bicycle usage sequences.

The irregular convolutional computation is analogous to the traditional convolution:

$$y(i, j) = b(i, j) + \sum_{c=1}^{C_{in}} \sum_{s=1}^{S} x_c^s(i, j) w_c^s \tag{7}$$

Here C_{in} denotes the number of channels in input $x(i, j)$; S denotes the size of convolutional kernel; $x_c^s(i, j)$ represents the semantic neighbor s associated with central cell $x(i, j)$ in the channel c; w_c^s denotes the weight in the convolutional kernel corresponding to the semantic neighbor $x_c^s(i, j)$; $b(i, j)$ denotes the learnable bias. Technically, the implementation of irregular convolution is a weighted summation, as shown in Eq. (7). To identify the elements in each convolution computation, one of the necessary procedures is to reconstruct a tensor that assigns semantic neighbors to their corresponding central cells through a look-up table. This look-up table

is built based on the similarity metric of bicycle usage patterns (Pearson/DTW) and documents the central cell's id with its semantic neighbors' ids. More details about the implementation of the proposed model are illustrated in the code that is open-sourced and available online.[2] We also released a sample dataset (station-based system in New York) used in our study, facilitating readers to better understand the mechanism of the proposed model.

4.3. Definitions of trend, period and closeness

As mentioned in 4.1, given a target period (for which the prediction is made), we identify three key periods from historical observations according to the levels of recency to the target period. We name them as *Closeness*, *Period*, and *Trend* (Fig. 2). The purpose of selecting observations from these periods as inputs is to reduce the negative impact of redundant information in the whole training data on model performance while lowering the training time complexity. For example, the usage patterns of shared bicycles during peak hours may be similar during weekdays, and bicycle usage might be associated with that at the same time in the previous week. This strategy of identifying key periods as training input has proved to achieve better performance than models trained using full historical observations (Zhang et al., 2017; Ren et al., 2020; Li et al., 2021; Wu et al., 2018). In this study, the definitions of such three key periods are given in Eq. (8) to Eq. (10).

$$\mathcal{X}_t^{closeness} = \{X_{t-l_c}, X_{t-(l_c-1)}, \ldots, X_{t-1}\} \tag{8}$$

$$\mathcal{X}_t^{period} = \{X_{t-24*l_p}, X_{t-24*(l_p-1)}, \ldots, X_{t-24}\} \tag{9}$$

$$\mathcal{X}_t^{trend} = \{X_{t-7*24*l_q}, X_{t-7*24*(l_q-1)}, \ldots, X_{t-7*24}\} \tag{10}$$

We select l_c as 24 to capture historical shared bicycle usage data in the past 24 h. The value of l_p is set as 7 to select the usage data at the same time for each day in the past week. l_q is chosen as 2 to provide historical information at the same time in the past week and the week before the past. The time complexity of training a model using the above mentioned key periods will be lower than that of training with a lengthy period of data.

The outputs of three separate modules are fused for the final forecast. We adopt the weighted element-wise addition method to merge the three spatial–temporal features, including $\mathcal{F}_{trend}$, $\mathcal{F}_{period}$ and $\mathcal{F}_{closeness}$. Then, the feature map is activated by a *tanh* function to generate the prediction values $\hat{X}_t$ that participates in the loss and backpropagation with the actual bicycle usage X_t. The computation of the prediction values $\hat{X}_t$ is shown Eq. (11).

$$\hat{X}_t = tanh(\mathcal{F}) = tanh(W_t \circ \mathcal{F}_{trend} + W_p \circ \mathcal{F}_{period} + W_c \circ \mathcal{F}_{closeness}) \tag{11}$$

where W_t, W_p and W_c denote the learnable parametric vectors with the same shape size of the corresponding feature, $\mathcal{F}$ is the feature map after fusion, and *tanh* denotes the activation function.

4.4. Hyperparameter settings and benchmark models

Since we adopt two metrics to quantify the similarity of temporal usage patterns, the performance of two variants of IrConv+LSTM is evaluated in this study. We name them as IrConv+LSTM:P and IrConv+LSTM:D, respectively. The hyperparameters for both variants are the same and refer to several existing studies (Liu et al., 2018; Yang et al., 2019; Zhang et al., 2018). Three irregular convolutional layers in each separate module operation (shown in Fig. 2) are adopted with 32 filters, 16 filters, and 1 filter, respectively. The last filter in the irregular architecture is to aggregate the high-dimensional information into one channel. The convolutional kernel size for two variants of our model is set as nine.

We select five benchmark models that are commonly used to predict bike-sharing demand and can be applied to dockless and station-based systems, including one parametric model (ARIMA) and four deep learning models (LSTM, GRU, STRN, and CNN+LSTM). These five benchmark models are adopted for performance comparison with two IrConv+LSTM variants.

- ARIMA: Auto-Regressive Integrated Moving Average model is a parametric model widely used for time series forecasting. ARIMA is a combination of the differenced autoregressive model (AR) with the moving average model (MA) (Kumar and Jain, 1999). ARIMA can handle non-stationary sequences by replacing the data values with the difference between their values and the previous values to obtain stationary sequences. Thus, ARIMA is widely adopted to predict traffic status that is dynamically changed over time, such as traffic accidents, traffic speed, and traffic volume (Avuglah et al., 2014; Billings and Yang, 2006; Lee and Fambro, 1999).
- LSTM: Long Short-Term Memory is a widely used deep learning architecture in the field of traffic prediction (Zhao et al., 2017; Fu et al., 2016; Shin et al., 2020). As a variant of Recurrent Neural Network (RNN), one of the advantages of LSTM is that it can capture the information for both long and short periods of time by introducing gate theory (with three gates). LSTM is effective for processing long sequences of input data because it tackles the gradient vanishing and explosion problems that can be encountered when training traditional RNN models.

---

[2] https://github.com/joeyleehk/IrConv-LSTM

- GRU: Gated Recurrent Unit, another variant of RNN, is also commonly used deep learning model for traffic prediction (Cho et al., 2014; Fu et al., 2016; Zhang and Kabuka, 2018; Dai et al., 2019). Compared to LSTM, the architecture of GRU is simplified, and the speed of training is faster. In specific, GRU only adopts two gates (update and reset gates) to model the short- and long-term dependency in temporal sequences. Although the architecture of GRU is simpler than LSTM, many studies find that the performance of GRU is similar to or even better than LSTM networks (Wang and Kim, 2018; Chen et al., 2020).
- STRN: Spatial–Temporal Residual Network is a hybrid deep learning prediction model initially proposed to forecast short-term pedestrian volume (Zhang et al., 2017). It consists of three Residual Convolutional Neural Networks (ResNet). The spatial features captured by ResNets from multiple fragments of historical data are fused together for the final prediction. The strategy of adopting fragments of historical data to train deep learning prediction models has been referenced in many studies (Liu et al., 2019; Ren et al., 2020; Li et al., 2021; Du et al., 2019b).
- CNN+LSTM: CNN+LSTM is a hybrid deep learning model for spatial–temporal prediction. This model couples a traditional convolution and an LSTM model to extract historical spatial–temporal information for the prediction. It has been adopted for traffic prediction by several studies (Kim and Cho, 2019; Cao et al., 2020). The difference between this model and our proposed model is that CNN+LSTM adopts traditional convolution involving spatial neighbors.

