

Article

A Spatiotemporal Multi-View-Based Learning Method for Short-Term Traffic Forecasting

Shifen Cheng ^{1,2,3}, Feng Lu ^{1,2,3,4,*} , Peng Peng ^{1,2} and Sheng Wu ^{3,5}

- ¹ State Key Lab of Resources and Environmental Information System, Institute of Geographic Sciences and Natural Resources Research, Chinese Academy of Sciences, Beijing 100101, China; chengsf@reis.ac.cn (S.C.); pengp@reis.ac.cn (P.P.)
- ² University of Chinese Academy of Sciences, Beijing 100049, China
- ³ Fujian Collaborative Innovation Center for Big Data Applications in Governments, Fuzhou 350003, China; ws0110@163.com
- ⁴ Jiangsu Center for Collaborative Innovation in Geographical Information Resource Development and Application, Nanjing 210023, China
- ⁵ Spatial Information Research Center of Fujian Province, Fuzhou University, Fuzhou 350002, China
- * Correspondence: luf@reis.ac.cn; Tel.: +86-10-6488-8966

Received: 24 April 2018; Accepted: 13 June 2018; Published: 14 June 2018



Abstract: Short-term traffic forecasting plays an important part in intelligent transportation systems. Spatiotemporal k-nearest neighbor models (ST-KNNs) have been widely adopted for short-term traffic forecasting in which spatiotemporal matrices are constructed to describe traffic conditions. The performance of the models is closely related to the spatial dependencies, the temporal dependencies, and the interaction of spatiotemporal dependencies. However, these models use distance functions and correlation coefficients to identify spatial neighbors and measure the temporal interaction by only considering the temporal closeness of traffic, which result in existing ST-KNNs that cannot fully reflect the essential features of road traffic. This study proposes an improved spatiotemporal k-nearest neighbor model for short-term traffic forecasting by utilizing a multi-view learning algorithm named MVL-STKNN that fully considers the spatiotemporal dependencies of traffic data. First, the spatial neighbors for each road segment are automatically determined using cross-correlation under different temporal dependencies. Three spatiotemporal views are built on the constructed spatiotemporal closeness, periodic, and trend matrices to represent spatially heterogeneous traffic states. Second, a spatiotemporal weighting matrix is introduced into the ST-KNN model to recognize similar traffic patterns in the three spatiotemporal views. Finally, the results of traffic pattern recognition under these three spatiotemporal views are aggregated by using a neural network algorithm to describe the interaction of spatiotemporal dependencies. Extensive experiments were conducted using real vehicular-speed datasets collected on city roads and expressways. In comparison with baseline methods, the results show that the MVL-STKNN model greatly improves short-term traffic forecasting by lowering the mean absolute percentage error between 28.24% and 46.86% for the city road dataset and, between 53.80% and 90.29%, for the expressway dataset. The results suggest that multi-view learning merits further attention for traffic-related data mining under such a dynamic and data-intensive environment, which owes to its comprehensive consideration of spatial correlation and heterogeneity as well as temporal fluctuation and regularity in road traffic.

Keywords: short-term traffic forecasting; spatiotemporal k-nearest neighbor model; spatiotemporal dependencies; multi-view based learning; traffic patterns

1. Introduction

Accurate and reliable short-term traffic prediction has long been a focus of intelligent transportation systems and location-based services. However, the dynamic, heterogeneous, and nonlinear characteristics of traffic conditions lead to extremely unpredictable changes in the entire traffic network, which makes it difficult to accurately model traffic conditions. Consequently, short-term traffic prediction remains a challenging issue [1].

Researchers have proposed a series of short-term traffic prediction models in past decades [2,3] such as ARIMA [4], Kalman filters [5,6], k-nearest neighbor (KNN) [7–10], and neural network models [11,12] among others. However, these models often fail to simultaneously consider both the spatial and temporal characteristics of traffic and neglect the spatial heterogeneity and temporal non-stationarity of traffic influence, which leads to large deviations in prediction accuracy [13,14]. As a typical system with spatiotemporal distribution, road traffic has the essential characteristics of spatiotemporal dependence compared with general time-series data and static spatial data [15,16]. Therefore, giving synchronous consideration to spatial and temporal information within a prediction model can better identify the traffic conditions on road segments [17,18]. With regard to this, a series of methods have been proposed for short-term traffic prediction [19], which can be roughly divided into parametric and nonparametric spatiotemporal modeling methods.

The most representative parametric spatiotemporal method is ST-ARIMA [20–22]. Considering that spatiotemporal parametric models usually quantitatively express spatiotemporal relationships using explicit parameterization functions and need to make strong assumptions in the modeling process. These approaches are not suitable for simulating real traffic application scenarios. In contrast, non-parametric spatiotemporal models are data-driven and, therefore, require no prior knowledge or explicit mathematical expression. Consequently, they can easily achieve satisfactory portability and comparatively greater prediction accuracy. Therefore, non-parametric spatiotemporal models are more popular for short-term traffic prediction problems [23].

As a typical non-parametric spatiotemporal modeling method, the ST-KNN model is widely used in traffic prediction. Wu et al. introduced spatial and temporal information to the traditional KNN model to achieve more accurate short-term traffic predictions [24]. Yu et al. considered the time-varying property and continuity of traffic conditions and realized multi-step prediction of short-term traffic conditions [25]. Xia et al. optimized the search mechanism of a KNN model by considering the spatiotemporal correlation and the trend of traffic flow and by implementing the KNN model with spatiotemporal weights [26]. Cai et al. constructed the spatiotemporal state matrix rather than the time series of the traditional KNN model and defined the distance function using a Gaussian weighting function to select candidate neighbors. The resulting spatiotemporal KNN model was used to realize multi-step short-term traffic prediction [27].

However, the previous studies have some shortcomings. With regard to spatial dependence, existing ST-KNN models mainly employ two methods to capture spatial information when defining spatiotemporal state space. The first includes selecting several road segments neighboring the target road segment such as upstream and downstream segments [24,26]. This approach assumes the existence of widespread spatial auto-correlation on adjacent road segments. However, when constructing the state space, it is difficult or even impossible to know exactly which—and how many—road segments should be included. This means the spatiotemporal relationship between traffic data cannot be clearly quantified [19]. In addition, when the time series problem is transformed into a supervised machine learning problem, the number of neighboring road segments determines the number of selected features. Consequently, artificial selection of adjacent road segments easily leads to dimensional problems, which makes the model's performance difficult to guarantee [28]. The second method includes utilizing the spatial correlation of road traffic to determine the neighbors [27]. This method can directly describe the impact of the surrounding road segments on the target road segment in order to construct the spatiotemporal state space more effectively. However, the number of spatial neighbors still cannot be automatically determined. Consequently, model performance is

heavily reliant on the threshold that is set to filter out those road segments with lower correlation coefficients. In addition, the dimensions of the spatiotemporal state space defined by the existing ST-KNN models are usually globally fixed. After selecting appropriate thresholds for the spatial neighbors and time windows, the road segments share the spatiotemporal dimensions across the whole road network. However, spatial neighbors of road segments should rely on the current traffic conditions, which are smaller during congestion and larger during non-peak periods [29,30]. Considering the heterogeneous characteristics of city road networks, sharing the dimensions of spatiotemporal state space globally across the entire networks is apparently unreasonable [31].

In terms of time dependence, existing ST-KNN models only consider the temporal closeness of traffic conditions when capturing information on the time dimension. This means the traffic conditions on the road segments during recent time intervals are selected to construct the spatiotemporal state space. The implied assumption is that the more recent the time interval, the greater the impact on the current road segment. However, in view of the obvious periodicity of traffic conditions, the constraints of temporal closeness may be broken in some cases. Normally, similar traffic conditions will be repeated daily on weekdays. Such a distinct periodicity should also be adopted besides that of temporal closeness. Moreover, the periodicity varies with changes in seasons, traffic control strategies, or other factors and shows this tendency especially with the change of seasons [32,33]. However, existing ST-KNN models usually only consider the temporal closeness of traffic in the spatiotemporal state space when characterizing traffic conditions and neglect the spatiotemporal dependence relationship of periodicity and tendency.

Furthermore, existing ST-KNN models usually employ weighted distance functions to choose K candidate neighbors such as Gaussian weight [27] and exponential weight with trend adjustment [26]. These methods will undoubtedly improve the accuracy of prediction models to some extent. However, the construction of distance functions often introduces excessive hyper parameters, which further exacerbates the difficulty of the parameter adjustment in the modeling process.

In recent years, multi-view learning methods have been widely used in the field of spatiotemporal modeling. Zheng et al. constructed a hybrid multi-view learning approach for fine-grained air quality prediction [34]. Liu et al. built a multi-view learning method to forecast urban water quality using multiple data sets in different fields [35]. Yi et al. combined empirical statistical models and data-driven algorithms to build a multi-view learning method to reconstruct missing data [36]. Considering the typical spatiotemporal characteristics of road traffic, we were inspired to further explore the application of multi-view learning to short-term traffic prediction.

In response to the shortcomings of existing ST-KNN models, we gave comprehensive consideration to spatiotemporal dependence and proposed an improved ST-KNN model based on multi-view learning (MVL-STKNN) for short-term traffic prediction. First, in the spatial dimension, considering the characteristics of spatial heterogeneity, we model each road segment individually and use a cross-correlation function to automatically determine the spatial neighbors of each road segment. In the time dimension, we consider temporal closeness, periodicity, and trend rates to describe the impacts of historical traffic on current traffic conditions. By integrating information from the spatial and temporal dimensions, we construct a spatiotemporal closeness matrix, spatiotemporal periodic matrix, and spatiotemporal trend matrix of the different dimensions. We built three spatiotemporal views to characterize current traffic conditions. Second, the spatiotemporal weighting matrix is introduced to improve the existing ST-KNN model, which avoids the introduction of additional parameters. This improved ST-KNN model is then used to mine similar traffic patterns, which obtained predictions according to the three spatiotemporal views. Finally, different weightings are assigned to the three spatiotemporal views by using a neural network algorithm to obtain the predicted traffic conditions.

