WILEY | Hindawi

Research Article

A Combined Deep Learning Method with Attention-Based LSTM Model for Short-Term Traffic Speed Forecasting

Pan Wu ⓘ, **Zilin Huang** ⓘ, **Yuzhuang Pian** ⓘ, **Lunhui Xu** ⓘ, **Jinlong Li** ⓘ, **and Kaixun Chen** ⓘ

School of Civil Engineering and Transportation, South China University of Technology, Guangzhou 510641, Guangdong, China

Correspondence should be addressed to Zilin Huang; ctjulian@mail.scut.edu.cn

Short-term traffic speed prediction is a promising research topic in intelligent transportation systems (ITSs), which also plays an important role in the real-time decision-making of traffic control and guidance systems. However, the urban traffic speed has strong temporal, spatial correlation and the characteristic of complex nonlinearity and randomness, which makes it challenging to accurately and efficiently forecast short-term traffic speeds. We investigate the relevant literature and found that although most methods can achieve good prediction performance with the complete sample data, when there is a certain missing rate in the database, it is difficult to maintain accuracy with these methods. Recent studies have shown that deep learning methods, especially long short-term memory (LSTM) models, have good results in short-term traffic flow prediction. Furthermore, the attention mechanism can properly assign weights to distinguish the importance of traffic time sequences, thereby further improving the computational efficiency of the prediction model. Therefore, we propose a framework for short-term traffic speed prediction, including data preprocessing module and short-term traffic prediction module. In the data preprocessing module, the missing traffic data are repaired to provide a complete dataset for subsequent prediction. In the prediction module, a combined deep learning method that is an attention-based LSTM (ATT-LSTM) model for predicting short-term traffic speed on urban roads is proposed. The proposed framework was applied to the urban road network in Nanshan District, Shenzhen, Guangdong Province, China, with a 30-day traffic speed dataset (floating car data) used as the experimental sample. Results show that the proposed method outperforms other deep learning algorithms (such as recurrent neural network (RNN) and convolutional neural network (CNN)) in terms of both calculating efficiency and prediction accuracy. The attention mechanism can significantly reduce the error of the LSTM model (up to 12.4%) and improves the prediction performance.

1. Introduction

Future short-term traffic speed information is critical for alleviating traffic congestion, predicting traffic incidents, organizing traffic travel, and controlling traffic [1, 2]. More importantly, it can promote intelligent transportation systems (ITSs) to make smarter decisions, effectively reduce traffic risks, and make the transportation system more intelligent and efficient. Therefore, short-term traffic speed prediction has become a hot topic in ITS and has also attracted numerous traffic practitioners and scholars to conduct deeper research. However, the traffic data imply spatiotemporal correlation and intricate periodicity and show strong chaos and randomness. This brings great difficulty in accurately predicting short-term traffic speeds. Finding a more efficient and accurate prediction method that can easily capture latent features of traffic data is still a challenging problem to be solved.

The traffic prediction methods proposed in early research are mainly divided into three categories: parameter-based methods, nonparameter-based methods, and hybrid methods. Parameter-based methods mainly include the time series method and the Kalman filter (KF) method [3, 4]. Prediction methods based on time series mainly focus on the automatic regression moving average (ARIMA) model and the improved variations in this model. Nonparametric methods mainly include the *K*-nearest neighbor (KNN) method, support vector regression (SVR) method, artificial

neural network (ANN) model, and other methods. Hybrid methods are mostly a combination of two or three methods. However, because of the effects of uncertain factors such as weather, the implicit correlation of traffic data captured by the above approaches is limited. These methods still can be improved in terms of prediction accuracy and generalization ability.

In recent years, with the rapid improvement in computer capabilities, many prediction methods based on deep learning algorithms have emerged. With good performance in other fields, many deep learning methods (such as convolutional neural network (CNN) models [5], recurrent neural network (RNN) models, and long short-term memory (LSTM) models [6–9]) have been introduced to predict short-term traffic flow and have achieved better prediction performance than traditional forecasting methods. In addition, the combined model often has a better predictive effect than the single model [10–13]. For example, Lu et al. [10] proposed a combined model of ARIMA and LSTM, and Zheng et al. proposed the Conv-LSTM model based on the attention mechanism [11]. Yu et al. [12] proposed a low-rank representation (LRR) and dynamic mode decomposition (DMD) combined model (LRDMD). Wu et al. [13] analyzed the prediction performance of combined RNN and CNN models. These methods can better compensate for the shortcomings of traditional methods in capturing the inherent temporal and spatial correlation of traffic data with good accuracy, which can handle incomplete data. Compared with the single method, although the calculation performance of these methods has improved significantly, there are still some weaknesses. Although accurate short-term traffic information can be obtained via these prediction methods, the training time of these models is too long, and it is prone to overfitting during training. Because of the intricate structure of traffic data, it is difficult to completely capture the inherent characteristics of the dataset. Furthermore, most of these studies on traffic prediction rarely focus on the imputation of missing data despite the fact that the accuracy of results is influenced by incomplete data to some extent.

Therefore, this study is devoted to propose an accurate and efficient prediction method for short-term traffic speed. Consulting the existing literature finds that the combined LSTM method shows outstanding performance in traffic prediction [14–16]. Moreover, LSTM, as a special form of RNN network, can solve well the impact of the RNN gradient disappearance on the accuracy of the prediction model. Consequently, in this paper, a hybrid deep learning method that combined attention mechanism and LSTM model (ATT-LSTM) is proposed for the prediction of short-term traffic speed, which can alleviate the loss, dilution, or coverage of the model details, thereby increasing the quality of decoding.

Finally, experimental data were collected from the urban road network. The contributions of this paper include the following three aspects:

(1) Different from the previous short-term traffic speed prediction methods, we design an entire forecasting framework of short-term traffic speed that is the combination of the data preprocessing module and ATT-LSTM prediction module, which can achieve high-prediction accuracy performance on the urban road network.

(2) To overcome the problem of missing data, we propose a new data preprocessing module that is composed of the naive Bayesian method and a dynamic time warping algorithm to handle raw dataset with a certain degree of missing and proved that the module can further improve the quality of the data and enlarge the data sample, ultimately providing a high-quality dataset for short-term traffic speed forecasting.

(3) To solve the problem that it is difficult to accurately predict the short-term traffic speed on the complex urban road network, we propose a speed prediction method which is especially suitable for traffic data characteristics, namely, the ATT-LSTM model. It uses the local attention vector calculation method to assign weights to traffic speed sequences and distinguish their importance. As a result, it effectively reduces the calculation for model training and improves the efficiency of the model.

The subsequent sections are arranged as follows. The next section discusses the related research. The third section is about problem description. The fourth section introduces the models and theories used in this research. The fifth section discusses the results of the case studies, verifies the prediction methods proposed in the study, and compares different methods in terms of prediction performance. The last section summarizes the conclusions and outlines further research.

2. Related Work

This section summarizes related research on short-term traffic prediction. As early as the 1980s, short-term traffic flow prediction had been an important topic in the research of ITS [17], and it has nearly 40 years of history. In early studies, statistical methods as main means were used to predict single traffic characteristics (such as traffic volume, speed, density, and travel time) at a special point [18]. Later, with the rapid progress in computer technology, many data-driven methods and intelligent algorithms based on empirical calculations (including neural networks and Bayesian networks, fuzzy algorithms, and evolutionary techniques) were represented. Recently, deep learning algorithms have prevailed in transportation, most of which are used to forecast short-term traffic flow with good results.

According to related literature [12], short-term traffic flow prediction methods are mainly divided into three categories: statistical learning methods, machine learning methods, and combined methods. The statistical method that has been proposed and applied for many years is to explore the implicit relationship between traffic time series through a statistical model, finding the optimal parameters of the fitting process using historical data. Typical methods mainly include the KF method and ARIMA, both of which are common linear time series models. In 1960, Kalman

proposed a linear prediction method called the KF method, which was widely used in predicting traffic flow [19–21]. Guo et al. [3] proposed an adaptive KF method, which can significantly improve the prediction performance of the original method. The ARIMA model is a well-known linear model and a popular parameter regression model [22]. However, it cannot accurately describe the randomness and nonlinear characteristics of traffic data. To increase its prediction performance, researchers have proposed many improved models such as SARIMA [23] and STARIMA [24].

The machine learning method is used to predict future traffic by training with historical traffic data. This method includes the genetic algorithm [25], KNN algorithm [26], ANN algorithm, BP neural network, support vector regression [27], LSTM [28], DNN model, and CNN model [29]. The hybrid method refers to a reasonable combination of machine learning methods and statistical methods. In recent years, with the extensive application of deep learning methods in traffic flow prediction, an increasing number of traffic researchers have been committed to proposing combined prediction models with excellent performance and efficiency. Initial results have been achieved, and many combined models have been proposed, such as CNN combined with LSTM [11, 13], the combined model of CNN and ARIMA [10], and other combined models [30].