To make the prediction results of IrConv+LSTM and CNN+LSTM comparable, we set similar hyperparameters and structures for both of them. Specifically, three hybrid modules in CNN+LSTM extract spatial–temporal information from their respective key historical periods for modeling. The definitions of such three key periods in CNN+LSTM is the same as them in our model. Each separate module also contains three traditional convolution layers with 32 filters in the 1st layer, 16 filters in the 2nd layer, 1 filter in the 3rd layer. The convolutional kernel size adopted in CNN+LSTM is also set as nine. The pooling layer is not adopted in either CNN+LSTM or the proposed model since the pooling process will reduce the output size. The hyperparameter settings of STRN are referred to the settings in article (Zhang et al., 2017). The parameters of all benchmarks have been calibrated to achieve optimal prediction results.

We use the first 80% of the hourly usage data as the training data and the last 20% data to validate the performance of models in each city, as shown in Table 1. In addition, the training data for each city are used to evaluate the similarity of temporal bicycle usage patterns among cells. The loss function of all deep learning models in this study is the MSELoss function. Eq. (12) defines the MSELoss function that is used to measure the errors between the prediction value and the ground truth (Wang et al., 2016). The optimization algorithm for updating parameters in deep learning models is important for backpropagation. This study employs RMSProp (Root Mean Square Prop) as the optimization algorithm across all deep learning models (Graves, 2013). Moreover, three indicators are employed to evaluate the performance of models, including Mean Absolute Percentage Error (MAPE), Root Mean Square Error (RMSE), and Mean Absolute Error (MAE) (Zhang et al., 2017; Ren et al., 2020; Zhang et al., 2020b). The definitions of them are shown from Eq. (13) to Eq. (15).

$$MSE = \frac{1}{n} \sum_{i=1}^{n} (\hat{y}_i - y_i)^2 \tag{12}$$

$$MAPE = \frac{100\%}{n} \sum_{i=1}^{n} |\frac{\hat{y}_i - y_i}{y_i}| \tag{13}$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (\hat{y}_i - y_i)^2} \tag{14}$$

$$MAE = \frac{1}{n} \sum_{i=1}^{n} |\hat{y}_i - y_i| \tag{15}$$

Here $\hat{y}_i$ denotes the prediction result and y_i denotes the actual value of bicycle usage. All deep learning frameworks are constructed on the Pytorch platform (Paszke et al., 2019). Also, the models are built on a server with NVIDIA Tesla V100 and a workstation with NVIDIA RTX 2070 Super Graphics Card.

5. Analysis results

5.1. Overall accuracy of the proposed model and benchmark models

In this section, we compare the overall accuracy between our proposed model and the benchmark models. Table 2 shows the overall accuracy of two variants of our proposed model and other benchmark models. Given a specific indicator, the model with the best performance is marked with ∗ in Table 2.

Generally, our proposed model achieves better performance than other baseline models across five cities based on the prediction accuracy. As a typical parametric model, ARIMA is hard to process the non-stationary sequence and cannot leverage any spatial dependency of bike-sharing usage. Thus, it achieves the lowest accuracy of forecasting bicycle demand across five cities. Unlike ARIMA, two variants of the RNN model (LSTM and GRU) are effective in extracting temporal correlation from historical observations in each urban area by using the recurrent process. Hence, the results of LSTM and GRU are better than that of ARIMA. However, these RNN-based deep learning models still cannot utilize the spatial correlation of bicycle usage among cells for prediction. Based on the results, the hybrid deep learning model that couples convolution architectures and LSTM model to extract spatial–temporal features

Table 2
Overall accuracy of all models across five cities.

City	Index	ARIMA	LSTM	GRU	STRN	CNN+LSTM	IrConv-LSTM:P	IrConv-LSTM:D
Singapore	MAPE	0.8488	0.6715	0.6853	0.7026	0.6696	0.5617*	0.5638
	MAE	2.6267	2.2411	2.4920	2.2488	2.0971	1.9727	1.9655*
	RMSE	4.0656	3.4403	3.8716	3.3602	3.2243	3.0911	3.0764*
Chicago	MAPE	1.3000	0.8615	0.7536	0.7131	0.7584	0.7240	0.6028*
	MAE	4.6156	3.6955	3.1127	2.5619	2.8373	2.4036	2.2047*
	RMSE	12.6296	10.1246	7.3725	4.7583	6.5034	4.7802	4.3356*
Washington, D.C.	MAPE	0.9672	0.7133	0.7089	0.6858	0.6492	0.5599*	0.5675
	MAE	3.7649	2.3535	2.4487	2.3327	2.4026	1.9758	1.9697*
	RMSE	7.2917	4.0227	4.1485	4.0657	4.0718	3.3669	3.3590*
New York	MAPE	2.9407	0.7479	0.7162	0.7770	0.7153	0.6243	0.6201*
	MAE	16.9310	7.5721	7.5848	7.2278	6.1815	5.9865	5.7769*
	RMSE	31.7418	15.3383	14.7946	13.7237	11.3954	11.3385	10.8141*
London	MAPE	1.7924	0.7483	0.7815	0.6753	0.6862	0.5852	0.5523*
	MAE	8.7771	4.8125	4.9858	4.2348	4.0423	3.7829	3.5785*
	RMSE	17.8831	9.4463	9.3810	7.1533	7.2279	6.6326	6.2960*

of bicycle usage generally outperforms LSTM/GRU in all cities. Moreover, STRN also performs better than LSTM. Although STRN and CNN+LSTM achieve good prediction results, the performance of our proposed model is still better than them in all indicators. Notably, our proposed model achieves an improvement of MAPE from 8% (in Washington, D.C.) to 12% (in London), compared to the model with the best performance in four benchmarks.