The remainder of this paper is organized as follows. Section 2 proposes the MVL-STKNN model to capture spatial and temporal dependencies and introduces the construction of the spatiotemporal cuboids, improvement of ST-KNN, and the multi-view-based learning method. In Section 3, we first test the heterogeneity of traffic data and then calibrate the parameters of the MVL-STKNN model.

A comprehensive assessment of the model performance is conducted using real road traffic datasets. The experimental results are discussed in Section 4. Section 5 concludes the paper and provides an outlook on future work.

2. Methodology

In this section, we construct the MVL-STKNN model for short-term traffic prediction for which the overall architecture is shown in Figure 1. In the model, ① represents the spatiotemporal cuboid, ② represents the modeling process of the ST-KNN model, ③ represents the sample construction of multi-view learning, ④ represents the multi-view learning process, ⑤ and ⑥ represent the candidate neighbor selection process for the ST-KNN model in training data sets and test data sets, and ⑦ represents the modeling process of different road segments. Elements ①, ②, ③, ⑤, and ⑥ are used to construct the model inputs and ④ is used for model training. The overall framework is a progressive relationship at the logical level, according to the construction of spatiotemporal cuboid and the improved ST-KNN model and the multi-view-based learning methods. The spatiotemporal cuboid is used to reorganize the original historical traffic data and it uses the stacked spatiotemporal state matrix to characterize the temporal and spatial dependence of traffic conditions. The improved ST-KNN model (the red-dashed box in ② is the improved part) exploited the constructed spatiotemporal cuboid as input to similar traffic patterns and, therefore, obtained predictions, according to the spatiotemporal closeness, periodicity, and trend views. The three predictions are used as training samples and input to the multi-view learning model for fusion. After developing the trained MVL-STKNN model, the test sample is the input to obtain predictions of traffic conditions. It should be noted that, since we consider spatial heterogeneity in the modeling process, the main difference between the modeling processes represented by ⑦ and ①–⑥ is that the dimensions of the spatiotemporal cuboids are different.

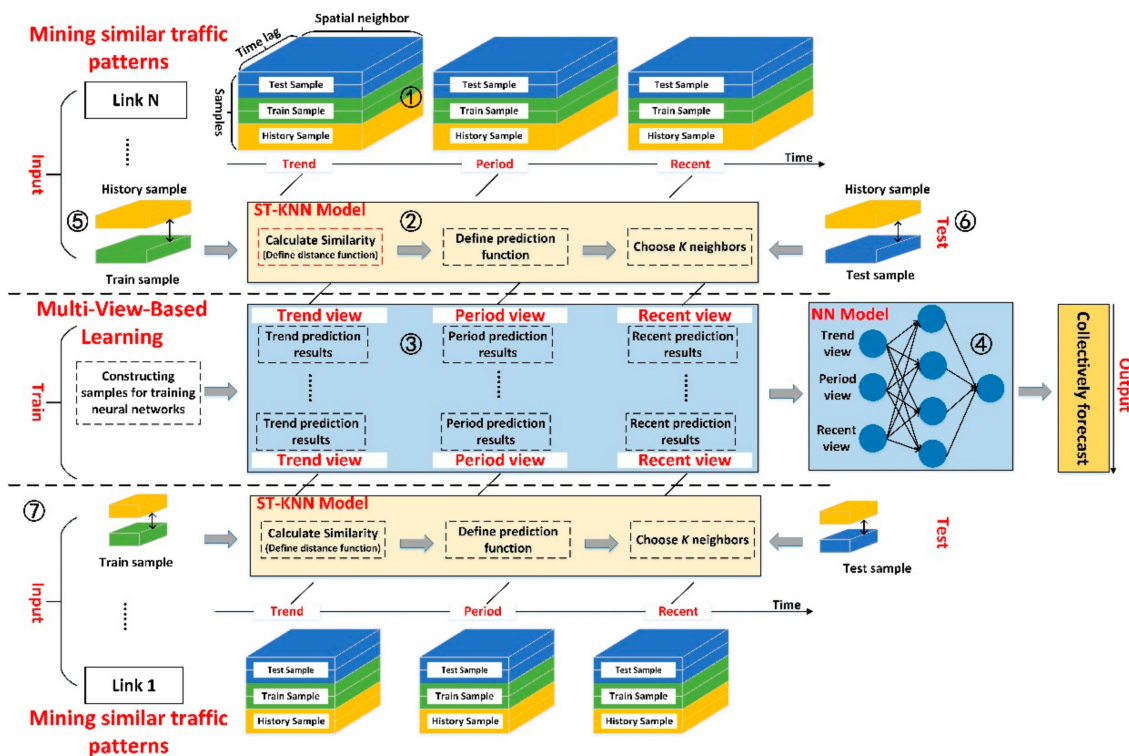


Figure 1. Schematic of the MVL-STKNN model.

2.1. Construction of Space-Time Cuboid

In the road network, traffic conditions on road segments are space–time constrained. In the spatial dimension, traffic conditions on road segments are usually influenced by the surrounding road segments. For example, congestion on one road segment may spread to surrounding segments over time, which leads to regional congestion [37]. In our research study, the surrounding road segments refer to the road segments within the third-level topological neighbors of the predicted road segment. This type of influence has spatial heterogeneity, which means the number of surrounding road segments affected by different road segments is inconsistent. In the time dimension, due to the existence of traffic patterns, the traffic state on road segments is usually associated with historical traffic conditions and displays temporal closeness, periodicity, and trend. Therefore, in the MVL-STKNN model, the time series of historical traffic conditions are reorganized by the fusion of spatiotemporal dimension information to form the spatiotemporal closeness matrix, the period matrix, and the trend matrix, which characterizes the traffic conditions on any road segment at any time. All of the spatiotemporal state matrices are then stacked according to the time necessary to form three spatiotemporal cuboids in order to depict the spatiotemporal dependence of road segments.

By taking the structure of the spatiotemporal closeness cuboid as an example, we assume that the time series representing traffic conditions is $\{v_t^{L_j}, t_0 \leq t \leq tc, 1 \leq j \leq m\}$, where t_0 and tc represent the start time-step and the current time-step of the time series, m is the number of road segments, and $v_t^{L_j}$ is the traffic conditions of the road segment L_j at time interval t . For each road segment L_j , the spatiotemporal closeness state matrix at t time interval can be expressed as $MC_t^{L_j}(lc, ln)$. The elements in the spatiotemporal closeness state matrix represent traffic conditions on the relevant surrounding road segments at historically recent time intervals, which is formally defined by the equation below.

$$MC_t^{L_j}(lc, ln) = \{v_{tk}^{L_k}, t - lc + 1 \leq tk < t, L_k \in \mathcal{R}_{L_j}\} \quad (1)$$

where lc is the length of the closeness-dependent sequence. The traffic conditions of lc historical time intervals adjacent to the t time interval are selected within the range $[t - lc + 1, t)$. \mathcal{R}_{L_j} represents the set of spatial neighbors of road segment L_j and each road segment has a different set of spatial neighbors. ln represents the number of spatial neighbors; $ln = \text{card}(\mathcal{R}_{L_j})$.

In a similar way, the formal definitions of spatiotemporal periodicity matrix $MP_t^{L_j}(lp, ln)$ and spatiotemporal trend matrix $MQ_t^{L_j}(lq, ln)$ are shown in the equations below.

$$MP_t^{L_j}(lp, ln) = \{v_{tk}^{L_k}, tk \in TP, L_k \in \mathcal{R}_{L_j}\} \quad (2)$$

$$MQ_t^{L_j}(lq, ln) = \{v_{tk}^{L_k}, tk \in TQ, L_k \in \mathcal{R}_{L_j}\} \quad (3)$$

where lp is the length of the period-dependent sequence and TP is the value set of the temporal period: $lp = \text{card}(TP)$. lq is the length of the trend-dependent sequence and TQ is the value set of temporal trends: $lq = \text{card}(TQ)$.

$$TP = \{t - 1440/t_interval * p * i, 1 \leq i \leq lp\} \quad (4)$$

$$TQ = \{t - 1440/t_interval * q * i, 1 \leq i \leq lq\} \quad (5)$$

In this scenario, the $t_interval$ is the sampling time interval of traffic conditions such as 5 min (where one day = 1440 min). p is the period span and q is the trend span. For example, $p = 1$ describes daily periodicity and $q = 7$ reveals the weekly trend.