Referring to the relevant literature mentioned above, it has been found that although the existing methods can be used to forecast short-term traffic flow, the prediction results are often affected by severe weather, sudden traffic accidents, and other uncertain factors. Therefore, the following are the limitations of these studies: (1) the traditional ARIMA algorithm cannot accurately track changes in traffic flow conditions under emergencies, which limits the extensive application of the algorithm. KF often has a residual error, which leads to a sharp drop in prediction accuracy. Effectively solving the residual problem is the key to improving the performance of the KF algorithm. (2) The traffic prediction algorithm based on machine learning is usually too dependent on the training data. Once a dataset with poor quality is encountered, the training time will become uncontrollable. In addition, overfitting reduces the prediction accuracy. (3) The popular deep learning algorithm for short-term traffic prediction also has the problem of data dependence. Even though the prediction accuracy is higher than most other algorithms, the computational efficiency of the multilayer structure needs to be improved. LSTM has attracted significant attention in deep learning algorithms because of its good generalization and lack of gradient vanishing problems. The attention mechanism can distinguish the importance of time series data by allocating weights. Therefore, this paper proposes a deep learning algorithm that combines the LSTM and attention model for short-term traffic speed prediction.

3. Problem Description

As the typical time series, traffic data also have the general characteristics of a nonlinear time series, which is reflected in nonstationarity, periodic distribution of traffic parameters and spatiotemporal correlation. Some recent studies indicate that the traffic time series exhibits stochasticity and uncertainty at different time periods [8, 10, 11, 31].

The main purpose of predicting short-term traffic speed is to provide the accurate traffic speed in the next five minutes, ten minutes, or fifteen minutes and to provide support for improving the operational efficiency of urban roads. V_τ^n is defined as the traffic velocity of the n-th observation location during the τ th time interval. And the n-th observation location refers to the road section designated as n. At current time t, the main task is to predict the traffic speed at points of interest (POI) for a certain prediction range δ in the prediction time interval $(t + d\delta)$ (for some prediction horizon δ given the historical traffic speed sequence of observation locations $\{V_\tau^n\}$), where $\tau = t - r\delta, \ldots, t - \delta, t, n \in N$, in which N is the set of n observation points in the road network. In this work, we consider $\delta = 5$ minutes and $d = 1$, 2, 3, which means the historical data are used to predict the traffic speed of the next 5, 10, and 15 minutes. To simplify the description, we use t-r represents $t - r\delta$ below.

Traffic data usually show strong space-time correlation and periodic characteristics; that is, the traffic speed data may be affected by the traffic speed of the adjacent POI observation position and the traffic speed at the previous moment. In 1990, Hoffman and Janko [32] proposed a historical trend model, which assumes that, within a day of the same historical trend, traffic has similar operating characteristics during the same time period. In other words, changes in the traffic speed on the same day for several consecutive weeks are similar, and the traffic speed shows a daily cycle pattern and a weekly cycle pattern. In this study, a deep learning model is proposed to use the temporal and spatial characteristics of traffic speed and periodically predict the future short-term traffic speed. By consulting related literature, it has been found that storing traffic speed data in matrix form can better exploit the temporal and spatial relationships and periodicity between the data for short-term traffic speed forecasting [11, 12]. Therefore, in this study, we have stored the traffic speed data in matrix form. If v_t^n is the traffic velocity of the n-th observation location at time t, then the historical traffic velocity of the n-th observation location from time t-r to t can be expressed as $V_t^n = [v_{t-r}^n, v_{t-(r-1)}^n, \ldots, v_t^n]^T$. The historical traffic velocity data of adjacent observation points (a total of n observation points) are combined to form a spatiotemporal correlation matrix:

$$\mathbf{V}_t^N = \begin{bmatrix} V_{t-r}^N \\ V_{t-(r-1)}^N \\ \vdots \\ V_t^N \end{bmatrix}^T = \begin{bmatrix} v_{t-r}^1 & v_{t-(r-1)}^1 & \cdots & v_t^1 \\ v_{t-r}^2 & v_{t-(r-1)}^2 & \cdots & v_t^2 \\ \vdots & \vdots & \ddots & \vdots \\ v_{t-r}^n & v_{t-(r-1)}^n & \cdots & v_t^n \end{bmatrix}, \quad (1)$$

where $V_t^N = [v_t^1, v_t^2, \ldots, v_t^n]$ indicates the traffic velocity of the prediction area at time t. Considering the periodic characteristics of traffic speed, the daily periodic traffic speed matrix and the weekly periodic traffic speed matrix are constructed as follows:

$$\mathbf{V}_{t^d}^{N} = \begin{bmatrix} V_{t^d-r}^{N} \\ V_{t^d-(r-1)}^{N} \\ \vdots \\ V_{t^d}^{N} \end{bmatrix}^{T} = \begin{bmatrix} v_{t^d-r}^{1} & v_{t^d-(r-1)}^{1} & \cdots & v_{t^d}^{1} \\ v_{t^d-r}^{2} & v_{t^d-(r-1)}^{2} & \cdots & v_{t^d}^{2} \\ \vdots & \vdots & \ddots & \vdots \\ v_{t^d-r}^{n} & v_{t^d-(r-1)}^{n} & \cdots & v_{t^d}^{n} \end{bmatrix},$$

$$\mathbf{V}_{t^w}^{N} = \begin{bmatrix} V_{t^w-r}^{N} \\ V_{t^w-(r-1)}^{N} \\ \vdots \\ V_{t^w}^{N} \end{bmatrix}^{T} = \begin{bmatrix} v_{t^w-r}^{1} & v_{t^w-(r-1)}^{1} & \cdots & v_{t^w}^{1} \\ v_{t^w-r}^{2} & v_{t^w-(r-1)}^{2} & \cdots & v_{t^w}^{2} \\ \vdots & \vdots & \ddots & \vdots \\ v_{t^w-r}^{n} & v_{t^w-(r-1)}^{n} & \cdots & v_{t^w}^{n} \end{bmatrix},$$

(2)

where t^d represents the same time as the time t of the last day and t^w is the same time and space as the time t of the last week.

4. Methodology

4.1. Analysis and Preprocessing of Traffic Speed Data. The urban road network traffic speed data have strong temporal and spatial correlation and periodicity and are greatly affected by external factors. This section analyzes the distribution characteristics of traffic speed data and proposes data preprocessing modules for missing data.

4.1.1. Distribution Characteristics of Traffic Data in Time and Spatial Dimension. Taking the traffic speed data of a weekday (May 10, 2017, Wednesday) and a weekend (May 20, 2017, Saturday) on a road network (including the expressway, arterial road, secondary road, and branch road) as an example, we analyze the distribution characteristics of traffic speed data in the time dimension. Dataset is processed separately and divided according to the 1-hour interval, and the average coverage intensity of different grades is obtained by statistics. The result is shown in Figure 1.

The coverage intensity is based on time h^{-1}. As can be seen in addition to changes in coverage intensity over time, there are significant differences in coverage strength at different road levels. The coverage intensity of express roads during peak hours on working days reaches more than 800 times h^{-1}, while the average coverage intensity of branch roads on working days does not exceed 200 times h^{-1} and the average coverage intensity of nonworking day does not exceed 250 times h^{-1}. The main reasons for the low-coverage intensity are the large number of road sections, the wide range of roads, and the combined effect of the travel willingness and the driving range of the floating vehicles. Therefore, high-grade roads have a large traffic volume and high coverage of floating car data, and the reliability of floating car data to estimate the average traffic speed of the road segment is higher than low-grade roads.

Taking the data from 7 am to 9 am on May 10, 2017, to analyze the spatial distribution characteristics of the road network in the study area, we match the data to the map and draw the distribution map of coverage frequency on the road network. Figure 2 uses color as a distinction to show the difference in the coverage frequency of traffic flow data over a long period of 2 hours.

Because this article uses 5 minutes as the sampling interval, the coverage frequency is up to 24 times within 2 hours. The coverage frequency is divided into 5 levels from 0 to 24 times. The thickness of the road section is from thin to thick, and the color of the road section is from green to red to indicate the coverage frequency from less to more. It can be seen that the coverage frequency is more than 20 times mostly on high-grade roads. Compared to the high-grade roads, the coverage frequency of secondary roads decreased significantly. And the coverage frequency of branch roads was still significantly lower than that of secondary roads or even missing. The above phenomenon shows that the uneven distribution of floating car data on different grades of roads is very obvious. On this basis, the 40th time period is taken as the sample for the same analysis. The results are shown in Figure 3. The solid line indicates that the current road section has complete data, and the dotted line indicates that the current road section data are missing. It can be seen that low-grade roads are much more likely to have missing data than high-grade roads. The missing data need to be repaired in advance for prediction of traffic speed.