Compared to CNN+LSTM, our proposed model only replaces the spatial neighbors involved in regular convolution with the semantic neighbors adopted in irregular convolution. However, the prediction accuracy of our model is much higher than CNN+LSTM across five cities. Specifically, among five cities, our approach achieves an improvement of MAPE by 8%–15% compared to CNN+LSTM. Such results imply that the semantic neighbors are more effective than spatial neighbors for predicting bike-sharing usage. The semantic neighbors are identified according to the similarity of temporal bicycle usage to their corresponding central cells. Although the semantic neighbors are not always spatially adjacent to their central cells, they can provide essential information for spatial–temporal modeling than the spatial neighbors. In other words, the semantic neighbors selected based on the similarity of bicycle usage patterns are more appropriate than spatial neighbors for short-term bike-sharing demand prediction. The possible reason is the uncertainty in the spatial relationship of bicycle usage between neighboring urban areas over two types of bike-sharing systems (Gong and Yamamoto, 2019; Du et al., 2019a). Some factors shaping users' travel behaviors affect this relationship uncertainty, such as built environment characteristics, urban functions, etc. Thus, this uncertainty can be introduced into the prediction model with regular convolution architecture, affecting the model performance. Instead, the selection of semantic neighbors depends on the similarity (relationship) of bicycle usage patterns among urban areas. Therefore, involving semantic neighbors in the irregular convolution is a good strategy for predicting shared-bicycle usage. We further analyze the differences between semantic and spatial neighbors in Section 5.4, facilitating our understanding of what affects the model performance.

We also find that the performance of the variant with DTW metric is better than the variant with Pearson measure in most cities. The variant with Pearson correlation coefficient only performs slightly better in two cities (Singapore and Washington D.C.) from the perspective of MAPE. Based on the characteristics of Pearson correlation coefficient, the usage variations of two measured sequences are simply reflected by their respective standard deviations, and the temporal offsets between such two sequences cannot be quantified by the coefficient. In contrast, DTW metric is effective to measure the similarity of usage variations between two sequences by calculating the distance of each record in one sequence to all other observations in the other. Moreover, the shortest warping distance between the sequences searched by DTW can quantify their similarity more precisely without the effects of temporal offsets. In sum, our proposed model achieves the best overall performance compared to other baseline models. Moreover, DTW metric is an effective way to select semantic neighbors involved in the irregular convolution.

5.2. Performance of models over cells with varying levels of bicycle usage

When operating bike-sharing systems, one of the important tasks is to satisfy users' travel needs. However, bike-sharing demand is not evenly distributed in urban areas. The number of areas with high user demand is relatively small but most users intend to ride bicycles in such areas. In contrast, areas with low demand require fewer bicycles, but they are widely distributed in the city. Therefore, there is a trade-off when allocating bicycles across urban areas. If the travel demand in areas with different usage can be precisely predicted, it is helpful for the deployment of bicycles.

We separate the cells into five quantiles in each city according to their respective hourly usage to assess the models' performance over these quantiles. Fig. 5 shows the MAE distributions of each quantile of bicycle usage in the five cities. We find that the performance of our model is better than baseline models across five quantiles of bicycle usage in most cities. Particularly, for these high-demand cells, the prediction accuracy of our proposed model is higher than other benchmarks. For example, the average MAE of our approach is less than 17 bicycles in Singapore, lower than that of the benchmarks. Also, the maximum MAE in Chicago is close to 10 bicycles, much smaller than other benchmarks. However, in Washington, D.C., and New York, the average MAE of
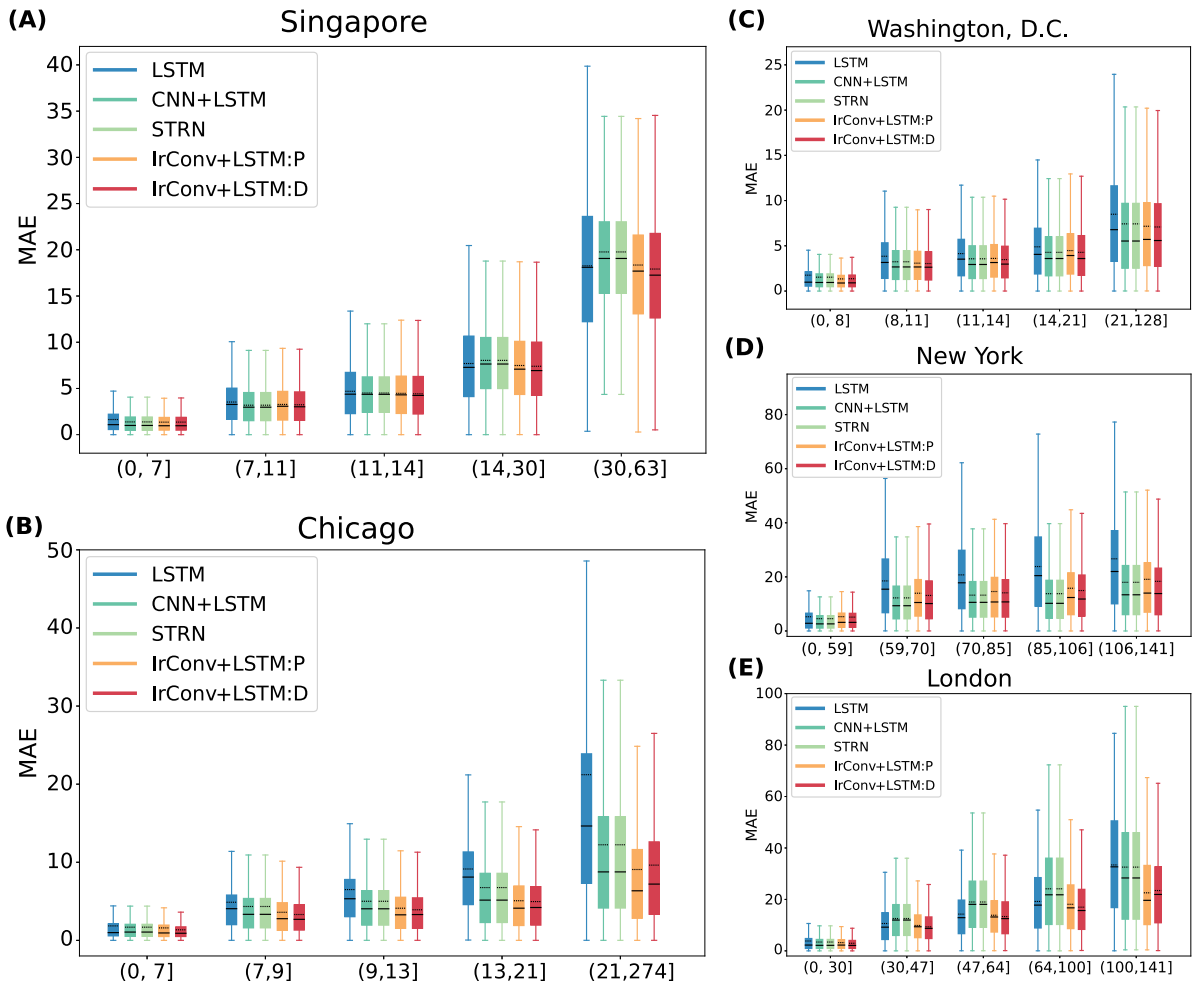
Fig. 5. Performance of models in cells with various levels of bicycle usage. The cells in each city are divided into five quantiles according to their respective hourly usage, in order from low to high usage shown in horizontal axes.

our approach is similar to that of other baseline models. This fact indicates that our approach has a similar prediction ability to the other models in such two cities with large bicycle usage. Additionally, our proposed model achieves better performance on the other quantiles of bicycle usage. Even in low-demand areas in most cities, our proposed model's average MAE is lower than the other baseline models. To sum up, IrConv+LSTM has achieved better performance over areas with varying levels of bicycle usage.