The spatial neighbor set \mathcal{R}_{L_j} of the road segment L_j is automatically obtained using cross-correlation. Considering the delay in spatiotemporal dependency between road segments, the traditional way to find

spatial neighbors using correlation coefficients does not satisfy this requirement. The cross-correlation function, as a delayed version of the correlation coefficient function, measures the correlation coefficients of two time series using a specific delay [38] and is better suited to describing the spatiotemporal dependence of traffic. Assume that the traffic condition time series of two road segments are, respectively: $U = \{v_t^{L_j}\}$, $Z = \{v_t^{L_v}\}$. Then the cross-correlation functions at delayed φ are defined by the equation below.

$$\begin{cases} ccf_{u,z}(\varphi) = \frac{\gamma_{u,z}(\varphi)}{\sigma_u \sigma_z}, \varphi = 0, \pm 1, \pm 2, \dots, \\ \gamma_{u,z}(\varphi) = E[(u_t - \mu_u)(z_{t+\varphi} - \mu_z)] \\ \sigma_u = \sqrt{\sum (u_t - \mu_u)^2} \\ \sigma_z = \sqrt{\sum (z_{t+\varphi} - \mu_z)^2} \end{cases} \quad (6)$$

where $\gamma_{u,z}(\varphi)$ is the cross-correlation coefficient of time series U and time series Z at delay φ . μ_u and μ_z are the mean values of U and Z , respectively. Parameters σ_u and σ_z are the standard deviations of U and Z , respectively. From the above definition, the cross-correlation function can be regarded as a function of time delay so that the maximum time delay value of the cross-correlation function is the average delay time of the surrounding road segments to the predicted road segment [29], which can be formally defined by Equation (7).

$$\psi_{L_v} = |\operatorname{argmax}(ccf_{u,z}(\varphi))|, v \in [1, m] \quad (7)$$

where ψ_{L_v} is the maximum time delay value of the surrounding road segment L_v for predicting the road segment L_j . Given the predicted road segment L_j and its prediction time range Δt , only those surrounding road segments that can be accessed within the maximum time delay are considered and the road segments beyond the time delay limit are excluded.

$$\mathcal{R}_{L_j} \leftarrow \{L_v | \forall 0 \leq \psi_{L_v} \leq \Delta t\} \quad (8)$$

The corresponding maximum time delay value ψ_{L_v} is obtained by calculating the cross-correlation $ccf_{u,z}(\varphi)$ between all the surrounding road segments and the predicted road segments. Finally, all the road segments satisfying the conditions $0 \leq \psi_{L_v} \leq \Delta t$ are added to the set \mathcal{R}_{L_j} .

After completing the construction of the spatiotemporal state matrix at all historical time intervals, we stacked three spatiotemporal state matrices in chronological order and constructed spatiotemporal closeness cuboids, periodic cuboids, and trend cuboids of road segments L_j . Their formal definitions are shown below.

$$XC_{L_j} = \{MC_t^{L_j}(lc, ln), ml + 1 \leq t \leq tc\} \quad (9)$$

$$XP_{L_j} = \{MP_t^{L_j}(lp, ln), ml + 1 \leq t \leq tc\} \quad (10)$$

$$XQ_{L_j} = \{MQ_t^{L_j}(lq, ln), ml + 1 \leq t \leq tc\} \quad (11)$$

where $ml = 1440/t_interval * q * lq$, which ensures that the temporal closeness matrix, the period matrix, and the trend matrix can be taken simultaneously at time interval t .

Finally, we divide the spatiotemporal cuboid $XC_{L_j}, XP_{L_j}, XQ_{L_j}$, respectively, including historical spatiotemporal cuboids and a historical template library for mining similar traffic patterns in the ST-KNN model. We also divide training spatiotemporal cuboids for input to the ST-KNN model in comparison with the historical template library to obtain forecasts from each view in order to construct the training samples for multi-view learning as well as test spatiotemporal cuboids for verification of model prediction accuracy, as shown in Figure 1. We take N as the total days of the time series of traffic conditions, the number of historical days as hd , the number of training days as td , and the number

of test days as sd . Then $N = hd + td + sd + q * lq$, where $q * lq$ is the number of days that the sample cannot be taken simultaneously. The number of historical samples is $n_{hd} = hd * 1440 / t_interval$, the number of training samples is $n_{td} = td * 1440 / t_interval$, and the number of test samples is $n_{sd} = sd * 1440 / t_interval$. Taking the partition of the temporal closeness cuboid as an example, the historical spatiotemporal closeness cuboid $XC_Ht_{L_j}$, the training spatiotemporal closeness cuboid $XC_Tr_{L_j}$, and the test spatiotemporal closeness cuboid $XC_Ts_{L_j}$ can be defined by the formulas below.

$$\left\{ \begin{array}{l} XC_{L_j} = \{XC_Ht_{L_j}, XC_Tr_{L_j}, XC_Ts_{L_j}\} \\ XC_Ht_{L_j} = \{MC_t^{L_j}(lc, ln), ml + 1 \leq t \leq n_{hd} + ml + 1\} \\ XC_Tr_{L_j} = \{MC_t^{L_j}(lc, ln), n_{hd} + ml + 1 < t \leq n_{td} + n_{hd} + ml + 1\} \\ XC_Ts_{L_j} = \{MC_t^{L_j}(lc, ln), n_{td} + n_{hd} + ml + 1 < t \leq tc\} \end{array} \right. \quad (12)$$

The same division method can be used to obtain the historical spatiotemporal period cuboid $XP_Ht_{L_j}$, the training spatiotemporal period cuboid $XP_Tr_{L_j}$, the test spatiotemporal period cuboid $XP_Ts_{L_j}$, the historical spatiotemporal trend cuboid $XQ_Ht_{L_j}$, the training spatiotemporal trend cuboid $XQ_Tr_{L_j}$, and the test spatiotemporal trend cuboid $XQ_Ts_{L_j}$.

2.2. ST-KNN Model

The basic principle of the ST-KNN model is to compare the distance between the spatiotemporal state space, select K closest historical spatiotemporal state matrices, and use predictive functions to integrate the traffic conditions at the next time interval for the target road segment to obtain a final predicted value. In Section 2.1, we obtained the spatiotemporal state matrix of the three views. Therefore, the key here is how to define the distance function and the prediction function.

When defining the distance function, different weightings need to be introduced to describe the influence of traffic conditions at different time intervals and different spatial neighbors on predicting road segments in the spatiotemporal state matrix. In the traditional method, too many parameters are often introduced when setting the weightings, which makes it difficult to adjust the parameters in the modeling process and inhibits global optimal prediction results. Considering that the cross-correlation function characterizes the correlation between the surrounding road segments and the predicted road segments without additional parameters, we use the cross-correlation coefficient to represent the spatial weighting value $w_s^{L_j}(si, sj)$ for the predicted road segment L_j in the spatial dimension, which becomes more closely related to the predicted road segment with greater weighting. In the time dimension, the time weight is assigned according to the linear distribution of time. The closer it is to the predicted time, the greater the assigned weighting. Taking the spatiotemporal closeness matrix as an example, the weighting allocation method is shown below.

$$w_t^{L_j}(ti, tj) = \begin{cases} ti / \sum_{ti=1}^{lc} ti, & ti = tj \\ 0, & ti \neq tj \end{cases} \quad (13)$$

$$w_s^{L_j}(si, sj) = \begin{cases} cc f_{L_v, L_j}^{si} / \sum_{si=1}^{ln} cc f_{L_v, L_j}^{si}, & si = sj \\ 0, & si \neq sj \end{cases} \quad (14)$$

where $cc f_{L_v, L_j}^{si}$ is the cross-correlation coefficient between the $si - th$ spatial neighborhood (whose road segment is L_v) and the time series of the predicted road segment L_j . By introducing space–time weightings into the original spatiotemporal closeness matrices, the spatiotemporal neighboring

weight matrices $\Gamma_t^{L_j}$ and $\Gamma_{tk}^{L_j}$ at time interval t and at a historical time interval tk are represented as Equations (15) and (16).

$$\begin{cases} \Gamma_t^{L_j} = w_s^{L_j}(si, sj) \times MC_t^{L_j}(lc, ln) \times w_t^{L_j}(ti, tj) \\ MC_t^{L_j}(lc, ln) \in \{XC_{Tr_{L_j}}, XC_{Ts_{L_j}}\} \end{cases} \quad (15)$$

$$\begin{cases} \Gamma_{tk}^{L_j} = w_s^{L_j}(si, sj) \times MC_{tk}^{L_j}(lc, ln) \times w_t^{L_j}(ti, tj) \\ MC_{tk}^{L_j}(lc, ln) \in \{XC_{Ht_{L_j}}\} \end{cases} \quad (16)$$

It should be noted that $MC_{tk}^{L_j}(lc, ln)$ belongs to the historical sample library of the ST-KNN model. Therefore, its value range is limited to the set $\{XC_{Ht_{L_j}}\}$. Furthermore, $MC_t^{L_j}(lc, ln)$ is used for comparing with $MC_{tk}^{L_j}(lc, ln)$ to obtain the candidate neighbors. Therefore, its value range is $\{XC_{Tr_{L_j}}, XC_{Ts_{L_j}}\}$. By calculating the distance $d(\Gamma_t^{L_j}, \Gamma_{tk}^{L_j})$ between the spatiotemporal closeness matrix at time interval t and the historical spatiotemporal closeness matrix, it can be used to select K candidate neighbors. The formula is below.

$$d(\Gamma_t^{L_j}, \Gamma_{tk}^{L_j}) = \sqrt{\text{trac}((\Gamma_t^{L_j} - \Gamma_{tk}^{L_j}) \times (\Gamma_t^{L_j} - \Gamma_{tk}^{L_j})^T)} \quad (17)$$

where trac represents the trace of the matrix. The distance function of the spatiotemporal periodic and trend matrix can be defined in a similar way.