4.1.2. Data Preprocessing. If the characteristics of traffic flow are regarded as signals that change over time, they are likely to be disturbed by noise signals, thereby masking the actual trend of traffic flow. Referring to the related literature [31] using wavelet transform to decompose the traffic time series into two frequency signals, the low-frequency series is named as trend signals, and the noise series is considered as residual signals. As shown in Figure 4, the trend signal exhibits sufficiently clear periodic characteristics, preserving the basic trend of the traffic flow and constituting a stable part of the traffic flow. The residual signal does not show obvious periodic and frequent changes. Furthermore, the traffic flow is a nonstationary series, which may be affected by road structure, traffic demand, and weather conditions.

After the wavelet transform, it is easy to pay attention to the average characteristics of trend signals. As shown in Figure 5, the average value of trend signals for all working days over multiple weeks is very consistent from 7:00 to 24:00. Their inflection points are roughly the same. On this basis, for the incomplete dataset, the imputation method proposed in [31, 33] is used to repair missing data in the traffic speed sample dataset, so as to provide complete data for subsequent forecasting research.

4.2. Overview of Proposed Model. Hochreiter and Schmidhuber proposed the LSTM model [34], which is a special form of an RNN, specializing in natural language processing at its initial stage. It can effectively solve the problem of gradient disappearance and the long-term dependence of learning in RNNs. Subsequently, the model has been widely used in the analysis of time series datasets and has good performance in traffic flow prediction [8]. Therefore, this study uses the LSTM neural network to study the prediction of short-term traffic speed on an urban road network, merging the attention mechanism to optimize the model structure, alleviating the loss, dilution, and coverage of
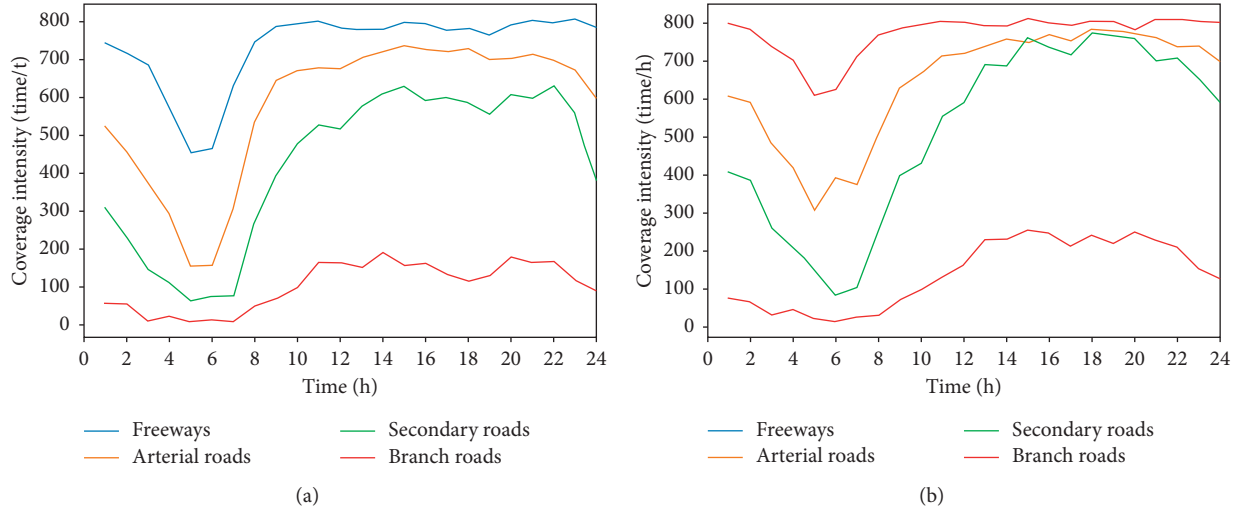
FIGURE 1: Changes in average coverage intensity within 24 h via grade roads: (a) weekday; (b) off day.

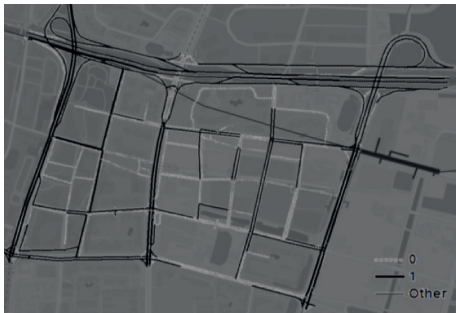FIGURE 2: Spatial coverage of floating vehicle data within 2 h.

FIGURE 3: Space coverage of floating car data at the 40th interval.

model details, increasing the decoding accuracy, and finally, building an attention-based LSTM prediction model. The local attention mechanism has been selected to calculate the attention vector in the variant, which improves the efficiency of the model.

4.2.1. LSTM Network for Short-Term Traffic Speed Forecasting. The structure of the LSTM is shown in Figure 6. And taking the traffic flow speed of a certain observation point as an example, the working principle of the repeated module of LSTM is explained, where V_t represents the input

traffic flow speed at the current moment, h_t is the corresponding output speed at the current moment, V_{t-1} represents the input speed data at the previous moment, h_{t-1} is the corresponding output speed, V_{t+1} is the input traffic speed at the next moment, and h_{t+1} is the output speed corresponding to the next moment. The flow chart of the short-term traffic speed prediction algorithm based on the LSTM is shown in Figure 7. First, a series of matrices and vectors were initialized to save the model parameters and intermediate calculation results. The purpose of this was to enable the neural network to learn effectively and obtain useful information during the training process.

4.2.2. Attention Mechanism. By imitating human thinking, different attention is allocated to the target, and features of different importance are matched. This study improves the classical attention mechanism, replacing the intermediate vector with a sequence of vectors, as shown in Figure 8. The model no longer needs to compress all the information into a fixed-dimensional vector, which greatly alleviates the problems of incomplete information representation and information dilution and coverage of the original model. When decoding, a subset of the vector can be selected for processing in the vector sequence. When the output is generated, the information conveyed by the input sequence can be fully utilized and interpreted.

After introducing the attention mechanism, each output is affected by the intermediate vector and the previous output, as follows:

$$Y_1 = f(C_1), Y_2 = f(C_2, Y_1), \ldots, Y_t = f(C_t, Y_{t-1}). \quad (3)$$

Among them, f represents a certain transformation function of the encoder to the input data and C_t is an extremely important parameter, which represents the probability distribution of attention distribution corresponding to different elements in the input sequence, which is called the attention vector. Generally, the variants in the attention mechanism are mainly carried out from two different
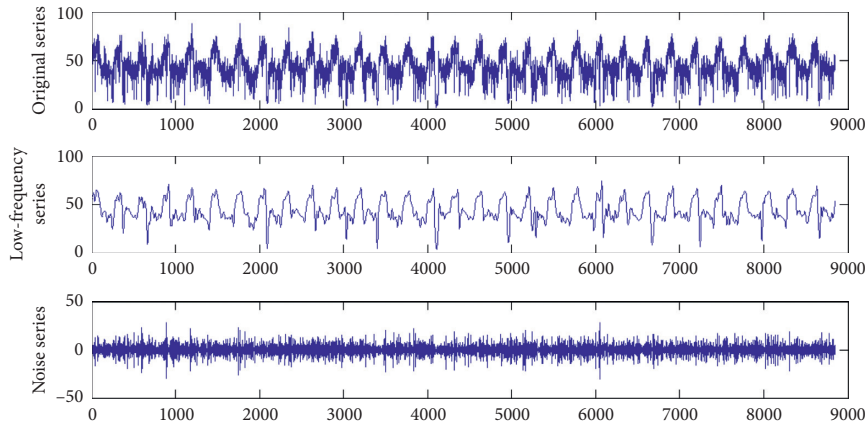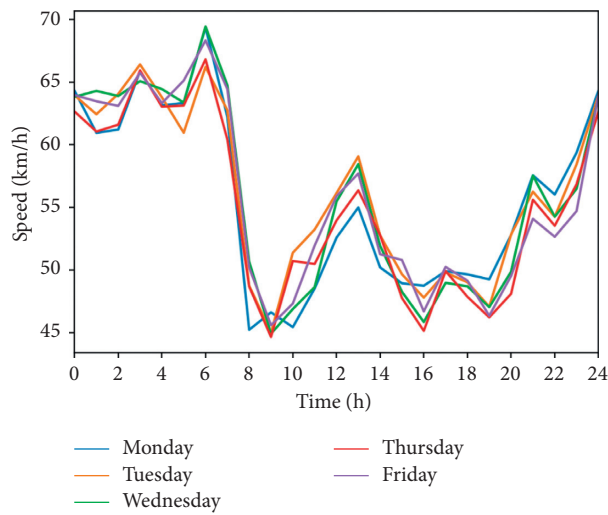
FIGURE 4: Signal after wavelet transform.