5.3. Performance of models during peak hours

Another important task in operating bike-sharing systems is meeting the users' needs during the morning and evening commuting hours. For example, in some cities, there are many users who prefer to ride bicycles to address first- and/or last-mile problems during the morning and evening peak hours (Xu et al., 2019; Midgley, 2011). The operators need to allocate enough available bicycles for users during such peak hours. Accurate bicycle demand forecasting can help them develop proper scheduling plans to satisfy users with as little cost as possible.

We select the morning peak (7:00–10:00 AM) and evening peak (5:00–8:00 PM) observations from the validation datasets to assess the models' performance in each city. Fig. 6 represents the prediction accuracy of each model during peak hours in five cities based on the two indicators (MAPE and MAE). The prediction error of our proposed model is the lowest across five cities during both peak hours. Specifically, in London, the MAE of IrConv+LSTM:D is smaller than five during the evening period, better than the benchmarks. In Chicago, the MAPE of our approach is close to 50% during both peak periods, which is smaller than other baseline models. Although IrConv+LSTM performs similarly to CNN+LSTM in Singapore during both peak hours, the performance of IrConv+LSTM is stable and better across all five cities. Comparing the two variants of our proposed model, the performance of the variant adopted DTW metric is more robust in most cities than the variant with the Pearson measure. To sum up, our model outperforms the benchmarks during peak hours in five cities.
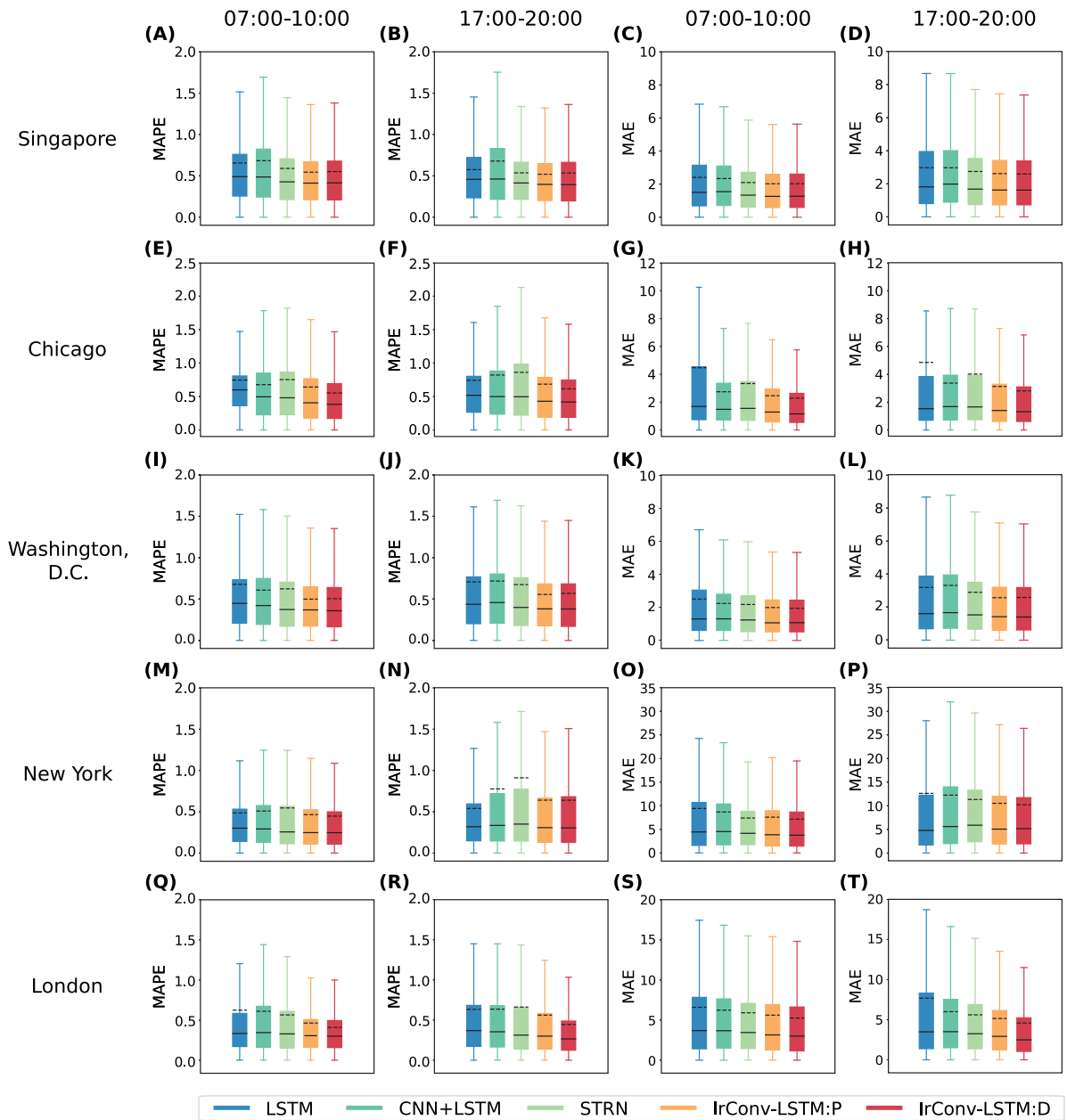
Fig. 6. Performance of models during peak hours: **(A–D)** Singapore; **(E–H)** Chicago; **(I–L)** Washington D.C.; **(M–P)** New York; and **(Q–T)** London.

5.4. Comparative analysis between semantic and spatial neighbors

As mentioned in Section 5.1, our approach involving semantic neighbors in irregular convolution has a notable improvement in model performance compared to the benchmark (CNN+LSTM) incorporating spatial neighbors in traditional convolution. Here we perform an additional analysis to gain insights into the spatial relationship between cells' semantic and spatial neighbors.