When defining the predictive function, we used the same strategy as in Reference [27] and utilized the Gaussian function to assign different weightings to the selected K candidate neighbors to obtain the predicted value for the target road segment L_j . Taking the spatiotemporal closeness cuboid as an example, the form is defined below.

$$\widehat{v_{t+1,sc}^{L_j}} = \frac{\sum_{k=1}^K v_{tk+1}^{L_j,k} \times \omega_{L_j}(k)}{\sum_{k=1}^K \omega_{L_j}(k)} \quad (18)$$

where $\widehat{v_{t+1,sc}^{L_j}}$ represents the prediction value of road segment L_j at time interval $t + 1$ and $v_{tk+1}^{L_j,k}$ represents the traffic condition of k -th candidate neighbors at the next time interval. $\omega_{L_j}(k)$ represents the weighting of the k -th candidate neighbors of the prediction road segment L_j . Its form is defined below.

$$\omega_{L_j}(k) = \frac{1}{4\pi a^2} \exp\left(-\frac{|d_k^{L_j}|^2}{4a^2}\right) \quad (19)$$

where $d_k^{L_j}$ is the distance between the k -th candidate neighbor and the predicted road segment L_j ; a is the spatiotemporal parameter. Using the same prediction function, we can obtain the predicted values $\widehat{v_{t+1,sp}^{L_j}}$ and $\widehat{v_{t+1,sq}^{L_j}}$ of the periodic view and the trend view of the road segment L_j at time interval t , respectively.

2.3. Multi-View-Based Learning

The basic principle of the multi-view learning model is to build a supervised learning method, which takes the prediction results of the spatiotemporal closeness, periodicity, and trend views as inputs to neural network models. The real values for traffic conditions as outputs.

The process of training the MVL-STKNN model is shown in Algorithm 1. First, we use the ST-KNN model to obtain predicted values for the training spatiotemporal closeness matrix $MC_t^{L_j}(lc, ln)$,

the training spatiotemporal periodic matrix $MP_t^{L_j}(lp, ln)$, and the training spatiotemporal trend matrix $MQ_t^{L_j}(lq, ln)$ for each spatiotemporal state matrix in training spatiotemporal cuboids (lines 4–6 lines in Algorithm 1). Afterward, the predictive values of the three views are used as the feature vector and the true values are used as label values to construct the training sample (line 7). Finally, the training samples are input to the neural network model for supervised learning (line 9) and the MVL-STKNN model \mathcal{M} can be obtained (line 10).

Algorithm 1: Training of MVL-STKNN

Input: Near spatiotemporal cuboids: $\{XC_Ht_{L_j}, XC_Tr_{L_j}\}$;
 Periodic spatiotemporal cuboid: $\{XP_Ht_{L_j}, XP_Tr_{L_j}\}$;
 Trend spatiotemporal cuboid: $\{XQ_Ht_{L_j}, XQ_Tr_{L_j}\}$;
 Lengths of closeness, period, trend: lc, lp, lq ;
 Number of candidate neighbors: K ;
 Parameter of Gaussian function: a .
Output: MVL-STKNN model \mathcal{M} .
 // construct training instances
 1 $\mathcal{D} \leftarrow \emptyset$
 2 **For** all time interval t in the training spatiotemporal cuboids
 3 // $n_{hd} + ml + 1 < t \leq n_{td} + n_{hd} + ml + 1$
 4 $\widehat{v}_{t+1,sc}^{L_j} = \text{ST-KNN}(XC_Ht_{L_j}, MC_t^{L_j}(lc, ln), K, a)$ // $MC_t^{L_j}(lc, ln) \in \{XC_Tr_{L_j}\}$
 5 $\widehat{v}_{t+1,sp}^{L_j} = \text{ST-KNN}(XP_Ht_{L_j}, MP_t^{L_j}(lp, ln), K, a)$ // $MP_t^{L_j}(lp, ln) \in \{XP_Tr_{L_j}\}$
 6 $\widehat{v}_{t+1,sq}^{L_j} = \text{ST-KNN}(XQ_Ht_{L_j}, MQ_t^{L_j}(lq, ln), K, a)$ // $MQ_t^{L_j}(lq, ln) \in \{XQ_Tr_{L_j}\}$
 7 Put a training instance $\left(\left\{ \widehat{v}_{t+1,sc}^{L_j}, \widehat{v}_{t+1,sp}^{L_j}, \widehat{v}_{t+1,sq}^{L_j} \right\}, v_{t+1}^{L_j} \right)$ into \mathcal{D}
 8 **End for**
 // Training the model
 9 $\mathcal{M} \leftarrow \text{Multi_view_learning}(\mathcal{D})$ // Neural network training
 10 Output the learned MVL-STKNN model \mathcal{M}

After obtaining the trained MVL-STKNN model \mathcal{M} , we can predict all samples of the test spatiotemporal cuboid. The prediction process is shown in Algorithm 2. First, the spatiotemporal closeness matrix $MC_t^{L_j}(lc, ln)$, spatiotemporal periodic matrix $MP_t^{L_j}(lp, ln)$, and spatiotemporal trend matrix $MQ_t^{L_j}(lq, ln)$ in all the test spatiotemporal cuboids are input to the ST-KNN model to obtain the predicted values of the three views (lines 3–5 in Algorithm 2). Note that the range of the spatiotemporal state matrix is different from the training process. Additionally, the predicted values obtained by the ST-KNN model are input to the trained MVL-STKNN model \mathcal{M} and we get the final prediction value of the road segment at the next time interval (line 6). Finally, the predicted values for all the test samples are saved to set χ_{L_j} for evaluating the accuracy of the MVL-STKNN model (line 7).

Algorithm 2: Prediction of MVL-STKNN

Input: Near spatiotemporal cuboids: $\{XC_Ht_{L_j}, XC_Ts_{L_j}\}$;
 Periodic spatiotemporal cuboid: $\{XP_Ht_{L_j}, XP_Ts_{L_j}\}$;
 Trend spatiotemporal cuboid: $\{XQ_Ht_{L_j}, XQ_Ts_{L_j}\}$;
 Lengths of closeness, period, trend: lc, lp, lq ;
 Number of candidate neighbors: K ;
 Parameter of Gaussian function: a .
Output: Set of test sample predictions: χ_{L_j} .

1 For all time interval t in the test spatiotemporal cuboids
2 // $n_{td} + n_{hd} + ml + 1 < t \leq tc$
3 $\widehat{v}_{t+1,sc}^{L_j} = \text{ST-KNN}(XC_Ht_{L_j}, MC_t^{L_j}(lc, ln), K, a)$ // $MC_t^{L_j}(lc, ln) \in \{XC_Ts_{L_j}\}$
4 $\widehat{v}_{t+1,sp}^{L_j} = \text{ST-KNN}(XP_Ht_{L_j}, MP_t^{L_j}(lp, ln), K, a)$ // $MP_t^{L_j}(lp, ln) \in \{XP_Ts_{L_j}\}$
5 $\widehat{v}_{t+1,sq}^{L_j} = \text{ST-KNN}(XQ_Ht_{L_j}, MQ_t^{L_j}(lq, ln), K, a)$ // $MQ_t^{L_j}(lq, ln) \in \{XQ_Ts_{L_j}\}$
6 $\widehat{v}_{t+1}^{L_j} \leftarrow \mathcal{M}\left(\widehat{v}_{t+1,sc}^{L_j}, \widehat{v}_{t+1,sp}^{L_j}, \widehat{v}_{t+1,sq}^{L_j}\right)$ // Obtain the predicted values
7 Put $\widehat{v}_{t+1}^{L_j}$ into χ_{L_j} // Save the predicted values into set χ_{L_j}
8 End for
9 Return the set of predictions χ_{L_j}

3. Performance Evaluation

In this section, we use floating car-speed data collected for the Beijing road network and the California Freeway and Expressway systems to evaluate the proposed MVL-STKNN model. First, we pre-process the raw traffic data and demonstrate its heterogeneous attributes. Afterward, in the Beijing data set, we adjust the parameters of the MVL-STKNN model to obtain a set of optimal parameter combination values. On this basis, we compare existing baseline methods to verify the efficiency of the proposed model. At the same time, the generalizability of the MVL-STKNN model was tested by further comparing the accuracy of different traffic prediction models using the Caltrans Performance Measurement System (PeMS) datasets. Finally, taking the Beijing dataset as an example, the impact of each component on the overall prediction accuracy of the model is explored including the time-space weighting and spatiotemporal dependence components proposed in this study.

3.1. Data Preparation**3.1.1. Data Sources**

We used floating car-speed data collected from the Beijing road network and the California Freeway and Expressway systems to evaluate the performance of the models by predicting the vehicular speed of the road segment. These data points are widely applied in the field of transportation [14,23,39,40]. Among these, PeMS continuously collects real traffic data from more than 8100 locations on the California Freeway and Expressway systems. These data are integrated into multiple time intervals and are freely available on the network [41]. PeMS travel speed data were downloaded for 59 consecutive locations on the US Route 101 and the data distribution is shown in Figure 2. The data were collected at 5-min intervals for 60 days (15 August 2016 to 14 October 2016), which was shown in Table 1. The Beijing dataset comes from the driving track of a GPS-equipped floating car by uploading driving-related information to the server at one-minute intervals, including travel time, direction, and speed. The time period is from 1 March 2012 to 30 April 2012, which is shown in Table 1. We integrated the data into 5-min intervals, calculate the average speed of each road segment, and finally selected 30 representative road segments for subsequent experiments. The data distribution is shown in Figure 3. In both data sets, data from the last five days were used as

3.1.2. Data Processing

Due to equipment failure and other factors, some data were missing from the original traffic-speed data sets. Considering the spatiotemporal characteristics of traffic data, we used the existing spatiotemporal interpolation algorithm to fill in the missing data in order to better reconstruct the original traffic conditions [42]. The basic principle is to consider the missing data pattern in the interpolation process and use coarse-grained interpolation to obtain partial reconstruction results in order to eliminate the effect of missing continuous block data on the subsequent interpolation process. On this basis, considering the spatiotemporal heterogeneity, fine-grained interpolation was applied to the spatial and temporal dimensions. Finally, the unbiased estimation values of the missing data were obtained by integrating the interpolation results of the spatiotemporal dimensions in a nonlinear way. Considering that travel speeds on different types of road segments vary greatly across the road network, we normalized the original traffic data and used the ratio of the average speed to a maximum speed limit on each road segment to represent the traffic conditions of road segments. This is expressed in Formula (20).

$$\overline{v}_t^{L_j} = \frac{v_t^{L_j}}{f_{j,max}}, \quad j \in [1, m], t \in [t_0, tc] \quad (20)$$

In this study, $\overline{v}_t^{L_j}$ is the normalized speed of the j -th road segment at time interval t , $v_t^{L_j}$ is the average speed of the road segment and $f_{j,max}$ is the speed limit of the j -th road segment.