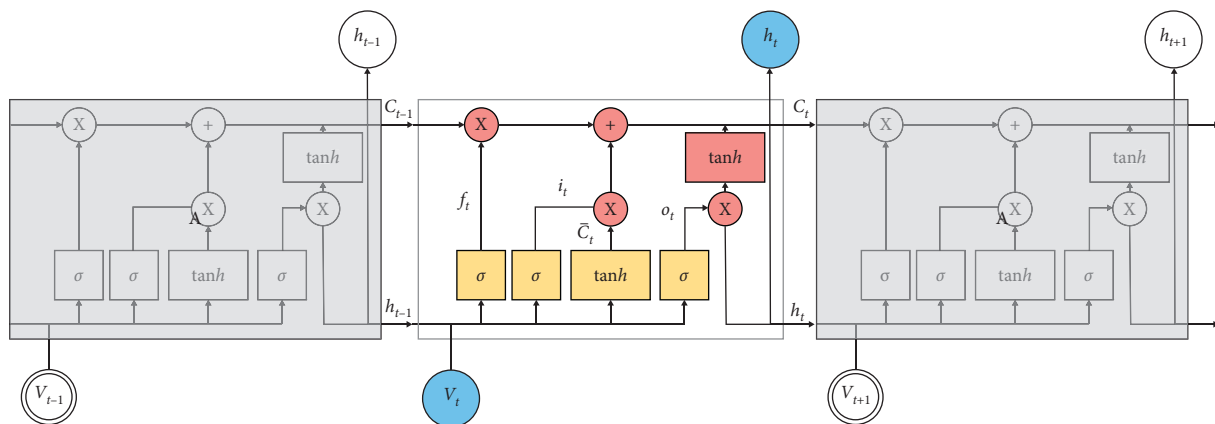FIGURE 5: Comparison of trend signals on weekdays.

FIGURE 6: Repeated modules in LSTM.

directions. The first is to study the variants in the calculation method based on the attention matching degree. The other is a variant of the weighted sum calculation method based on the attention vector. This article conducts an in-depth study on the second type of variant, and choosing local attention

for traffic flow prediction, compared to other variants, the calculation is smaller and the efficiency is higher. Specifically, the local attention method generates an alignment position p_t in the source sequence for the output at time t. Then, taking the window $[p_t - D, p_t + D]$ in the source
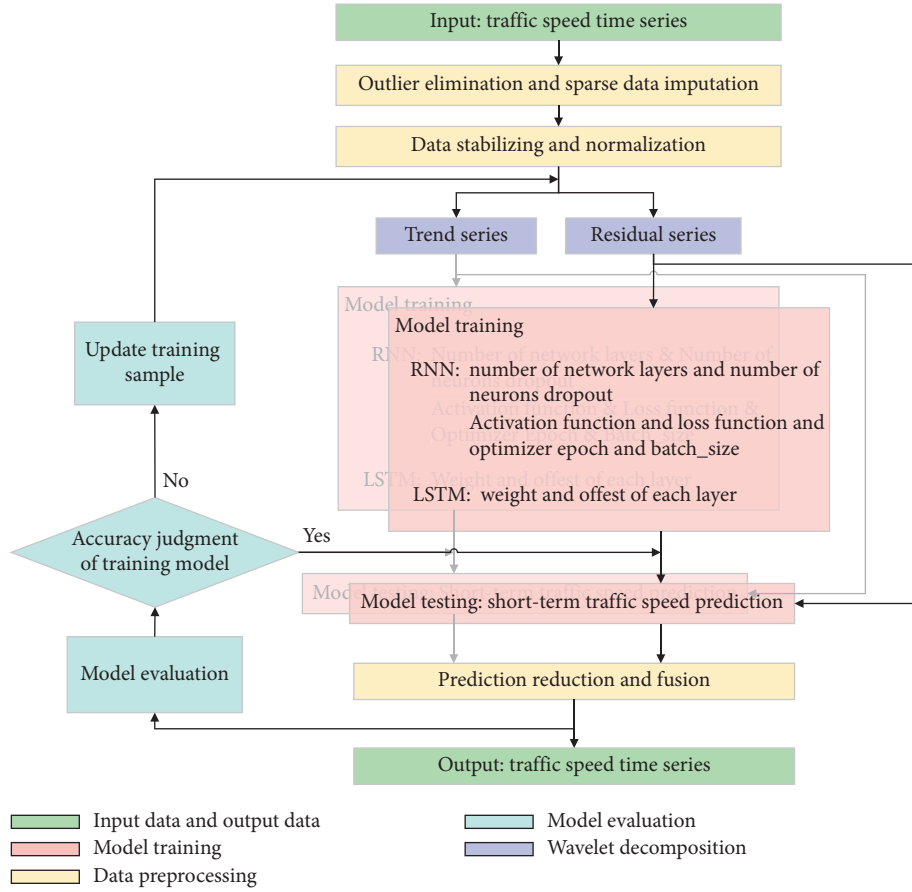
FIGURE 7: Algorithm flow chart of prediction model for short-term traffic velocity based on LSTM. "Epoch" represents the frequency of the data sample in rounds, and "Batch_size" refers to the optimal sample batch set to obtain a stable precision gradient descent.
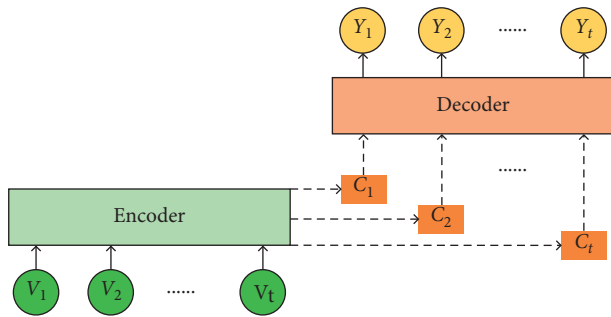

FIGURE 8: Schematic diagram of the attention model.

sequence, the intermediate vector C_t is obtained by calculating the weighted average of the hidden layer state in the window. When the range of the window exceeds the boundary of the source sequence, the boundary of the sequence shall prevail. Local attention finds p_t and calculates alpha in two ways.

The monotonic alignment (local-m) method assumes that the alignment position is $p_t = t$ (linear alignment), and then calculating the softmax inside the window, the alpha outside the window takes 0. The formula is as shown in formula (4), where the score () function in theory can be any comparison function, and dot product can also be used as a scoring function:

$$\alpha_{ti} = \frac{\exp\left(\text{score}\left(h_i, s_t\right)\right)}{\sum_i \exp\left(\text{score}\left(h_i, s_t\right)\right)}. \tag{4}$$

The predictive alignment (Local-p) method is to predict its alignment position in the source sequence for each target output, in other words, predicting p_t between [0, T] through a function. This article uses this method to find p_t and to calculate the alpha formula as follows:

$$p_t = T \cdot \text{sigmoid}\left(v_t^T \tan h\left(w_t s_t\right)\right), \tag{5}$$

where w_t and v_t are model parameters and T is the length of the source sequence. Then, we introduced a Gaussian distribution subject to $N(p_t, D/2)$ to set the alignment weight. The calculation formula for the alignment probability between the target position t and the source sequence position i is as follows:

$$\alpha_{ti} = \frac{\exp\left(\text{score}\left(h_i, s_t\right)\right)}{\sum_i \exp\left(\text{score}\left(h_i, s_t\right)\right)} \exp\left(-\frac{\left(i - p_t\right)^2}{2\sigma^2}\right). \tag{6}$$

4.2.3. Proposed Attention-Based LSTM (ATT-LSTM). The introduction of the attention mechanism is mainly to optimize the LSTM structure, that is, to add high-impact features to the sequence to compensate for the lack of

learning ability of the ultralong sequence. The structure of the ATT-LSTM model is shown in Figure 9. The ATT-LSTM is roughly divided into four layers: the hybrid input layer, hidden layer, attention mechanism layer, and output layer. Taking the traffic speed data of an observation location as an example, the input layer sequence is $V = (V_1, V_2, \ldots, V_t)$, and $h = (h_1, h_2, \ldots, h_t)$ is the hidden layer of the LSTM. Firstly, at the attention mechanism layer, the local attention mechanism is used to predict its alignment p_t of the output y_t in the input sequence. Then, in the input sequence selection window $[p_t - D, p_t + D]$, the output value s_{t-1} of the hidden layer node at $t - 1$ before the output Y_t is used to match the state of the hidden layer node corresponding to each element in the input sequence one by one. The function $F(h_i, s_{t-1})$ is used to obtain the alignment possibility of Y_t and each corresponding input element, namely, the weighted alpha. The matching process only needs to calculate the elements within the window, and the weight of the elements outside the window is 0. Finally, the output is processed through the normalized exponential function softmax, to obtain the required attention distribution probability within the range of the probability distribution, and the input Y_t encoded by the newly added LSTM unit is obtained.