First, we analyze the difference in temporal similarity between the central cell and spatial neighbors vs. the central cell and semantic neighbors. As shown in Fig. 7, the similarity of temporal usage patterns between central cells and their semantic neighbors is generally higher than that between central cells and their spatial neighbors. Combined with the prediction accuracy of our model and CNN+LSTM, the similarity of semantic neighbors is more effective in enhancing the accuracy of deep learning models. However, the traditional convolution fails to leverage such similarity due to its structural limitations. Moreover, the similarity differences between the two types of neighbors imply that directly applying traditional convolution involving spatial neighbors for shared bicycle
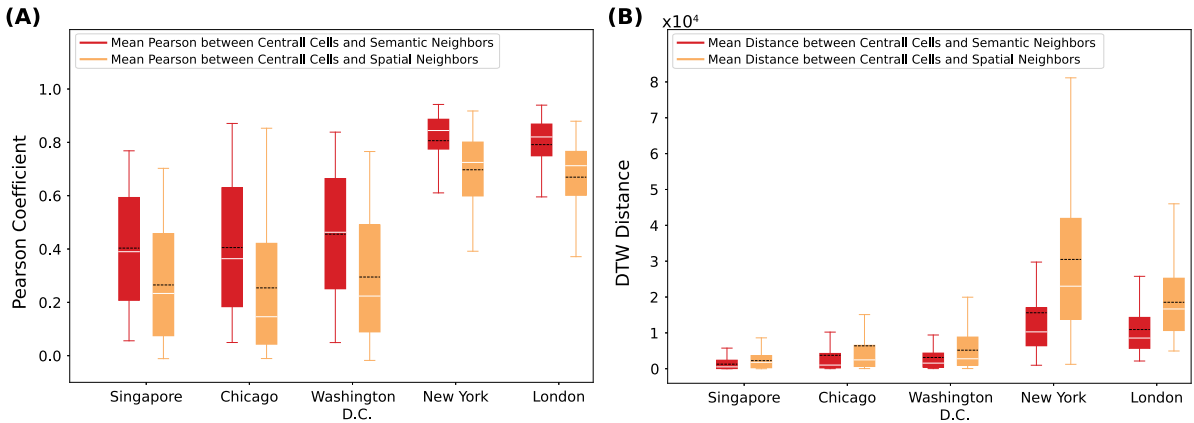
Fig. 7. **(A)** Distribution of average Pearson correlation coefficients between the central cell and spatial neighbors vs. semantic neighbors; **(B)** Distribution of average DTW distance between the central cell and spatial neighbors vs. semantic neighbors.
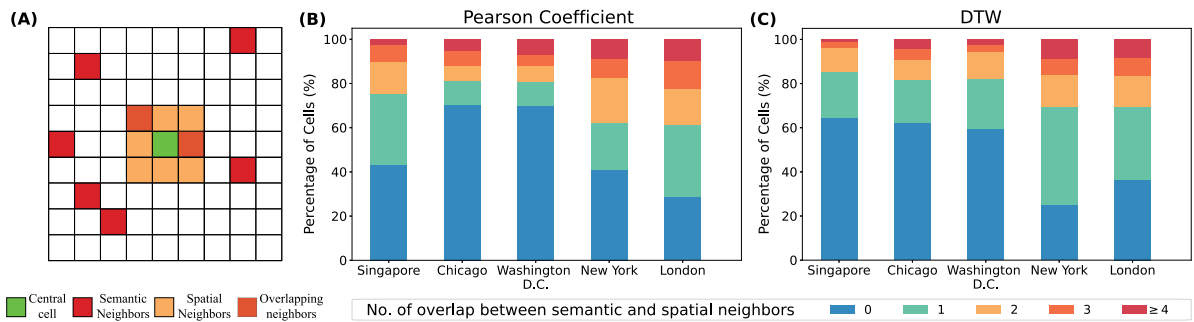

Fig. 8. Number of overlapping cells between a central cell's spatial and semantic neighbors: **(A)** an example representing overlapping cells between spatial and semantic neighbors; **(B)** results based on Pearson correlation coefficient; **(C)** results based on Dynamic Time Warping (DTW).

usage prediction might not always achieve desirable performance. In the five study sites evaluated in this research, CNN+LSTM based on regular convolution architecture performs worse than our proposed model.

From a spatial point of view, we evaluate the extent to which semantic neighbors overlap with spatial neighbors. Fig. 8 shows the distributions of overlap cells between spatial and semantic neighbors in five cities. We find that more than 70% of central cells have their spatial and semantic neighbors sharing less than two cells, which can be observed under both Pearson and DTW metrics. This fact indicates that the semantic neighbors are mainly distributed at non-spatially adjacent areas to the central cells, further illustrating the difference between semantic and spatial neighbors. In other words, some areas in a city, though far apart, could show relatively similar temporal bicycle usage patterns than spatially adjacent areas. This is potentially attributed to built environment characteristics, urban functions, and other factors that shape users' travel behavior across urban areas.

6. Discussion and conclusion

This paper proposes a hybrid deep learning model coupling irregular convolution architecture and LSTM modules to predict bicycle usage demand across one dockless bike-sharing system in Singapore and four station-based systems in Chicago, Washington D.C., New York, and London. The irregular convolution is applied over semantic neighbors that refer to places with temporal usage patterns similar to those of the areas being forecast. To measure the cells' similarity and build the semantic neighbors, one variant of our model uses the Pearson correlation coefficient while the other adopts Dynamic Time Warping (DTW). Our proposed model and four benchmark models are evaluated in five cities. Based on the prediction results, the proposed model generally outperforms the benchmarks across five cities. The prediction accuracy of the proposed model is also higher than that of the benchmarks in areas with different bicycle usage levels and during peak hours. Comparing the two variants of our model, the performance of the variant with the DTW metric is better than that of the variant with the Pearson coefficient metric in most of the cities.

We find that the semantic neighbors adopted in the irregular convolution are quite different from the spatial neighbors involved in the traditional convolution. Specifically, the similarity of temporal usage patterns between the central cells and their semantic neighbors is generally higher than that between the central cells and their spatial neighbors under both the Pearson and DTW metrics. Unlike the spatial neighbors, the semantic neighbors are mainly distributed in areas that are not spatially adjacent to their central cells. The model comparison suggests that relating areas that share similar temporal usage patterns through irregular convolution

is helpful to improving the prediction accuracy. The findings also indicate that the neighbors involved in the convolution and the metrics for quantifying similarity (e.g., Pearson correlation vs. DTW) among urban areas tend to influence the performance of the prediction model.

The study suggests that "thinking beyond spatial neighbors" can inspire new solutions that improve short-term travel demand prediction. The implications go beyond applications of bike-sharing systems. In general, a reliable travel demand or traffic prediction model requires a good understanding on the spatial–temporal dependency of historical travel demand. Although the subjects being studied could vary (e.g., vehicles, cyclist, pedestrians and goods), it is reasonable to assume that there exists spatial autocorrelation in travel patterns. This partially explains why deep learning architecture with typical CNN (i.e., by capturing spatial dependency) is helpful to improving travel demand prediction in different transportation applications. However, given that travel patterns are affected by spatial variations of built environment characteristics, urban functions and socioeconomic characteristics, areas close to each other might show different travel dynamics that are sometimes not correlated over time. Therefore, the typical CNN architecture can sometimes introduce "noise" into the prediction process. The irregular convolution with semantic neighbors is one example of many possible strategies to overcome this limitation.