3.2. Evaluation Metrics

Similarly, with regard to other traffic prediction studies [9,10,26,27], the present study uses the mean absolute percentage error (MAPE) as a measure of performance, which reflects the percentage difference between predicted and actual traffic conditions. The prediction accuracy of different models is depicted by averaging the percentage error of traffic conditions for all road segments during the test time interval. A smaller MAPE value indicates higher accuracy of the prediction model. The formal definition is below.

$$\left\{ \begin{array}{l} MAPE = \frac{1}{m \times n_{sd}} \sum_{j=1}^m \sum_{t=ts}^{tc} \frac{|v_t^{L_j} - \widehat{v}_t^{L_j}|}{\widehat{v}_t^{L_j}} \\ ts = n_{td} + n_{hd} + ml + 1 \\ n_{sd} = tc - ts + 1 \end{array} \right. \quad (21)$$

where ts is the start time interval of the test data and $v_t^{L_j}$ and $\widehat{v}_t^{L_j}$ are the actual vehicular speed and the predicted vehicular speed at the next time step of the j -th predicted road segment at the current time step. After dividing the test data set for the original traffic time series, the equations $ts = n_{td} + n_{hd} + ml + 1$ and $n_{sd} = tc - ts + 1$ are satisfied.

3.3. Variable Estimation

The MVL-STKNN hyper parameters include parameter a , the number of candidate neighbors K , the length of the closeness-dependent sequence lc , the length of the periodic-dependent sequence lp , and the length of the trend-dependent sequence lq . We set up a reasonable range of values for each parameter in order to find the combination of parameter values that achieves the lowest MAPE where $a \in [0.001, 0.003, \dots, 0.013, 0.015]$, $K \in [5, 10, \dots, 25, 30]$, $lc \in [1, 2, \dots, 5, 6]$, $lp \in [1, 2, 3, 4]$, and $lq \in [1, 2, 3]$. During this estimation process, the parameters were adjusted in a progressive manner. First, a and K were determined to obtain the optimal ST-KNN model. On this basis, the effect of temporally dependent sequence length on model accuracy was tested to determine the values of lc , lp , and lq .

3.3.1. Calibrating the Parameters of ST-KNN Model

Figure 4 shows the effect of the combined values of parameters a and K on the prediction accuracy of the model. We fixed $lc = 3$, $lp = 4$, and $lq = 2$ while varying the values of parameters a and K , which is shown in Figure 4. It can be seen that the MVL-STKNN model has the smallest MAPE value when $a = 0.009$ and $K = 5$. MAPE shows a small overall change, which indicates that the values of parameters a and K have little influence on the overall prediction accuracy of the MVL-STKNN model.

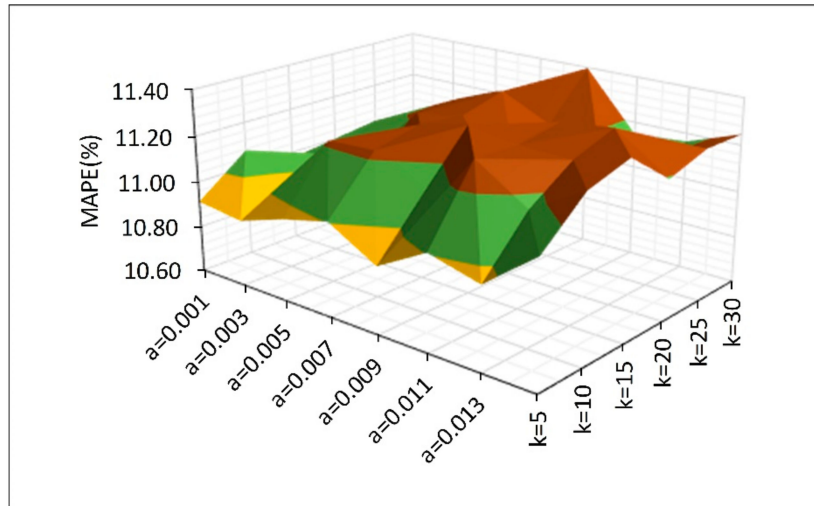


Figure 4. Impact of ST-KNN model parameters.

3.3.2. Calibrating the Temporally Dependent Parameters

Furthermore, we verified the impacts of temporally dependent parameters on model performance, which is shown in Figure 5. First, we fixed $lp = 4$, $lq = 2$, and varied the value of the temporal closeness parameter lc to test its effect on prediction accuracy. As shown in Figure 5a, by changing lc , the prediction accuracy first increased and then decreased. The lowest MAPE value was recorded for $lc = 2$. Afterward, we fixed $lc = 2$, $lq = 2$, and varied the value of the temporal period parameter lp . The resulting prediction accuracies are shown in Figure 5b. The MVL-STKNN model has the highest prediction accuracy when $lp = 1$. As lp increases, prediction accuracy gradually decreases. The results show that short-range periods are helpful for prediction models while long-range periods are difficult to model. Finally, we fixed $lc = 2$, $lp = 1$, and varied the time trend parameter lq . As shown in Figure 5c, the model has a nearly equal prediction performance when $lq = 1$ or $lq = 2$, which indicates that the short-range trend is easier to capture. When $lq = 3$, the value of MAPE rose sharply. The reason is that the dynamic changes in traffic conditions obscure the long-range trend, which affects the prediction accuracy of the model.

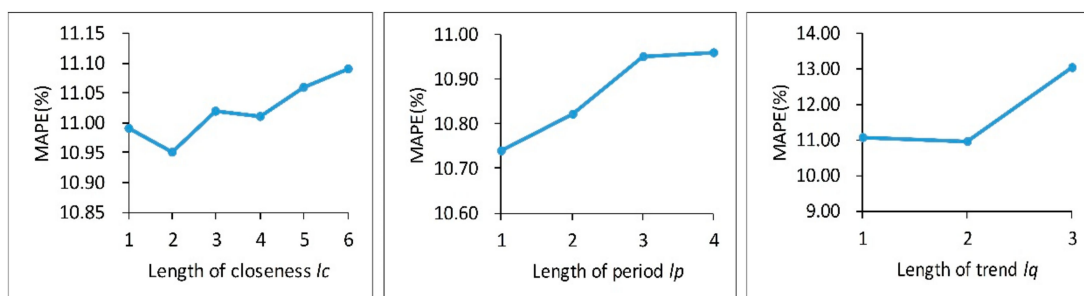


Figure 5. Impact of temporal dependent parameters.

Through the above parameter-debugging process, the parameter-calibration results of the MVL-STKNN model are shown in Table 2.

Table 2. Calibrated model parameters.

Parameters	Values
a	0.009
K	5
lc	2
lp	1
lq	2

3.4. Test of Spatial Heterogeneity

After completing the data preprocessing, we tested the heterogeneous nature of traffic. Taking 30 road segments in Beijing as an example, we used cross-correlation to automatically determine the spatial neighbors of each road segment, according to the selection strategy of spatial neighbors. The result is shown in Figure 6. The abscissa represents the IDs of the road segments and the ordinate represents the number of spatial neighbors of the road segments. For example, Link 1 has 15 spatial neighbors and Link 2 has eight spatial neighbors. It can be observed that the number of spatial neighbors of different road segments is inconsistent. The above results show that the set of spatial neighbors and the number of spatial neighbors differ for each road segment, which reflects the obvious heterogeneity of urban road traffic. Therefore, when constructing the spatiotemporal state matrix, previous studies used global fixed spatiotemporal dimensions to represent traffic conditions, which cannot reflect the heterogeneity of traffic conditions on the road network. It also confirms that the selection strategy of adaptive spatial neighbor adopted in this research study is reasonable.

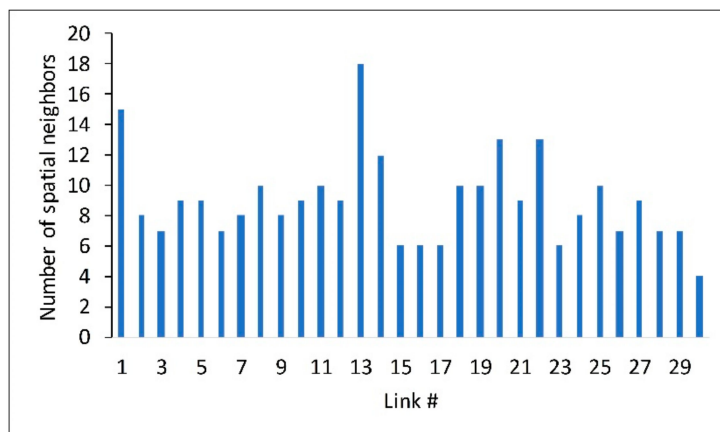


Figure 6. Adaptive spatial neighbors of each road segment.

