BASED ON THE SPATIO-TEMPORAL CHARACTERISTICS AND 1D CNN-BiLSTM OF THE SHORT-TERM PREDICTION AND EVALUATION OF URBAN ROAD TRAFFIC

基于时空特性和 1D CNN-BiLSTM 的城市道路交通短时预测与评价

A THESIS

SUBMITTED TO SCHOOL OF MATHEMATICS & PHYSICS OF

XI'AN JIAOTONG-LIVERPOOL UNIVERSITY

IN PARTIAL FULFILMENT FOR THE AWARD OF THE DEGREE OF

BSC APPLIED MATHEMATICS

By

Keyu Chen 2037673

Supervisor: Dr. Arodh Lal Karn

May 8, 2024

Abstract

This paper analyzes the characteristics and limitations of the existing short-term traffic parameter prediction model. Based on the spatiotemporal information, the 1D CNN-BiLSTM model exploits the periodicity and similarity of traffic parameters in the temporal dimension, as well as the similarity transfer in the spatial dimension. In this model, the 1D CNN model inputs data with spatial characteristics (merging two traffic parameter data of adjacent roads), and can effectively extract the spatial characteristics of the data through a series of steps. The BiLSTM model can consider bidirectional timing information simultaneously, which means capture more temporal features than a single LSTM. Therefore, 1D CNN-BiLSTM short-term traffic parameter predictive model has the spatio-temporal information capacity of road traffic in the actual situation. Meanwhile, traffic and average speed, which are inherently related, are used as two parameters to construct the data set. Following model training and testing, it is demonstrated that the model has effective learning ability. Finally, the congestion prediction accuracy of the model is high, as demonstrated by the calculation of the congestion index with the prediction results. This indicates that the model has excellent congestion prediction ability.

KEY WORDS: 1D CNN-BiLSTM, Short-term traffic forecasting, Traffic flow, Average speed, Congestion index

摘要

本文通过分析现有短时交通参数预测模型的特点和局限性。针对交通参数在时间维度上具有的周期性和相似性，在空间维度上具有近域相似传递性，研究提出一种基于增强道路时空信息的 1D CNN-BiLSTM 模型。其中，1D CNN 模型输入具有空间特性的数据(相邻道路的两个交通参数数据进行合并)，经一系列步骤可有效提取数据的空间特征。BiLSTM 模型可以同时考虑双向的时序信息，比起单一的 LSTM 能够捕获到更多的时间特征。因此，1D CNN-BiLSTM 短时交通参数预测模型具备全面捕获实际情况中道路交通的时空信息能力。同时，采用有内在联系的流量和平均速度作为两个参数，构建数据集。通过模型训练和测试后，发现模型学习能力优良。最后用预测结果计算拥堵指数得到模型的拥堵预测准确率较高，表明该模型预测拥堵能力优秀。

关键词： 1D CNN-BiLSTM模型，短时交通预测，交通流量，平均速度，拥堵指数

Acknowledgements

Here, I will take this opportunity to thank my supervisor Dr. Arodh Lal Karn. In the stage of thesis topic selection and research framework construction, The teacher provided me with his paper "DESIGNING A DEEP LEARNING-BASED FINANCIAL DECISION SUPPORT SYSTEM FOR FINTECH TOSUPPORT CORPORATE" CUSTOMER's CREDIT EXTENSION "and asked me to output ppt according to the paper. In this learning process, I learned about EM algorithm, maximum likelihood algorithm, RNN, LSTM, GRU, CNN, DNN, hyperparameter calculation and binary classification model evaluation analysis. This provided an idea for me to choose the machine learning model for the research paper topic later. After that, the teacher shared relevant literature on short-term traffic prediction, which made me further understand how to deal with outliers in traffic flow prediction and how to detrend analysis, thus gradually forming the preliminary framework of the paper. Thanks to the teacher's careful guidance!

Contents

Chapter 1

Introduction

For the past few years, the number of urban motor vehicles has increased rapidly, which leads to the inability of urban road traffic capacity to meet the growth of the number of motor vehicles. Due to the imbalance between road supply and demand, a number of traffic jams have occurred, seriously affecting the safety and operation efficiency of urban traffic. To solve urban traffic problems, city managers analyze urban traffic flow data and predict urban traffic conditions in advance by building smart city data platforms. On the one hand, it can guide the optimisation of urban traffic and formulate targeted traffic policies to effectively alleviate traffic congestion (Zhang and Kabuka, 2018). On the other hand, it can provide a reference for public travel operations. Hence, the accurate and reliable prediction of urban traffic data is particularly significant in the field of intelligent transportation.

The difficulty of the traffic flow prediction model is that the traffic flow has both dynamic time variation characteristics and typical spatial variation characteristics. Therefore, two important dependencies need to be considered: spatial dependence and temporal dependence (Kumar, Panwar and Chaurasiya, 2023). Most of the existing research on traffic flow prediction has concentrated on the use of statistical models and machine learning techniques. Kalman filter and ARIMA statistical models can be used to predict the dynamic traffic parameters with time, but it is challenging to accurately describe the data characteristics of non-linear traffic flow. Machine learning has advantages in extracting complex nonlinear features and dealing with complex system problems, and can better predict traffic flow. As an extension of machine learning, deep learning techniques can effectively extract large data with spatiotemporal dependencies in high-dimensional data mining (Zhang and Kabuka, 2018), which significantly improves traffic flow forecasts.