The findings of this study have broad implications for other types of deep learning prediction models. As described earlier, leveraging the hidden linkage among urban areas exhibiting similar bicycle usage patterns facilitates the prediction model performance. This strategy also can be applied to other types of deep learning prediction models. For example, the similarity of bicycle usage patterns could be involved in the definition of adjacency matrix when adopting GCN to capture the relationship of bicycle usage for bike-sharing demand prediction. As another novel deep learning technique, attention mechanism is widely adopted for bike-sharing demand prediction (Zi et al., 2021; Ma et al., 2022; Lee and Ku, 2022). This mechanism facilitates the model selectively concentrating on a few input data/features that are important for the prediction. The similarity of bicycle usage patterns could be incorporated into the attention mechanism, allowing the prediction model to better identify which data/features can further improve the model's performance. Therefore, we believe that the strategy introduced in this paper can also be applied in the context of other novel deep learning models for achieving better performance.

In this study, the semantic neighbors are identified solely based on the temporal similarity in historical bicycle usage patterns. These areas with similar temporal travel patterns could signify potential similarities in built environment characteristics and other factors that shape cycling behaviors. However, these environmental and socioeconomic characteristics are not directly utilized in our proposed model. A possible direction for future research is to consider both travel patterns and environmental factors for defining similarities and identifying semantic neighbors. Combining these static (e.g., built environment characteristics) and dynamic features (i.e., bicycle usage patterns) may further improve the robustness of the prediction model.

CRediT authorship contribution statement

Xinyu Li: Formal analysis, Methodology, Visualization, Writing – original draft; Writing – review & editing. **Yang Xu:** Conceptualization, Funding acquisition, Methodology, Investigation, Writing – original draft, Writing – review & editing. **Xiaohu Zhang:** Data curation, Investigation, Writing – review & editing. **Wenzhong Shi:** Methodology, Writing – review & editing. **Yang Yue:** Investigation, Writing – review & editing. **Qingquan Li:** Investigation, Writing – review & editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

Appendix. The structure of long short-term memory network

We adopt the Long Short-Term Memory (LSTM) model to extract temporal information from the features captured by the irregular convolutional network (Xiangxue et al., 2019; Fu et al., 2016; Du et al., 2019b; Hochreiter and Schmidhuber, 1997). The structure of LSTM is shown in Fig. A.9. As a variant of Recurrent Neural Network, LSTM adopts the gate theory to control the information captured from both long and short periods of time for avoiding gradient descent or explosion. The computations of LSTM are shown from Eq. (A.1) to Eq. (A.6).

$$f_t = \sigma(W_{if}x_t + b_{if} + W_{hf}h_{t-1} + b_{hf}) \tag{A.1}$$

$$i_t = \sigma(W_{ii}x_t + b_{ii} + W_{hi}h_{t-1} + b_{hi}) \tag{A.2}$$

$$\widetilde{C}_t = tanh(W_{ic}x_t + b_{ic} + W_{hc}h_{t-1} + b_{hc}) \tag{A.3}$$

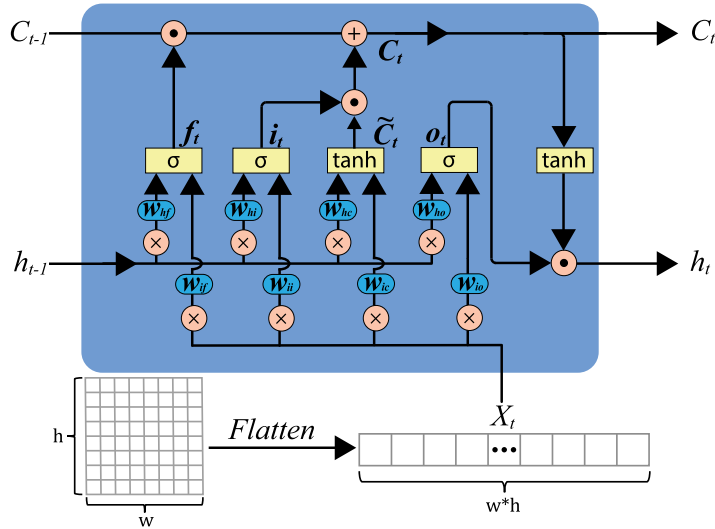**Fig. A.9.** The structure of Long Short-Term Memory Network.

$$o_t = \sigma(W_{io}x_t + b_{io} + W_{ho}h_{t-1} + b_{ho}) \tag{A.4}$$

$$C_t = f_t \odot C_{t-1} + i_t \odot \widetilde{C}_t \tag{A.5}$$

$$h_t = o_t \odot tanh(C_t) \tag{A.6}$$

Here f_t, i_t and o_t are the forget, input and output gates, respectively. W_* and b_* are respectively learnable weights and bias. $\odot$ is the Hadamard product, and σ denotes the sigmoid function. Particularly, the input x_t is the feature captured by irregular convolutional network at time t; C_t denotes the cell state at time t aggregating the information from forget, input gates, cell state and hidden state of the previous layer; h_t represents the hidden state that incorporates the information from C_t and the output gate. According to the equations of LSTM, the hidden state is essential for controlling short-term temporal information, while the cell state stores the long-term temporal information. Therefore, LSTM achieves good performance in prediction by capturing short- and long-term temporal information. Besides, LSTM only accepts vectors as the input in each time interval. As a result, the information captured by the irregular convolution are flattened as a vector then input to LSTM model (shown as Fig. A.9).