3.5. Accuracy Comparison

In order to verify the efficiency of the proposed MVL-STKNN model, we compared it with four existing traffic predicting models in Beijing and PeMS data sets, which are comprised of the historical average model (HA), the Elman neural network (Elman-NN), the KNN model (Original-KNN), and the spatiotemporal KNN model (ST-KNN). Of these, the HA model uses the average value of the historical time window as the prediction value of traffic condition at the next time interval. The Elman-NN model is extended to short-term traffic prediction by introducing delay operators into the model to adapt traffic time-varying features [43]. The Original-KNN model is an instance-based nonparametric supervised learning algorithm that heuristically compares the similarity of historical traffic time series with the current observed time series. Many researchers have successfully applied the KNN model

to short-term traffic prediction [25,44,45]. Based on the Original-KNN model, the ST-KNN model is formed by considering the spatiotemporal relationships among multiple road segments in the road network.

Figure 7 shows the prediction accuracy of the four models when using the Beijing dataset. It can be seen that because the HA, the Elman-NN, and the Original-KNN models consider traffic forecasting as a time series modeling problem, they ignore the influence of spatial factors on the predicted road segments. Therefore, their predictive performances are worse than those of the ST-KNN model and the proposed MVL-STKNN model. The ST-KNN model introduces the spatiotemporal closeness matrix rather than the traditional time series to characterize traffic conditions and constructs the distance function and prediction function using the Gaussian function to improve the predictive performance of the model. However, in the modeling process, the spatiotemporal dependence relationship is not fully characterized. For example, in the spatial dimension, the heterogeneity of road network traffic is not considered, which results in a globally fixed dimension of its spatiotemporal adjacency matrix. In the temporal dimension, the influence of temporal periodicity and trend on the prediction of road segment traffic conditions is ignored. In addition, when the distance function is constructed, too many parameters are artificially introduced, which exacerbates the difficulty of the parameter adjustment. Consequently, the predictive accuracy of the model is significantly lower than the method proposed here. In the MVL-STKNN model, the improved ST-KNN model is first introduced by using space-time weighting rather than Gaussian weighting. On this basis, the spatiotemporal dependencies are comprehensively considered and three spatiotemporal views are constructed (spatiotemporal closeness, periodic, and trend matrices) to characterize the traffic conditions. Finally, short-term traffic prediction is achieved by using a multi-view learning method. It can be seen from Figure 7 that the MVL-STKNN model reduces the MAPE between 28.24% and 46.86% compared with the existing methods, which indicates that the proposed method solves the problems existing in the baseline method.

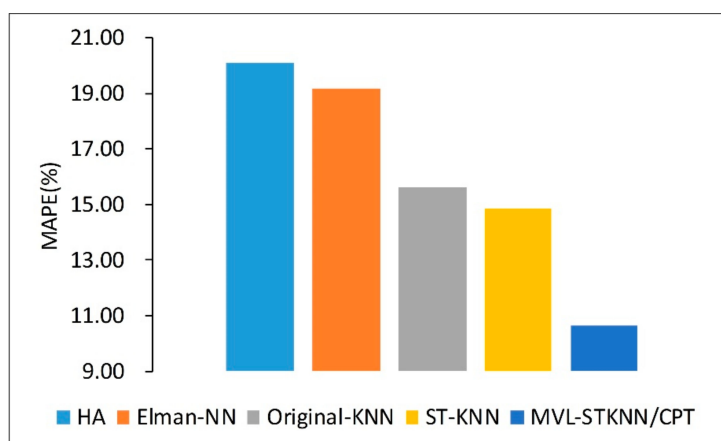


Figure 7. Comparison with baselines using the Beijing data set.

Since urban road networks and freeway exhibit different road network topological structures and have large differences in traffic patterns, we further validated the model's performance on PeMS datasets. The experimental results are shown in Figure 8. Overall, the accuracy trend of the model was similar to that of the Beijing dataset, but the predictive accuracy was more extensive. This is attributed to higher data quality. The PeMS data set is relatively complete and, because the data collection area is an expressway, the traffic pattern is relatively simple when compared with an urban road network. The change trend is small, which makes it easier to capture the regular traffic pattern. As can be seen from Figure 8, the MVL-STKNN model reduces the MAPE between 53.80% and 90.29%. When compared with the existing methods, this further demonstrates the effectiveness of the proposed method and its satisfactory generalization.

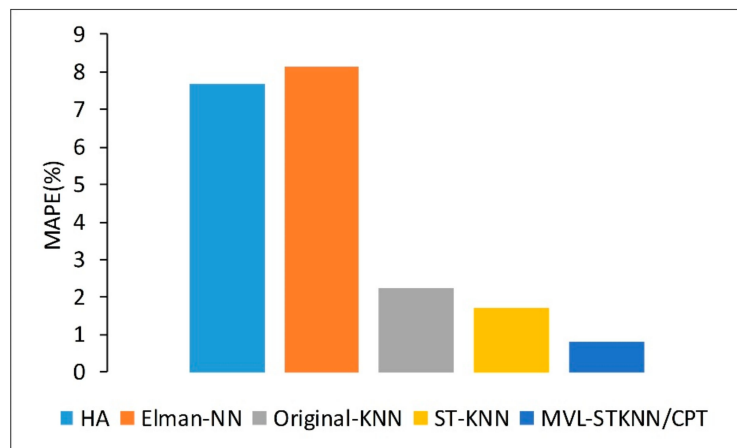


Figure 8. Comparison with baselines using the PeMS data set.

3.6. Impact of Space-Time Weighting Matrix

Based on the existing ST-KNN model, we use space–time weighting allocation instead of the traditional Gaussian function allocation method to construct the STKNN-W model. The experimental results in Figure 9 clearly demonstrate that the proposed method achieves superior performance, which is attributed to the MVL-STKNN model by employing the ST-KNN model to mine similar traffic patterns and to derive predictions for the spatiotemporal closeness, periodic, and trend views. In the training process, these three sets of predictions are then used to construct training samples for multi-view learning. During testing, the parameters trained through multi-view learning are used to integrate the three views to give predicted values for traffic conditions on the road segments. Therefore, the performance of the ST-KNN model will have a certain impact on the overall prediction results. In the ST-KNN model, the key is how to construct a suitable distance function to select the candidate neighbors. The existing ST-KNN constructs the distance function using a Gaussian function to allocate weightings in the time and space dimensions, respectively. Compared with Euclidean distance and correlation distance, the predictive accuracy of the model is somewhat improved [27]. However, the Gaussian weight function requires the introduction of additional parameters, which exacerbates the difficulty of parameter calibration. In contrast, the proposed weight allocation method does not require additional parameters. Moreover, the influence of traffic conditions with different time intervals and different spatial neighbors on the predicted road segments can be mined from the data itself, which achieves higher prediction accuracy.

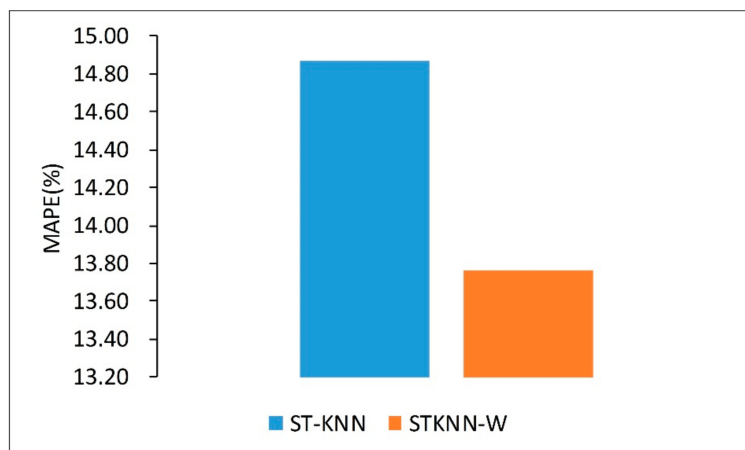


Figure 9. Impact of space-time weighting matrix.

3.7. Impact of Spatial and Temporal Dependencies

In this section, we further discuss the impacts of spatial and temporal dependencies on the overall results.

First, we test the effect of spatiotemporal closeness on predictive accuracy. Considering that the existing ST-KNN and STKNN-W models use the globally fixed spatiotemporal closeness matrix to characterize traffic conditions, they cannot cope with the heterogeneous nature of traffic conditions. Therefore, based on the STKNN-W model, we hold other variables fixed and automatically select the spatial neighbors for each road segment using cross-correlation so that the spatial neighbors of each road segment are different (see Figure 4). Afterward, the STKNN-W-C model is constructed using the adaptive spatiotemporal closeness matrix rather than the traditional globally fixed spatiotemporal closeness matrix. As can be seen from Figure 10, the STKNN-W-C model is more accurate than STKNN-W, which confirms the validity of the adaptive spatiotemporal closeness matrix.

Second, on the basis of the STKNN-W-C model, we further introduce temporal periodicity, which is combined with an adaptive spatial neighbor to form a spatiotemporal periodic matrix to represent traffic conditions from another view. Since the entire model contains two views, it can constitute a multi-view learning method known as MVL-STKNN/CP, which considers both spatiotemporal closeness and period. The training and prediction process is similar to that shown in Algorithms 1 and 2 except that the spatiotemporal trend view is not included. As shown in Figure 10, the performance of the MVL-STKNN/CP model is significantly improved, which demonstrates the usefulness of spatiotemporal periodicity in the modeling process. Multi-view learning has the potential to accurately predict short-term traffic.

Finally, on the basis of the MVL-STKNN/CP model, we further introduce the spatiotemporal trend matrix to form a complete MVL-STKNN/CPT model. Based on the introduction of space–time weightings, we simultaneously consider spatiotemporal closeness, the period, and the trend to form a multi-view learning method. It can be seen that the predictive accuracy of MVL-STKNN/CP is further improved. This demonstrates the validity of the spatiotemporal trend.

In addition, from comparing the improved models, comprising STKNN-W, STKNN-W-C, MVL-STKNN/CP, and MVL-STKNN/CPT, the introduction of periodicity induces the largest improvement in prediction accuracy, which induces the smallest improvement. Since the trend has a greater time span than periodicity, it is more difficult to capture the changing characteristics of traffic patterns.