5. Performance Evaluation

5.1. Data Description. In this study, the following case study was used to evaluate the performance of the proposed method, and we choose Nanshan District as the experimental area because Nanshan District is an important and typical downtown area in Shenzhen, Guangdong Province, China, where a representative regional road network composed of Nanhai Avenue, Binhai Avenue, Chuangye Road, and Houbinhai Road was chosen as the research area to cover all types of roads including expressways, arterial roads, secondary roads, and branch roads [31]. The sample data collected by the institute from May 1, 2017, to May 31, 2017, were all from the Shenzhen Urban Transportation Planning and Design Research Center, with a total of about 4 million samples. The process of converting the map into a road network, linking the floating car data, and selecting a suitable area for data extraction and research analysis is shown in Figure 10, and the detailed map of the selected study area is shown in Figure 11.

In this study, refer to literature [11, 12, 35], we used 5 minutes as time interval to collect experiment data and divided a day into 288 periods for collection, processing, and analysis. An example of floating vehicle data is shown in Table 1.

5.2. Data Quality Improvement. In order to improve the quality of the raw data to achieve more accurate prediction results, according to the missing characteristics, the original data are divided into accidental missing and multiple missing. Referring to related literature, it is found that the naive Bayesian method and dynamic time warping method can be used to repair these two types of missing data,
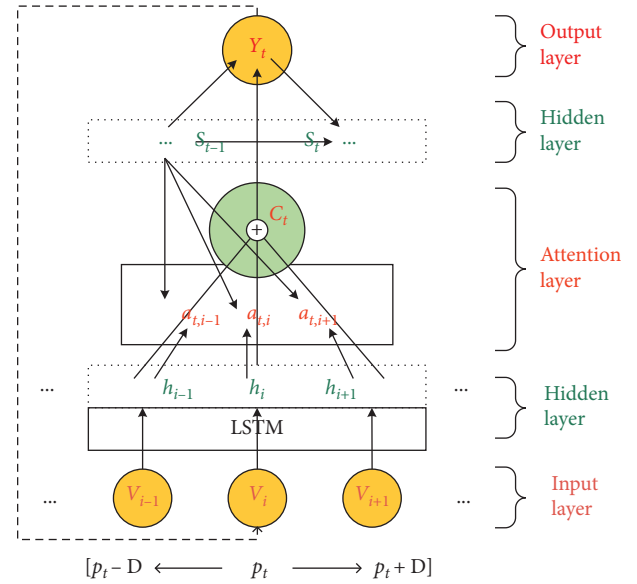

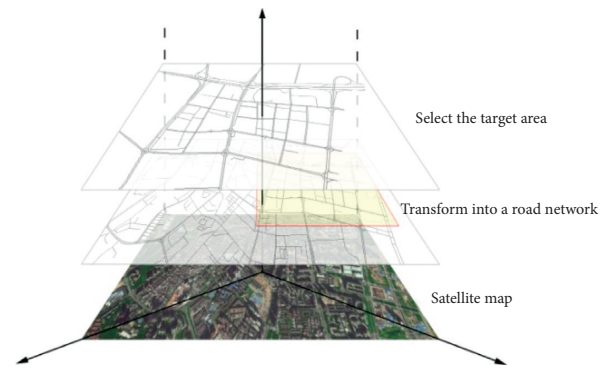FIGURE 9: Attention-based LSTM model.

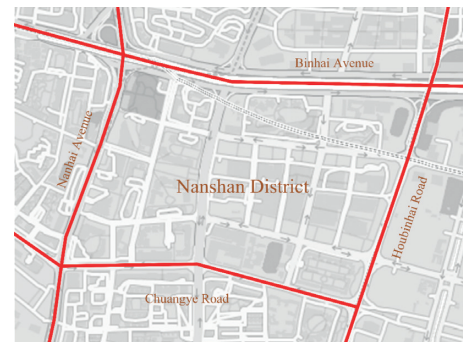FIGURE 10: Selection of research area.

FIGURE 11: The detailed map of the area.

respectively, with good performance [36–38]. Consequently, we choose the naive Bayesian method and dynamic time warping method to estimate the two types of missing data separately, obtaining a complete dataset without abnormal points, which lays a solid foundation for the subsequent prediction of short-term traffic speed.

TABLE 1: Data format of floating car.

Symbol	Definition	Data content	Remark
TIME	Data	20170507	20170507
PERIOD	Time interval	37	Time period 37, 185–190 min
LINKID	Road ID	380129	Unique section identification
FROMNODE	From the node	243022	Last node identification
TONODE	Destination node	254150	Next node identification
GOSPEED	Average road speed	42.936	Average speed of traffic in this section, km/h

5.3. Model Training and Parameter Optimization. According to the traffic speed prediction framework of the ATT-LSTM, the advanced neural network application programming interface Keras of TensorFlow is selected for experiments, picking up a sequence model and gradually covering the relevant neural network layers to establish a complete model. In the modeling process, the models with different depths are shown in Table 2. Taking the model training of trend signals as an example, the training loss functions and verification loss functions with different depths are shown in Figure 12.

Although the model structures with different depths have good performance, it can still be seen that, before the 4 depth layers, as the depth increases, the loss function greatly decreases, but after the 4 depth layers, the loss function increases slightly. Therefore, in this experiment, the performance of the model structure with the 4 depth layers is better than the model structure with other depths; it can be used for model training and further parameter adjustment.

In the model training module, through multiple experiments, the residual signal and the trend signal model have four depth layers and the same structure, but some parameters have subtle differences. The model structure and parameter settings are summarized in Table 3, and the model training parameter settings are summarized in Table 4.

5.4. Performance Evaluation Index. In order to evaluate the performance of the proposed prediction model, the evaluation indices include the average absolute error (MAE) and root mean square error (RMSE) are employed to measure the accuracy of the model prediction [11, 12, 39], in which the computational formulas are as follows:

$$MAE = \sum_{i=1}^{n} \frac{|obs_i - pre_i|}{n},$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (obs_i - pre_i)^2}, \quad (7)$$

where n is the total number of test samples, obs is the real traffic speed, and pre is the traffic speed predicted by the model. When verifying the prediction model, the test data are regarded as the target to be predicted, and the deviation between the prediction data and the real data is used as an evaluation of the accuracy of the prediction result. In addition, the efficiency of the model needs to be measured by the training time.

5.5. Validity Analysis of the Proposed Model. We use actual road network traffic flow speed data to verify and analyze the model proposed in this paper.

5.5.1. Evaluation of Effectiveness of Attention Mechanism. The purpose of introducing the attention mechanism is to find an accurate range of attention in the input sequence so that attention is only focused, or mostly focused, on the most relevant elements. In this study, this feature mainly works in two aspects: the selection of windows and the probability distribution of attention. Part of the dataset is randomly selected from the training set for verification, the model is trained, and the attention distribution is visualized. The results are shown in Figure 13.

As can be seen from the above figure, when selecting the window, the model finds the time $t = 126$ as the center position of the alignment window. From the distribution of attention weights, it can be seen that larger weights are distributed around this moment. The model has developed a strong focus on the key parts of the study; that is, it verifies the successful introduction and adaptation of the attention mechanism. At this time, discarding the data outside the window can greatly reduce redundant input, which is beneficial to the improvement of the model's efficiency. The models with and without the attention mechanism are trained separately, and the training time of each section on the road network is counted. The results are shown in Figure 14, where the light blue histogram is the training time of the model without introducing the attention mechanism and the blue dotted line represents the average training time of each section. The light orange histogram shows the model training time after introducing the attention mechanism. The orange dotted line represents the average training time for each road segment. After introducing the attention mechanism, it can be seen that the training time of most road sections was shortened to varying degrees. The average training time of the entire road network was shortened by approximately 1.7 s, which proved that the attention mechanism improved the efficiency of the model. This special case works, but it has a certain degree of universality and reliability.

5.5.2. Effectiveness Evaluation of Sequence Input Method. To verify the effectiveness of various sequence input prediction models, the prediction method of training the trend sequence and the residual sequence was verified separately.

The prediction result of the model using the trend sequence is shown in Figure 15(a), and the prediction result of

TABLE 2: Structures of model at different depths.