References

Ai, Y., Li, Z., Gan, M., Zhang, Y., Yu, D., Chen, W., Ju, Y., 2019. A deep learning approach on short-term spatiotemporal distribution forecasting of dockless bike-sharing system. Neural Comput. Appl. 31 (5), 1665–1677.
Antoniou, C., Ben-Akiva, M., Koutsopoulos, H.N., 2007. Nonlinear Kalman filtering algorithms for on-line calibration of dynamic traffic assignment models. IEEE Trans. Intell. Transp. Syst. 8 (4), 661–670.
Avuglah, R., Adu-Poku, K., Harris, E., 2014. Application of ARIMA models to road traffic accident cases in Ghana. Int. J. Stat. Appl. 4 (5), 233–239.
Benesty, J., Chen, J., Huang, Y., Cohen, I., 2009. Pearson correlation coefficient. In: Noise Reduction in Speech Processing. Springer, pp. 1–4.
Billings, D., Yang, J.S., 2006. Application of the ARIMA models to urban roadway travel time prediction-a case study. In: 2006 IEEE International Conference on Systems, Man and Cybernetics, Vol. 3. IEEE, pp. 2529–2534.
Cao, M., Li, V.O., Chan, V.W., 2020. A CNN-LSTM model for traffic speed prediction. In: 2020 IEEE 91st Vehicular Technology Conference. VTC2020-Spring, IEEE, pp. 1–5.
Chai, D., Wang, L., Yang, Q., 2018. Bike flow prediction with multi-graph convolutional networks. In: Proceedings of the 26th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems. pp. 397–400.
Chen, P.-C., Hsieh, H.-Y., Su, K.-W., Sigalingging, X.K., Chen, Y.-R., Leu, J.-S., 2020. Predicting station level demand in a bike-sharing system using recurrent neural networks. IET Intell. Transp. Syst. 14 (6), 554–561.
Cho, K., Van Merriënboer, B., Gulcehre, C., Bahdanau, D., Bougares, F., Schwenk, H., Bengio, Y., 2014. Learning phrase representations using RNN encoder-decoder for statistical machine translation. arXiv preprint arXiv:1406.1078.
Dai, G., Ma, C., Xu, X., 2019. Short-term traffic flow prediction method for urban road sections based on space–time analysis and GRU. IEEE Access 7, 143025–143035.
Dell'Amico, M., Hadjicostantinou, E., Iori, M., Novellani, S., 2014. The bike sharing rebalancing problem: Mathematical formulations and benchmark instances. Omega 45, 7–19.
Dell'Amico, M., Iori, M., Novellani, S., Subramanian, A., 2018. The bike sharing rebalancing problem with stochastic demands. Transp. Res. B 118, 362–380.
Du, Y., Deng, F., Liao, F., 2019a. A model framework for discovering the spatio-temporal usage patterns of public free-floating bike-sharing system. Transp. Res. C 103, 39–55.

Du, B., Peng, H., Wang, S., Bhuiyan, M.Z.A., Wang, L., Gong, Q., Liu, L., Li, J., 2019b. Deep irregular convolutional residual LSTM for urban traffic passenger flows prediction. IEEE Trans. Intell. Transp. Syst. 21 (3), 972–985.

Fu, R., Zhang, Z., Li, L., 2016. 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, pp. 324–328.

Gammelli, D., Wang, Y., Prak, D., Rodrigues, F., Minner, S., Pereira, F.C., 2022. Predictive and prescriptive performance of bike-sharing demand forecasts for inventory management. Transp. Res. C 138, 103571.

Gong, L., Yamamoto, T., 2019. Temporal and spatial pattern of shared bike trips—An empirical study of New York City. J. East. Asia Soc. Transp. Stud. 13, 1333–1347.

Graves, A., 2013. Generating sequences with recurrent neural networks. arXiv preprint arXiv:1308.0850.

Haghshenas, H., Vaziri, M., 2012. Urban sustainable transportation indicators for global comparison. Ecol. Indic. 15 (1), 115–121.

Hochreiter, S., Schmidhuber, J., 1997. Long short-term memory. Neural Comput. 9 (8), 1735–1780.

Hou, Y., Edara, P., Chang, Y., 2017. Road network state estimation using random forest ensemble learning. In: 2017 IEEE 20th International Conference on Intelligent Transportation Systems. ITSC, IEEE, pp. 1–6.

Hu, S., Xiong, C., Liu, Z., Zhang, L., 2021. Examining spatiotemporal changing patterns of bike-sharing usage during COVID-19 pandemic. J. Transp. Geogr. 91, 102997.

Jiang, W., 2022. Bike sharing usage prediction with deep learning: A survey. Neural Comput. Appl. 1–17.

Jobe, J., Griffin, G.P., 2021. Bike share responses to COVID-19. Transp. Res. Interdiscip. Perspect. 10, 100353.

Kim, K., 2021. Impact of CoVid-19 on usage patterns of a bike-sharing system: Case study of Seoul. J. Transport. Eng. A 147 (10), 05021006.

Kim, T.Y., Cho, S.B., 2019. Predicting residential energy consumption using CNN-LSTM neural networks. Energy 182, 72–81.

Kristoufek, L., 2014. Measuring correlations between non-stationary series with DCCA coefficient. Physica A 402, 291–298.

Kumar, K., Jain, V.K., 1999. Autoregressive integrated moving averages (ARIMA) modelling of a traffic noise time series. Appl. Acoust. 58 (3), 283–294.

Larsen, J., 2013. Bike-Sharing Programs Hit the Streets in over 500 Cities Worldwide. Earth Policy Institute, Washington, DC.

LeCun, Y., Bengio, Y., Hinton, G., 2015. Deep learning. Nature 521 (7553), 436–444.

Lee, S., Fambro, D.B., 1999. Application of subset autoregressive integrated moving average model for short-term freeway traffic volume forecasting. Transp. Res. Rec. 1678 (1), 179–188.

Lee, S.-H., Ku, H.-C., 2022. A dual attention-based recurrent neural network for short-term bike sharing usage demand prediction. IEEE Trans. Intell. Transp. Syst..

Li, X., Xu, Y., Chen, Q., Wang, L., Zhang, X., Shi, W., 2021. Short-term forecast of bicycle usage in bike sharing systems: A spatial-temporal memory network. IEEE Trans. Intell. Transp. Syst..

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., Zhu, Z., Kong, D., Xu, M., Zhao, Y., 2019. Learning heterogeneous spatial-temporal representation for bike-sharing demand prediction. In: Proceedings of the AAAI Conference on Artificial Intelligence, Vol. 33, no. 01. pp. 1004–1011.

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. Transp. Res. C 97, 258–276.

van Lint, H., Djukic, T., 2012. Applications of Kalman filtering in traffic management and control. In: New Directions in Informatics, Optimization, Logistics, and Production. informs, pp. 59–91.

Litman, T., Burwell, D., 2006. Issues in sustainable transportation. Int. J. Global Environ. Issues 6 (4), 331–347.

Liu, Y., Liu, Z., Jia, R., 2019. DeepPF: A deep learning based architecture for metro passenger flow prediction. Transp. Res. C 101, 18–34.

Liu, Q., Wang, B., Zhu, Y., 2018. Short-term traffic speed forecasting based on attention convolutional neural network for arterials. Comput.-Aided Civ. Infrastruct. Eng. 33 (11), 999–1016.

Ma, X., Dai, Z., He, Z., Ma, J., Wang, Y., Wang, Y., 2017. Learning traffic as images: A deep convolutional neural network for large-scale transportation network speed prediction. Sensors 17 (4), 818.

Ma, X., Yin, Y., Jin, Y., He, M., Zhu, M., 2022. Short-term prediction of bike-sharing demand using multi-source data: A spatial-temporal graph attentional LSTM approach. Appl. Sci. 12 (3), 1161.

Miao, H., Fei, Y., Wang, S., Wang, F., Wen, D., 2022. Deep learning based origin-destination prediction via contextual information fusion. Multimedia Tools Appl. 81 (9), 12029–12045.

Midgley, P., 2011. Bicycle-sharing schemes: enhancing sustainable mobility in urban areas. pp. 1–12, United Nations, Department of Economic and Social Affairs, 8.