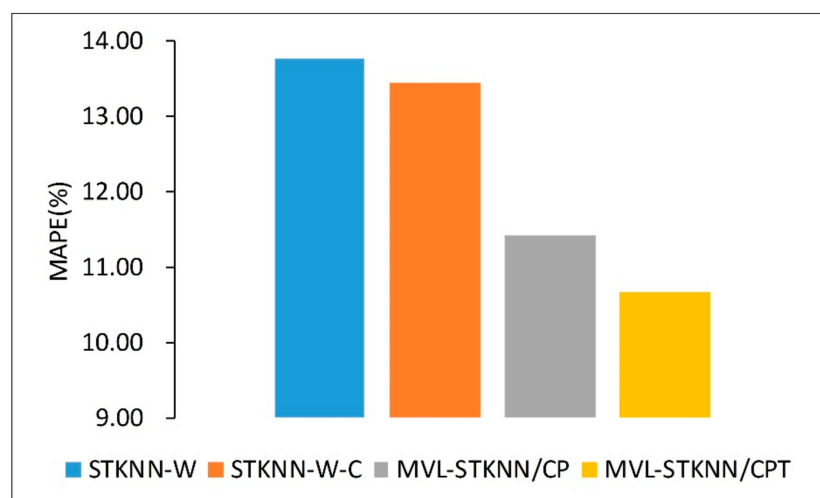


Figure 10. Impact of spatial and temporal dependencies.

4. Discussion

Accurate and robust short-term vehicular speed forecasting is a critical issue in superior transportation systems and traffic related applications. The existing ST-KNN model uses distance functions and correlation coefficients to identify spatial neighbors and measures the temporal interaction by only considering the temporal closeness of traffic, which means that existing ST-KNNs cannot fully reflect the essential features of road traffic. We address the shortcomings of such models by introducing an improved spatiotemporal KNN model integrated with multi-view learning. The proposed model greatly improves the performance of short-term traffic forecasting. However, there are still some issues that require further investigation.

Considering the spatial heterogeneity of city traffic, traffic patterns of different road segments show significant differences [46,47], which is reflected in different model structures including different spatial neighbors and spatiotemporal parameters. In the MVL-STKNN model, we have constructed an adaptive spatial neighbor for each road segment. However, in order to simplify the model (for example, in the modeling process, there is no need to adjust the parameters of each road segment), we adopt the mechanism of sharing parameters, which need further refinement in future work.

Yet, we consider the temporal proximity, periodicity, and trend to characterize the temporal non-stationarity of different road segments in the MVL-STKNN model. However, from the perspective of the more fine-grained variation characteristics of temporal non-stationarity, different road segments may have different fluctuation patterns in different time periods such as morning and evening peaks. Therefore, a very promising approach is to identify the traffic pattern of the entire road network through a clustering algorithm and then divide the time period for each traffic pattern. Next, the MVL-STKNN model is constructed for each time period in each cluster to further improve the accuracy and robustness of short-term traffic prediction.

Finally, we only validated the MVL-STKNN model with a small dataset. For example, 30 representative road segments and five days of data in the Beijing dataset were selected to test the performance of the proposed model. As mentioned above, different road segments show different traffic patterns at different time periods in order to use more road segments and a longer prediction period to comprehensively evaluate the accuracy and robustness of the MVL-STKNN model.

5. Conclusions

In this paper, we propose the MVL-STKNN model for short-term vehicular speed prediction. First, considering the heterogeneous nature of traffic, different road segments have different numbers of spatial neighbors. Therefore, we use cross-correlation to automatically determine the spatial neighbors of each road segment, which overcame the problem that existing methods can not automatically capture spatial information. Afterward, we consider temporal closeness, periodicity, and trend. This is further combined with an adaptive spatial neighbor to form three characterizations of traffic conditions including matrices for spatiotemporal closeness, period, and trend, which overcomes the problem that existing methods only use a global, fixed spatiotemporal closeness matrix to describe the traffic condition. Second, we introduce space-time weighting to improve the distance function in the ST-KNN model, which is then used to mine similar traffic patterns and obtain prediction results for the three spatiotemporal views, respectively. Finally, a multi-view learning method based on the neural network model is constructed from these three predicted spatiotemporal views with different weightings allocated to these views to obtain prediction values for traffic conditions.

In the experimental section, we employ two widely used traffic data sets comprised of floating car-speed data collected from the Beijing road network and from the California Freeway and Expressway systems to verify the efficiency of the proposed model. We first conducted a comparison with four existing baseline methods including HA, Elman-NN, KNN, and ST-KNN models. Compared the existing methods, the MVL-STKNN model decreases the MAPE index anywhere from 28.24% to 46.86% in the Beijing dataset and anywhere from 53.80% to 90.29% in the PeMS dataset. This demonstrates the efficiency of the multi-view learning method. Second, we further explore the influence of the components of

the MVL-STKNN model on predictive accuracy. For the space–time weighting matrix, the problem of parameter calibration of the existing ST-KNN model is solved by using space–time weighting instead of Gaussian weighting in the distance function. For spatiotemporal dependencies, the introduction of the adaptive spatial neighbor, temporal periodicity, and the trend in road traffic significantly improves the performance of short-term traffic forecasting, which demonstrates that the multi-view learning approach merits further attention in traffic-related data mining under such a dynamic and data-intensive environment. This needs to be considered with regard to the spatial correlation and heterogeneity as well as temporal closeness, periodicity, and trend in road traffic.

For the direction of future work, the following problems need to be investigated: (1) Further validation of the proposed model using more road segments and longer prediction periods, (2) comprehensive comparison of other spatiotemporal modeling methods such as ST-ARIMA and ST-ANN, and (3) incorporating external factors into the modelling process such as weather conditions and traffic accidents, and integrating these into the MVL-STKNN to achieve a more robust model that would further improve short-term traffic forecasting.

Author Contributions: S.C. and F.L. conceived the idea for the research and wrote the paper. S.C. implemented the ST-2SMR model and carried out the experimental validation. F.L. interpreted the results. P.P. and S.W. made important comments and suggestions for this paper.

Funding: This research was funded by the Key Research Program of the Chinese Academy of Sciences, grant number ZDRW-ZS-2016-6-3 and the State Key Research Development Program of China, grant number 2016YFB0502104. Their supports are gratefully acknowledged.

Acknowledgments: We also thank the anonymous referees for their helpful comments and suggestions.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Xia, J.; Chen, M.; Huang, W. A multistep corridor travel-time prediction method using presence-type vehicle detector data. *J. Intell. Transp. Syst. Technol. Plan. Oper.* **2011**, *15*, 104–113. [[CrossRef](#)]
2. Vlahogianni, E.I.; Golias, J.C.; Karlaftis, M.G. Short-term traffic forecasting: Overview of objectives and methods. *Transp. Rev.* **2004**, *24*, 533–557. [[CrossRef](#)]
3. Van Lint, J.W.C.; Van Hinsbergen, C. Short-term traffic and travel time prediction models. *Artif. Intell. Appl. Crit. Transp. Issues* **2012**, *22*, 22–41.
4. Levin, M.; Tsao, Y.-D. On forecasting freeway occupancies and volumes. *Transp. Res. Rec.* **1980**, *773*, 47–49.
5. Guo, J.; Williams, B.M. Real-Time Short-Term Traffic Speed Level Forecasting and Uncertainty Quantification Using Layered Kalman Filters. *Transp. Res. Rec. J. Transp. Res. Board* **2010**, *2175*, 28–37. [[CrossRef](#)]
6. Guo, J.; Huang, W.; Williams, B.M. Adaptive Kalman filter approach for stochastic short-term traffic flow rate prediction and uncertainty quantification. *Transp. Res. Part C Emerg. Technol.* **2014**, *43*, 50–64. [[CrossRef](#)]
7. Smith, B.L.; Williams, B.M.; Keith Oswald, R. Comparison of parametric and nonparametric models for traffic flow forecasting. *Transp. Res. Part C* **2002**, *10*, 303–321. [[CrossRef](#)]
8. Clark, S. Traffic Prediction Using Multivariate Nonparametric Regression. *J. Transp. Eng.* **2003**, *129*, 161–168. [[CrossRef](#)]
9. Zheng, Z.; Su, D. Short-term traffic volume forecasting: A k-nearest neighbor approach enhanced by constrained linearly sewing principle component algorithm. *Transp. Res. Part C Emerg. Technol.* **2014**, *43*, 143–157. [[CrossRef](#)]
10. Habtemichael, F.G.; Cetin, M. Short-term traffic flow rate forecasting based on identifying similar traffic patterns. *Transp. Res. Part C Emerg. Technol.* **2016**, *66*, 61–78. [[CrossRef](#)]
11. Khosravi, A.; Mazloumi, E.; Nahavandi, S.; Creighton, D.; Van Lint, J.W.C. A genetic algorithm-based method for improving quality of travel time prediction intervals. *Transp. Res. Part C Emerg. Technol.* **2011**, *19*, 1364–1376. [[CrossRef](#)]
12. Park, J.; Li, D.; Murphey, Y.L.; Kristinsson, J.; McGee, R.; Kuang, M.; Phillips, T. Real time vehicle speed prediction using a Neural Network Traffic Model. In Proceedings of the International Joint Conference on Neural Networks, San Jose, CA, USA, 31 July–5 August 2011; pp. 2991–2996. [[CrossRef](#)]
13. Koesdwiady, A.; Soua, R.; Karray, F. Improving Traffic Flow Prediction with Weather Information in Connected Cars: A Deep Learning Approach. *IEEE Trans. Veh. Technol.* **2016**, *65*, 9508–9517. [[CrossRef](#)]