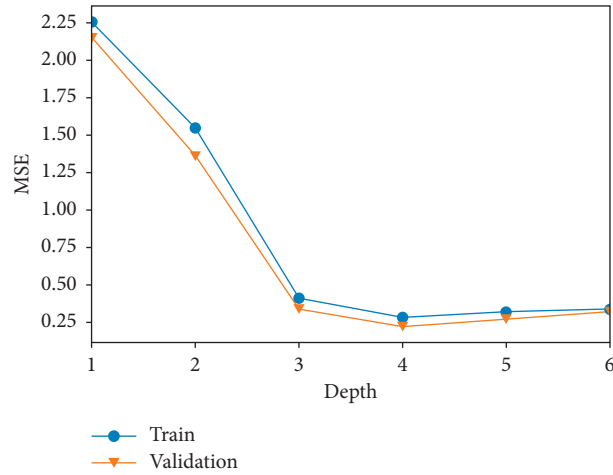| Depth	Structure of model
1	Input $\longrightarrow$ LSTM layer $\longrightarrow$ attention mechanism layer $\longrightarrow$ dense layer $\longrightarrow$ output
2	Input $\longrightarrow$ LSTM layer $\longrightarrow$ dropout layer $\longrightarrow$ attention mechanism layer $\longrightarrow$ dense layer $\longrightarrow$ output
3	Input $\longrightarrow$ LSTM layer $\longrightarrow$ dropout layer $\longrightarrow$ LSTM layer $\longrightarrow$ attention mechanism layer $\longrightarrow$ dense layer $\longrightarrow$ output
4	Input $\longrightarrow$ LSTM layer $\longrightarrow$ dropout layer $\longrightarrow$ LSTM layer $\longrightarrow$ dropout layer $\longrightarrow$ attention mechanism layer $\longrightarrow$ dense layer $\longrightarrow$ output
5	Input $\longrightarrow$ LSTM layer $\longrightarrow$ dropout layer $\longrightarrow$ LSTM layer $\longrightarrow$ dropout layer $\longrightarrow$ LSTM layer $\longrightarrow$ attention mechanism layer $\longrightarrow$ dense layer $\longrightarrow$ output
6	Input $\longrightarrow$ LSTM layer $\longrightarrow$ dropout layer $\longrightarrow$ LSTM layer $\longrightarrow$ dropout layer $\longrightarrow$ LSTM layer $\longrightarrow$ dropout layer $\longrightarrow$ attention mechanism layer $\longrightarrow$ dense layer $\longrightarrow$ output

FIGURE 12: Performance of loss function in different depth models.

TABLE 3: Structure and parameter of trend series and residual series of prediction models.

Layer	Series name of model (trend/residual)	Parameter	Size (trend/residual)
1	LSTM layer	Input_dim	1
		Neurons	50/100
		Param	10400/40800
2	Dropout (0.3/0.2) layer	Param	0
3	LSTM layer	Neurons	100
		Param	60400/80400
4	Dropout (0.3/0.2) layer	Param	0
5	Attention layer	Param	2116
6	Dense layer	Param	101
		Output_dim	1

TABLE 4: Training parameters of trend series and residual series of prediction models.

Parameter	Matching
Loss	Mse
Optimizer	Adam
Batch_size	128
Epoch	10

the model using the residual sequence is shown in Figure 15(b). Figure 15(c) shows the prediction result of the model utilizing the trend sequence plus the residual sequence. It can be seen that the trend sequence is very smooth and changes more coherently when forecasting separately, and the prediction accuracy is very high. In contrast, the variation rule of the residual series is not as obvious as that of the trend series, and the prediction results have a larger deviation than the trend series, but the accuracy of the model using the residual series is still relatively reliable. After combining the prediction results of the trend sequence and the residual sequence, it can be seen that the final prediction result reaches an ideal prediction accuracy. Compared with the result of direct prediction through the sequence without decomposition, which is shown in Figure 15(d), it is more advantageous to subdivide the data according to traffic speed and regularity for mining the internal characteristics of the data.
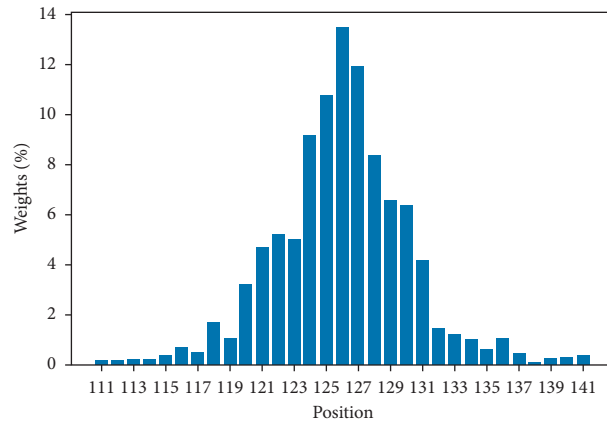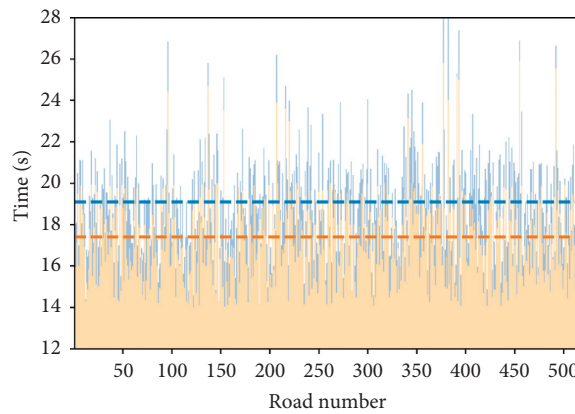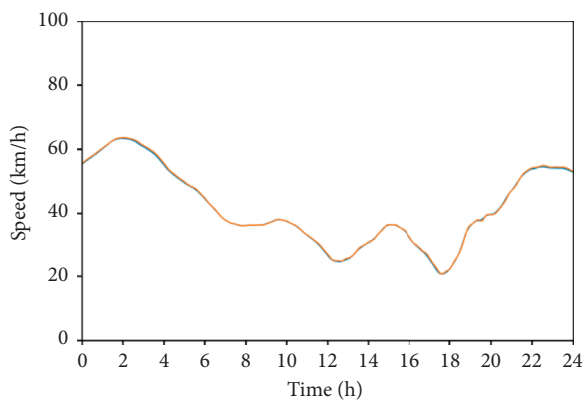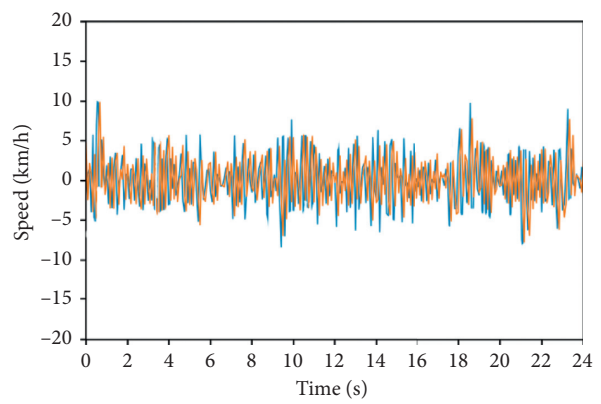
FIGURE 13: Attention distribution.

FIGURE 14: Training time of proposed model on each road segment with and without the attention mechanism.
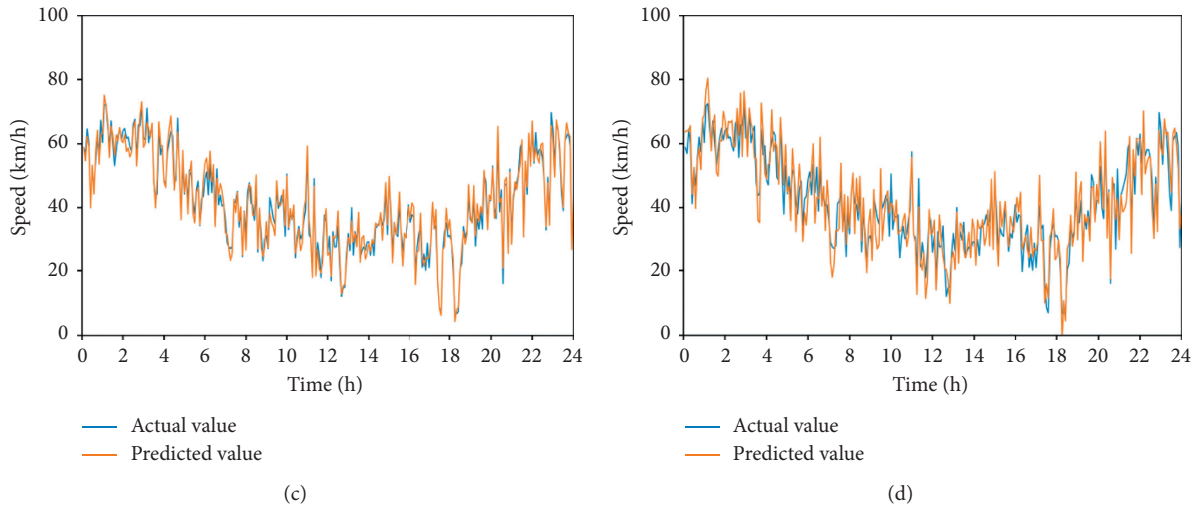

(a)

(b)

FIGURE 15: Continued.