Müller, M., 2007. Dynamic time warping. In: Information Retrieval for Music and Motion. Springer, pp. 69–84.

Nagy, A.M., Simon, V., 2018. Survey on traffic prediction in smart cities. Pervasive Mob. Comput. 50, 148–163.

Pan, Y., Zheng, R.C., Zhang, J., Yao, X., 2019. Predicting bike sharing demand using recurrent neural networks. Procedia Comput. Sci. 147, 562–566.

Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., Chanan, G., Killeen, T., Lin, Z., Gimelshein, N., Antiga, L., Desmaison, A., Kopf, A., Yang, E., DeVito, Z., Raison, M., Tejani, A., Chilamkurthy, S., Steiner, B., Fang, L., Bai, J., Chintala, S., 2019. PyTorch: An imperative style, high-performance deep learning library. In: Wallach, H., Larochelle, H., Beygelzimer, A., d' Alché-Buc, F., Fox, E., Garnett, R. (Eds.), Advances in Neural Information Processing Systems, Vol. 32. Curran Associates, Inc., pp. 8024–8035.

Raviv, T., Kolka, O., 2013. Optimal inventory management of a bike-sharing station. Iie Trans. 45 (10), 1077–1093.

Ren, Y., Chen, H., Han, Y., Cheng, T., Zhang, Y., Chen, G., 2020. A hybrid integrated deep learning model for the prediction of citywide spatio-temporal flow volumes. Int. J. Geogr. Inf. Sci. 34 (4), 802–823.

Sathishkumar, V., Park, J., Cho, Y., 2020. Using data mining techniques for bike sharing demand prediction in metropolitan city. Comput. Commun. 153, 353–366.

Shin, D., Chung, K., Park, R., 2020. Prediction of traffic congestion based on LSTM through correction of missing temporal and spatial data. IEEE Access 8, 150784–150796.

Singhvi, D., Singhvi, S., Frazier, P.I., Henderson, S.G., O'Mahony, E., Shmoys, D.B., Woodard, D.B., 2015. Predicting bike usage for New York City's bike sharing system. In: Workshops At the Twenty-Ninth AAAI Conference on Artificial Intelligence.

Steg, L., Gifford, R., 2005. Sustainable transportation and quality of life. J. Transp. Geogr. 13 (1), 59–69.

Sun, S., Zhang, C., Yu, G., 2006. A Bayesian network approach to traffic flow forecasting. IEEE Trans. Intell. Transp. Syst. 7 (1), 124–132.

Sun, S., Zhang, C., Zhang, Y., 2005. Traffic flow forecasting using a spatio-temporal Bayesian network predictor. In: International Conference on Artificial Neural Networks. Springer, pp. 273–278.

Szeto, M.W., Gazis, D.C., 1972. Application of Kalman filtering to the surveillance and control of traffic systems. Transp. Sci. 6 (4), 419–439.

Taylor, J., Zhou, X., Rouphail, N.M., Porter, R.J., 2015. Method for investigating intradriver heterogeneity using vehicle trajectory data: A dynamic time warping approach. Transp. Res. B 73, 59–80.

Tian, C., Zhu, X., Hu, Z., Ma, J., 2021. A transfer approach with attention reptile method and long-term generation mechanism for few-shot traffic prediction. Neurocomputing 452, 15–27.

Wang, B., Kim, I., 2018. Short-term prediction for bike-sharing service using machine learning. Transp. Res. Procedia 34, 171–178.

Wang, S., Liu, W., Wu, J., Cao, L., Meng, Q., Kennedy, P.J., 2016. Training deep neural networks on imbalanced data sets. In: 2016 International Joint Conference on Neural Networks. IJCNN, IEEE, pp. 4368–4374.

Wang, Z., Su, X., Ding, Z., 2020. Long-term traffic prediction based on LSTM encoder-decoder architecture. IEEE Trans. Intell. Transp. Syst..

Wu, Y., Tan, H., Qin, L., Ran, B., Jiang, Z., 2018. A hybrid deep learning based traffic flow prediction method and its understanding. Transp. Res. C 90, 166–180.

Xiangxue, W., Lunhui, X., Kaixun, C., 2019. Data-driven short-term forecasting for urban road network traffic based on data processing and LSTM-RNN. Arab. J. Sci. Eng. 44 (4), 3043–3060.

Xu, Y., Chen, D., Zhang, X., Tu, W., Chen, Y., Shen, Y., Ratti, C., 2019. Unravel the landscape and pulses of cycling activities from a dockless bike-sharing system. Comput. Environ. Urban Syst. 75, 184–203.

Xu, D., Wang, Y., Jia, L., Qin, Y., Dong, H., 2017. Real-time road traffic state prediction based on ARIMA and Kalman filter. Front. Inf. Technol. Electron. Eng. 18 (2), 287–302.

Yang, Y., Heppenstall, A., Turner, A., Comber, A., 2020. Using graph structural information about flows to enhance short-term demand prediction in bike-sharing systems. Comput. Environ. Urban Syst. 83, 101521.

Yang, D., Li, S., Peng, Z., Wang, P., Wang, J., Yang, H., 2019. MF-CNN: traffic flow prediction using convolutional neural network and multi-features fusion. IEICE Trans. Inform. Syst. 102 (8), 1526–1536.

Zhang, J., Chen, F., Cui, Z., Guo, Y., Zhu, Y., 2020a. Deep learning architecture for short-term passenger flow forecasting in urban rail transit. IEEE Trans. Intell. Transp. Syst. 1–11. http://dx.doi.org/10.1109/TITS.2020.3000761.

Zhang, Y., Cheng, T., Ren, Y., Xie, K., 2020b. A novel residual graph convolution deep learning model for short-term network-based traffic forecasting. Int. J. Geogr. Inf. Sci. 34 (5), 969–995.

Zhang, D., Kabuka, M.R., 2018. Combining weather condition data to predict traffic flow: A GRU-based deep learning approach. IET Intell. Transp. Syst. 12 (7), 578–585.

Zhang, Y., Liu, Y., 2009. Traffic forecasting using least squares support vector machines. Transportmetrica 5 (3), 193–213.

Zhang, C., Zhang, H., Yuan, D., Zhang, M., 2018. Citywide cellular traffic prediction based on densely connected convolutional neural networks. IEEE Commun. Lett. 22 (8), 1656–1659.

Zhang, J., Zheng, Y., Qi, D., 2017. Deep spatio-temporal residual networks for citywide crowd flows prediction. In: Thirty-First AAAI Conference on Artificial Intelligence.

Zhao, Z., Chen, W., Wu, X., Chen, P.C., Liu, J., 2017. LSTM network: A deep learning approach for short-term traffic forecast. IET Intell. Transp. Syst. 11 (2), 68–75.

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. Inform. Sci. 561, 274–285.