14. Ma, X.; Dai, Z.; He, Z.; Ma, J.; Wang, Y.Y.; Wang, Y.Y. Learning traffic as images: A deep convolutional neural network for large-scale transportation network speed prediction. *Sensors* **2017**, *17*, 1–16. [[CrossRef](#)] [[PubMed](#)]
15. Vlahogianni, E.I.; Karlaftis, M.G.; Golias, J.C. Spatio-temporal short-term urban traffic volume forecasting using genetically optimized modular networks. *Comput. Civ. Infrastruct. Eng.* **2007**, *22*, 317–325. [[CrossRef](#)]
16. Vlahogianni, E.I.; Karlaftis, M.G.; Golias, J.C. Short-term traffic forecasting: Where we are and where we're going. *Transp. Res. Part C Emerg. Technol.* **2014**, *43*, 3–19. [[CrossRef](#)]
17. Zou, Y.; Zhu, X.; Zhang, Y.; Zeng, X. A space-time diurnal method for short-term freeway travel time prediction. *Transp. Res. Part C Emerg. Technol.* **2014**, *43*, 33–49. [[CrossRef](#)]
18. Asif, M.T.; Dauwels, J.; Goh, C.Y.; Oran, A.; Fathi, E.; Xu, M.; Dhanya, M.M.; Mitrovic, N.; Jaillet, P. Spatiotemporal patterns in large-scale traffic speed prediction. *IEEE Trans. Intell. Transp. Syst.* **2014**, *15*, 794–804. [[CrossRef](#)]
19. Ermagun, A.; Levinson, D. Spatiotemporal Traffic Forecasting: Review and Proposed Directions. *Transp. Rev.* **2017**, 1–29. [[CrossRef](#)]
20. Kamarianakis, Y.; Prastacos, P. Spatial Time-Series Modeling: A review of the proposed methodologies. Available online: <https://ideas.repec.org/p/crt/wpaper/0604.html> (accessed on 14 June 2018).
21. Min, X.; Hu, J.; Chen, Q.; Zhang, T.; Zhang, Y.Y. Short-term traffic flow forecasting of urban network based on dynamic STARIMA model. In Proceedings of the IEEE Conference on Intelligent Transportation Systems, St. Louis, MO, USA, 4–7 October 2009; Volume 100084, pp. 461–466. [[CrossRef](#)]
22. Min, X.; Hu, J.; Zhang, Z. Urban traffic network modeling and short-term traffic flow forecasting based on GSTARIMA model. In Proceedings of the IEEE Conference on Intelligent Transportation Systems, Funchal, Portugal, 19–22 September 2010; pp. 1535–1540. [[CrossRef](#)]
23. Yu, H.; Wu, Z.; Wang, S.; Wang, Y.; Ma, X. Spatiotemporal recurrent convolutional networks for traffic prediction in transportation networks. *Sensors* **2017**, *17*, 1501. [[CrossRef](#)] [[PubMed](#)]
24. Wu, S.; Yang, Z.; Zhu, X.; Yu, B. Improved k-nn for Short-Term Traffic Forecasting Using Temporal and Spatial Information. *J. Transp. Eng.* **2014**, *140*, 1–9. [[CrossRef](#)]
25. Yu, B.; Song, X.; Guan, F.; Yang, Z.; Yao, B. k-Nearest Neighbor Model for Multiple-Time-Step Prediction of Short-Term Traffic Condition. *J. Transp. Eng.* **2016**, *142*, 4016018. [[CrossRef](#)]
26. Xia, D.; Wang, B.; Li, H.; Li, Y.; Zhang, Z. A distributed spatial-temporal weighted model on MapReduce for short-term traffic flow forecasting. *Neurocomputing* **2016**, *179*, 246–263. [[CrossRef](#)]
27. Cai, P.; Wang, Y.; Lu, G.; Chen, P.; Ding, C.; Sun, J. A spatiotemporal correlative k-nearest neighbor model for short-term traffic multistep forecasting. *Transp. Res. Part C Emerg. Technol.* **2016**, *62*, 21–34. [[CrossRef](#)]
28. Hong, H.; Huang, W.; Zhou, X.; Du, S.; Bian, K.; Xie, K. Short-term traffic flow forecasting: Multi-metric KNN with related station discovery. In Proceedings of the 12th International Conference on Fuzzy Systems and Knowledge Discovery, Zhangjiajie, China, 15–17 August 2015; pp. 1670–1675.
29. Yue, Y.; Yeh, A.G.O. Spatiotemporal traffic-flow dependency and short-term traffic forecasting. *Environ. Plan. B Plan. Des.* **2008**, *35*, 762–771. [[CrossRef](#)]
30. Cheng, T.; Haworth, J.; Wang, J. Spatio-temporal autocorrelation of road network data. *J. Geogr. Syst.* **2012**, *14*, 389–413. [[CrossRef](#)]
31. Cheng, T.; Wang, J.; Haworth, J.; Heydecker, B.; Chow, A. A Dynamic Spatial Weight Matrix and Localized Space-Time Autoregressive Integrated Moving Average for Network Modeling. *Geogr. Anal.* **2014**, *46*, 75–97. [[CrossRef](#)]
32. Zhang, J.; Zheng, Y.; Qi, D.; Li, R.; Yi, X. DNN-Based Prediction Model for Spatio-Temporal Data. In Proceedings of the 24th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems, Burlingame, CA, USA, 31 October–3 November 2016; p. 92.
33. Zhang, J.; Zheng, Y.; Qi, D. Deep Spatio-Temporal Residual Networks for Citywide Crowd Flows Prediction. In Proceedings of the AAAI, San Francisco, CA, USA, 4–9 February 2017; pp. 1655–1661.
34. Zheng, Y.; Yi, X.; Li, M.; Li, R.; Shan, Z.; Chang, E.; Li, T. Forecasting Fine-Grained Air Quality Based on Big Data. In Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD '15), Sydney, Australia, 10–13 August 2015; pp. 2267–2276.
35. Liu, Y.; Zheng, Y.; Liang, Y.; Liu, S.; Rosenblum, D.S. Urban water quality prediction based on multi-task multi-view learning. In Proceedings of the International Joint Conference on Artificial Intelligence Google Scholar, New York, NY, USA, 9–15 July 2016.

36. Yi, X.; Zheng, Y.; Zhang, J.; Li, T. ST-MVL: Filling Missing Values in Geo-Sensory Time Series Data Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences. In Proceedings of the 25th International Joint Conference on Artificial Intelligence (IJCAI 2016), New York, NY, USA, 9–15 July 2016; pp. 2704–2710.
37. Ma, X.; Yu, H. Large-scale transportation network congestion evolution prediction using deep learning theory. *PLoS ONE* **2015**, *10*, e0119044. [[CrossRef](#)] [[PubMed](#)]
38. Box, G.E.P.; Jenkins, G.M.; Reinsel, G.C. *Time Series Analysis: Forecasting and Control*; John Wiley & Sons: Hoboken, NJ, USA, 2013; ISBN 1118619064.
39. Huang, W.; Song, G.; Hong, H.; Xie, K. Deep Architecture for Traffic Flow Prediction: Deep Belief Networks With Multitask Learning. *IEEE Trans. Intell. Transp. Syst.* **2014**, *15*, 2191–2201. [[CrossRef](#)]
40. Duan, Y.; Lv, Y.; Kang, W.; Zhao, Y. A deep learning based approach for traffic data imputation. In Proceedings of the 17th International IEEE Conference on Intelligent Transportation Systems (ITSC), Qingdao, China, 8–11 October 2014; pp. 912–917.
41. Fouladgar, M.; Parchami, M.; Elmasri, R.; Ghaderi, A. Scalable Deep Traffic Flow Neural Networks for Urban Traffic Congestion Prediction. In Proceedings of the International Joint Conference on Neural Networks (IJCNN), Anchorage, AK, USA, 14–19 May 2017. [[CrossRef](#)]
42. Cheng, S.; Lu, F. A Two-Step Method for Missing Spatio-Temporal Data Reconstruction. *ISPRS Int. J. Geo-Inf.* **2017**, *6*, 187. [[CrossRef](#)]
43. Dang, X.-C.; Hao, Z.-J. Prediction for network traffic based on modified Elman neural network. *J. Comput. Appl.* **2010**, *30*, 2648–2652. [[CrossRef](#)]
44. Li, S.; Shen, Z.; Xiong, G. A k-nearest neighbor locally weighted regression method for short-term traffic flow forecasting. In Proceedings of the 15th International IEEE Conference on Intelligent Transportation Systems, Anchorage, AK, USA, 16–19 September 2012; pp. 1596–1601.
45. Xiaoyu, H.; Yisheng, W.; Siyu, H.; Hou, X.; Wang, Y.; Hu, S. Short-term Traffic Flow Forecasting based on Two-tier K-nearest Neighbor Algorithm. *Procedia Soc. Behav. Sci.* **2013**, *96*, 2529–2536. [[CrossRef](#)]
46. Gao, P.; Liu, Z.; Tian, K.; Liu, G. Characterizing Traffic Conditions from the Perspective of Spatial-Temporal Heterogeneity. *ISPRS Int. J. Geo-Inf.* **2016**, *5*, 34. [[CrossRef](#)]
47. Liu, K.; Gao, S.; Qiu, P.; Liu, X.; Yan, B.; Lu, F. Road2Vec: Measuring Traffic Interactions in Urban Road System from Massive Travel Routes. *ISPRS Int. J. Geo-Inf.* **2017**, *6*, 321. [[CrossRef](#)]



© 2018 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).