(c)

(d)

FIGURE 15: Comparison of prediction results by different sequence input methods: (a) trend series forecast; (b) residual sequence prediction; (c) trend and residual mixed forecast; (d) undecomposed sequence prediction.

5.5.3. Evaluation of Model Universality. Universality is also important for constructing a prediction model. In order to verify the applicability of the research method in this paper to the short-term traffic speed prediction of different grades of roads, the road grades in the regional road network were divided, as shown in Figure 16. The prediction and evaluation of each road grade are carried out, and the final results are shown in Table 5. It can be seen that the research method used shows a good prediction effect for all road grades from expressways to branch roads, but the accuracy of the prediction results of high-grade roads is usually slightly better than that of low-grade roads. The main reason for this may be that the lack of data on low-level roads is more serious, and the dataset used for prediction may have had more estimates, which could make the source data deviate from the real situation to a certain extent in the trend. In addition, the high-level road datasets are more complete; there are only a few estimates, and the trend is closer to the true periodic law. In addition, compared to lower grade roads, high-grade roads have better road conditions, with fewer interference factors and less random interference. However, the overall error of the repaired dataset has little effect on the performance of research methods on all levels of roads.

5.5.4. Comparison of the Performance of Different Prediction Methods. As shown in Table 6, MAE and RMSE were used to evaluate the prediction accuracy of the model in steps of 5 min, 10 min, and 15 min. It turned out that the MAE and RMSE of the method used in this study are the smallest among all methods. The attention effectively reduces the error of this model. The MAE error is reduced by up to 12%, and the RMSE error is reduced by up to 5.56%. The calculation accuracy of CNN and LSTM-CNN is the closest to the method used in this study, but there is still a gap between the accuracy values. It can be seen that the research in this paper includes a data processing module and LSTM

FIGURE 16: Classification of road network, where the red sections are expressways, the orange sections are the arterial roads, the green sections are the secondary roads, and the gray sections are the branch roads.

TABLE 5: Evaluation of the prediction results for each road grade.

Road type	Expressway	Arterial road	Secondary road	Branch way
MAE (km/h)	2.02	2.16	2.27	2.47
RMSE (km/h)	3.12	3.15	3.53	3.76

prediction model based on the attention mechanism, which has good accuracy and robustness in practical applications.

Figure 17 shows a comparison of the training time of different algorithms before and after data processing on the entire road network. In each histogram, the dark color indicates that the input data are the data processed by the method used in this study, and the light color indicates that the input data are the data processed by the interpolation method. It can be seen that although the training time of each algorithm is different, the training time has been

TABLE 6: Average prediction error of each algorithm.

Model	5 min		10 min		15 min	
	MAE (km/h)	RMSE (km/h)	MAE (km/h)	RMSE (km/h)	MAE (km/h)	RMSE (km/h)
ATT-LSTM	**2.23**	**3.39**	**2.75**	**4.08**	**3.26**	**4.37**
LSTM	2.51	3.47	3.14	4.32	3.60	4.53
RNN	3.54	4.73	4.36	5.29	4.85	5.70
LSTM-CNN	2.77	3.93	3.48	4.81	3.93	5.04
CNN	2.69	3.88	3.45	4.38	4.21	5.16
ANN	4.53	5.69	5.10	5.83	5.48	6.59
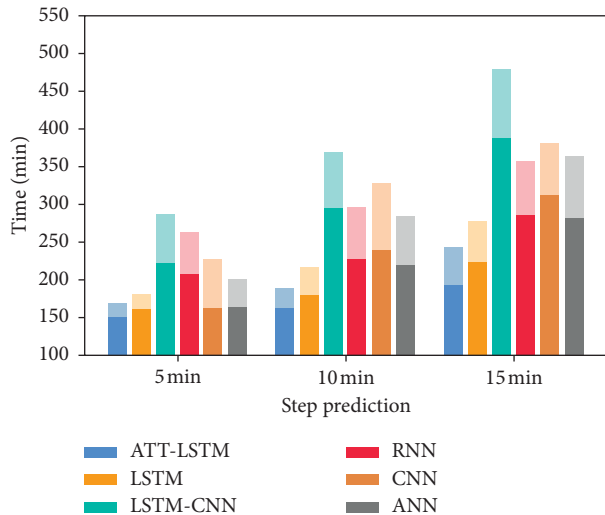
FIGURE 17: Average training time of various algorithms.

FIGURE 18: The distribution of real traffic velocity on the road network.

shortened after using this method to process the data. This is due to the data processing and repair method adopted in this study, which proves that this method can not only optimize the model prediction accuracy but also effectively reduce the computational cost of model training and improve the efficiency of the model. Compared with other models, the LSTM model consumes less time. After introducing the attention mechanism, the efficiency of the LSTM model can be further improved. In addition, in the verification of prediction accuracy, the performance of CNN network-related algorithms that are close to the method used in this study has poor performance in training time and is less practical because of overfitting. The above example verification proves that the research method in this study is not only more accurate than other models in terms of prediction accuracy but also has a shorter model training time and higher model efficiency.

TransCAD is used to import data for map matching and speed classification and to compare the real speed distribution of the road network and the prediction results. Figure 18 shows the traffic conditions at a certain moment in the morning rush hour as taken from real data. The speed values are divided into 20 groups with an interval of 5 km/h, with different colors indicating the speed group. The closer the color is to red, the lower the traffic speed of the road section. The closer the color is to green, the higher the traffic speed of the road section. A black line indicates that the data of the section are temporarily missing. It can

be seen from the figure that high-grade roads have more traffic travel demand than low-grade roads, but higher grade roads have better road environments and real-time road conditions. There is less traffic flow in low-speed areas than low-grade roads, and the traffic flow is smoother. In contrast, low-grade roads have obvious congestion in many areas during rush hour; that is, the traffic flow is low. Figure 19 shows the prediction effect of road network speed distribution during the peak hour period. It is found that the distribution map is almost synchronized with the real distribution, and even some original sections with missing data were repaired and given specific values in the forecast.

In summary, through the verification and comparative analysis of the hybrid prediction method (ATT-LSTM), it is found that this method has outstanding performance and applicability for short-term traffic speed prediction on urban road networks. Specifically, different from other studies, the application scenario in this article is an urban road network with varied road levels (including expressways, main roads, secondary roads, and branch roads). Especially, for the raw data with a certain missing, this paper firstly fills out incomplete data according to its missing type, which effectively improves the quality of the data, enhances the usability of sample data, and improves the accuracy of the model to a certain extent. At the same time, the attention mechanism is used to effectively assign weights to distinguish the importance of traffic sequences, which helpfully reduces the training time of the model and improves the computational efficiency of the model. The experiment results have proved that the proposed method is superior to other advanced methods both in predicting accuracy and computational efficiency.

FIGURE 19: The distribution of predicted traffic speed on the road network.

6. Conclusion

In this paper, we propose an entire forecasting framework of short-term traffic speed combined with the data preprocess module and prediction module. In the data preprocessing module, in order to improve the sample data quality, we use the naive Bayes and dynamic time warping methods to fill the sparse traffic speed data to provide a complete dataset for the prediction work. In the prediction module, for the sake of improving the accuracy of shorttime traffic velocity prediction, attention mechanism is introduced. An ATT-LSTM traffic speed prediction model is proposed. Firstly, a window is selected in the input sequence according to the target prediction value. Next, the window is matched to obtain the attention weight. And then, we calculate the predicted value encoded by the LSTM through the attention distribution probability. Finally, the model is verified using road network data from Nanshan District, Shenzhen. Compared with deep learning algorithms such as RNN, CNN, and LSTM-CNN, the ATT-LSTM model has more advantages in terms of prediction accuracy and calculating efficiency. The attention mechanism can further improve the computational efficiency of the prediction model. In addition, after introducing the attention mechanism, the error of the prediction model is significantly reduced. The MAE is reduced by up to 12.4%, and the RMSE is reduced by up to 5.5%. This demonstrates that the attention mechanism can effectively improve the accuracy of the prediction results.

Due to the limitations in the objective conditions during the research period, the research content needs to be further improved. In future work, we plan to conduct a deeper discussion on related work, including the following two aspects: (1) while subdividing missing types and optimizing models, we should look back at longer historical data to obtain more accurate estimates; (2) in the next prediction model, we should consider more integrated data sources and add more dimensional factors that affect the traffic operation state of urban road networks, so as to further improve the accuracy and practicability of prediction.

Data Availability

The data used to support the findings of this study are included within the article. The source and composition of the experimental data are explained in Section 5.1. All the experimental data in this paper are provided by the Shenzhen Urban Transportation Planning and Design Research Center Shenzhen, Guangdong China.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

Acknowledgments

References

[1] A. Jalali, R. Mallipeddi, and M. Lee, "Sensitive deep convolutional neural network for face recognition at large standoffs with small dataset," *Expert Systems with Applications*, vol. 87, pp. 304–315, 2017.

[2] T. Zhang, L. Sun, L. Yao, and J. Rong, "Impact analysis of land use on traffic congestion using real-time traffic and POI," *Journal of Advanced Transportation*, vol. 2017, pp. 1–8, 2017.

[3] J. Guo, W. Huang, and B. M. Williams, "Adaptive Kalman filter approach for stochastic short-term traffic flow rate prediction and uncertainty quantification," *Transportation Research Part C: Emerging Technologies*, vol. 43, pp. 50–64, 2014.

[4] Z. Huang, L. Xu, and Y. Lin, "Multi-stage pedestrian positioning using filtered WiFi scanner data in an urban road environment," *Sensors*, vol. 20, no. 11, p. 3259, 2020.

[5] C. Dong, "An improved deep learning model for traffic crash prediction," *Journal of Advanced Transportation*, vol. 2018, Article ID 3869106, 13 pages, 2018.

[6] L. Huang, H. Guo, R. Zhang, D. Zhao, and J. Wu, "A data-driven operational integrated driving behavioral model on highways," *Neural Computing and Applications*, vol. 32, no. 16, pp. 13017–13033, 2020.

[7] Z. Huang, L. Xu, Y. Lin, P. Wu, and B. Feng, "Citywide metro-to-bus transfer behavior identification based on combined data from smart cards and GPS," *Applied Sciences*, vol. 9, no. 17, p. 3597, 2019.

[8] X. Wang, L. Xu, and K. Chen, "Data-driven short-term forecasting for urban road network traffic based on data processing and LSTM-RNN," *Arabian Journal for Science and Engineering*, vol. 44, no. 4, pp. 3043–3060, 2019.

[9] L. Huang, H. Guo, R. Zhang, and J. Wu, "LSTM-based lane-changing behavior model for unmanned vehicle under environment of heterogeneous human-driven and autonomous vehicles," *China Journal of Highway and Transport*, vol. 33, no. 7, pp. 156–166, 2020.

[10] S. Lu, Q. Zhang, G. Chen et al., "A combined method for short-term traffic flow prediction based on recurrent neural network," *Alexandria Engineering Journal*, pp. 4–11, 2020.

[11] H. Zheng, F. Lin, X. Feng, and Y. Chen, "A hybrid deep learning model with attention-based conv-LSTM networks

for short-term traffic flow prediction," *IEEE Transactions on Intelligent Transportation Systems*, vol. 99, pp. 1–11, 2020.

[12] Y. Yu, Y. Zhang, S. Qian, S. Wang, Y. Hu, and B. Yin, "A low rank dynamic mode decomposition model for short-term traffic flow prediction," *IEEE Transactions on Intelligent Transportation Systems*, pp. 1–14, 2020.

[13] Y. Wu, H. Tan, L. Qin, B. Ran, and Z. Jiang, "A hybrid deep learning based traffic flow prediction method and its understanding," *Transportation Research Part C: Emerging Technologies*, vol. 90, no. 5, pp. 166–180, 2018.

[14] X. Luo, D. Li, Y. Yang, and S. Zhang, "Spatiotemporal traffic flow prediction with KNN and LSTM," *Journal of Advanced Transportation*, vol. 2019, Article ID 4145353, 10 pages, 2019.

[15] Y. Jia, J. Wu, and M. Xu, "Traffic flow prediction with rainfall impact using a deep learning method," *Journal of Advanced Transportation*, vol. 2017, Article ID 6575947, 10 pages, 2017.

[16] Q. Tang, M. Yang, and Y. Yang, "ST-LSTM: a deep learning approach combined spatio-temporal features for short-term forecast in rail transit," *Journal of Advanced Transportation*, vol. 2019, Article ID 8392592, 8 pages, 2019.

[17] M. S. Ahmed and A. R. Cook, "Analysis of freeway traffic time-series data by using Box-Jenkins techniques," *Transportation Research Record*, vol. 722, pp. 1–9, 1979.

[18] E. I. Vlahogianni, M. G. Karlaftis, and J. C. Golias, "Short-term traffic forecasting: where we are and where we're going," *Transportation Research Part C: Emerging Technologies*, vol. 43, no. 2, pp. 3–19, 2014.

[19] J. W. C. Van Lint, "Online learning solutions for freeway travel time prediction," *IEEE Transactions on Intelligent Transportation Systems*, vol. 9, no. 1, pp. 38–47, 2008.

[20] J. Xia, M. Chen, and W. Huang, "A multistep corridor travel-time prediction method using presence-type vehicle detector data," *Journal of Intelligent Transportation Systems*, vol. 15, no. 2, pp. 104–113, 2011.

[21] J. Guo and B. M. Williams, "Real-time short-term traffic speed level forecasting and uncertainty quantification using layered Kalman filters," *Transportation Research Record*, vol. 2175, pp. 28–37, 2010.

[22] M. M. Hamed, H. R. Al-Masaeid, and Z. M. B. Said, "Short-term prediction of traffic volume in urban arterials," *Journal of Transportation Engineering*, vol. 121, no. 3, pp. 249–254, 1995.

[23] B. M. Williams and L. A. Hoel, "Modeling and forecasting vehicular traffic flow as a seasonal ARIMA process: theoretical basis and empirical results," *Journal of Transportation Engineering*, vol. 129, no. 6, pp. 664–672, 2003.

[24] Q. Y. Ding, X. F. Wang, X. Y. Zhang et al., "Forecasting traffic volume with space-time ARIMA model," *Advanced Materials Research*, vol. 156-157, pp. 979–983, 2010.

[25] E. I. Vlahogianni and M. G. Karlaftis, "Testing and comparing neural network and statistical approaches for predicting transportation time series," *Transportation Research Record*, vol. 2399, pp. 9–22, 2013.

[26] H. Yu, N. Ji, Y. Ren, and C. Yang, "A special event-based *K*-nearest neighbor model for short-term traffic state prediction," *IEEE Access*, vol. 7, no. 99, pp. 81717–81729, 2019.

[27] W. C. Hong, Y. Dong, F. Zheng et al., "Hybrid evolutionary algorithms in a SVR traffic flow forecasting model," *Applied Mathematics and Computation*, vol. 217, no. 15, pp. 6733–6747, 2017.

[28] Z. Zhao, W. Chen, X. Wu, P. C. Y. Chen, and J. Liu, "LSTM network: a deep learning approach for short-term traffic forecast," *Iet Intelligent Transport Systems*, vol. 11, no. 2, pp. 68–75, 2017.

[29] X. Ma, Z. Dai, Z. He et al., "Learning traffic as images: a deep convolutional neural network for large-scale transportation network speed prediction," *Sensors*, vol. 17, no. 4, 2017.

[30] K. Y. Chan, T. S. Dillon, J. Singh, and E. Chang, "Neural-network-based models for short-term traffic flow forecasting using a hybrid exponential smoothing and Levenberg-Marquardt algorithm," *IEEE Transactions on Intelligent Transportation Systems*, vol. 13, no. 2, pp. 644–654, 2012.

[31] X. Wang and L. Xu, "Wavelet-based short-term forecasting with improved threshold recognition for urban expressway traffic conditions," *IET Intelligent Transport Systems*, vol. 12, no. 6, pp. 463–473, 2018.

[32] G. Hoffmann and J. Janko, "Travel times as a basic part of the LISB guidance strategy," in *Proceedings of the Third International Conference on Road Traffic Control*, pp. 6–10, London, UK, May 1990.

[33] P. Wu, L. Xu, and Z. Huang, "Imputation methods used in missing traffic data: a literature review," in *Proceedings of the International Symposium on Intelligence Computation and Applications*, pp. 662–677, Stockholm, Sweden, August 2019.

[34] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Computation*, vol. 9, no. 8, pp. 1735–1780, 1997.

[35] B. Yang, S. Sun, J. Li, X. Lin, and Y. Tian, "Traffic flow prediction using LSTM with feature enhancement," *Neurocomputing*, vol. 332, pp. 320–327, 2019.

[36] X. Chen, Z. He, and L. Sun, "A Bayesian tensor decomposition approach for spatiotemporal traffic data imputation," *Transportation Research Part C: Emerging Technologies*, vol. 98, pp. 73–84, 2018.

[37] A. J. Fowe and Y. Chan, "A microstate spatial-inference model for network-traffic estimation," *Transportation Research Part C: Emerging Technologies*, vol. 36, no. 11, pp. 245–260, 2013.

[38] J. Wang and M. Yun, "Development of urban road network traffic state dynamic estimation method," *Mathematical Problems in Engineering*, vol. 2015, Article ID 714149, 10 pages, 2015.

[39] Y. Zhao, S. V. Ukkusuri, and J. Lu, "Multidimensional scaling-based data dimension reduction method for application in short-term traffic flow prediction for urban road network," *Journal of Advanced Transportation*, vol. 2018, Article ID 3876841, 10 pages, 2018.