The above background information leads to the conclusion that this study will adopt a 1D

CNN-BiLSTM hybrid model for the spatio-temporal characteristics of urban traffic parameters. The CNN model can capture the spatial characteristics of the road network, while the BiLSTM output layer can receive backward and forward information simultaneously (thus obtain relatively complete data information). Through the construction of the hybrid model, the research aims to solve the following three problems:
(1) Does the 1D CNN-BiLSTM hybrid model effectively predict the daily change in traffic flow and average speed (Links the internal relationship between two parameters)?
(2) What is the prediction effect of the 1D CNN-BiLSTM hybrid model on adjacent continuous road sections?
(3) What is the prediction effect of the 1D CNN-BiLSTM hybrid model on traffic congestion?

This study proposes a new approach to short-term traffic parameter prediction, namely the CNN-BiLSTM model. The aim is to develop the short-term traffic prediction model capture the temporal and spatial condition comprehensively. The advantages of 1D CNN model in processing spatial data and BiLSTM model in bidirectional time series data are combined to enhance the model's capacity to represent temporal and spatial information about roads. In addition, the complete data sets for three continuous roads in Kunming Second Ring viaduct were used for training and testing. This enabled the prediction of short-term traffic flow, average speed and traffic congestion index for urban roads. Hence, the model have high applicability and prediction accuracy.

The content of this paper is organized as follows: The first chapter describes the research background and outlines the research questions; The second chapter provides an overview of the latest progress in traffic parameter prediction at both domestic and abroad. The third chapter outlines the technical route and research method. This paper describes the source of traffic data, data preprocessing, and the characteristics of traffic parameters. Chapter 4 outlines the construction of the CNN-BiLSTM model and the methodology for model training. The fifth chapter verifies the prediction effect of the model and evaluates the road traffic condition. Chapter 6 concludes the work of the whole paper.

Chapter 2

Literature Review

At present, the global urban traffic flow prediction methods are mainly divided into three categories: parametric models, non-parametric models, and hybrid models (Mario, 2022). In both theory and practice, the parametric model of short-term traffic flow prediction based on mathematical statistics is highly developed. The non-parametric model predicts the changing trend of traffic conditions by learning the data characteristics and their correlation. The hybrid model combines the strengths of different types of prediction models, handling the limitations of a single model and improves the precision of traffic flow prediction.

2.1 Parametric model

Parametric modeling relies on the dynamic properties of traffic flow over time, employing time series models such as ARIMA, SARIMA, SARIMAX, and NARX for flow prediction. ARIMA is predominantly utilized within this family, where its components can be employed individually or in combination. Alghamdi et al. (2019) used several ARIMA models to analyse the main factors affecting urban traffic congestion by ACF and PACF, and obtained a model with adequate predictive performance by comparison. Deretić et al. (2022) found that traffic accidents have seasonal time characteristics, and used the SIRMA model to predict and analyse the trend of traffic accidents. Despite the effectiveness of parameter models in predicting traffic flow, they assume constancy in flow mean and variance, struggling to accurately describe its random and nonlinear nature.

2.2 Nonparametric model

In transport, non-parametric models can better predict complex changes such as road conditions and traffic volumes, compared to the limitations of parametric models. Non-parametric prediction methods include SVR model, artificial neural network model and deep learning model. ANN method, KNN method, fuzzy logic system method and SVR

method are shallow models (Fang et al., 2023). Ghanim, D. Muley, M. Kharbeche (2022) used ANN model to analyse the effect of policy and other factors on traffic flow under the of COVID-19. A comparison between the model's predictions and actual observations revealed a linear relationship, with a slope nearing unity, indicating the ANN model's efficacy in forecasting traffic flow trends.

Due to the characteristics of high-dimensional data mining, deep learning models can capture the temporal and spatial characteristics of complex dynamic traffic flow and show advantages in the field of traffic prediction. RNN with its variants LSTM and GRU are extensively utilized to analyze the temporal dynamics within traffic data. While RNNs are proficient in modeling the temporal aspects of traffic flow, they struggle to encapsulate the spatial attributes of urban road networks. RNNs are capable of learning dependencies within short-term time series data, yet they face limitations in addressing issues related to the vanishing and exploding gradient problems that are common in traffic flow data (Erdem, 2020). Mackenzie, Roddick Zito (2019) introduced HTM as a potent method for short-term traffic flow prediction by comparing the prediction effects of HTM and LSTM.

Scholars employ CNN to analyse the spatial characteristics of urban road systems. Wu et al. (2018) use CNN model to obtain the spatial characteristics of traffic flow while considering the periodicity of traffic flow. The historical traffic data is fed into the prediction model, and the correlation between the historical data and the prediction data is detected. Ghanim, Muley and Kharbeche (2017) proposed a decentralised method of CNN to predict the congestion state of each road node according to the current state of its neighbouring points. Additionally, Dai, Ma and Xu (2019) utilized GRU to process the two-dimensional matrix with spatial characteristics to obtain the spatiotemporal characteristics of traffic flow and the best prediction algorithm.

2.3 Hybrid model

Due to the limitations of a single model in accurately capturing the spatio-temporal attributes of traffic flow data, a hybrid model is proposed. Du et al. (2020) proposed a hybrid deep learning method for traffic flow prediction, which used CNN to describe spatial features and GRU to describe temporal features. Consequently, the deep learning model represented by GCN can better solve the spatial characteristics of urban road

network with topological network structure.

In comparison to SARIMA, ANN is more adept at analyzing traffic flow trends, avoiding abrupt fluctuations during holidays. Mohammadzadeh, Choupani and Afshar (2023) combining ANN with LSTM, LSTM is used to analyze the daily and holiday traffic volume changes, so as to obtain the short-term and long-term time characteristics of traffic flow. Kumar, Panwar and Chaurasiya (2023) combined GCN and GRU to construct the STAGG model to solve the problem of spatial dependence and temporal dependence in traffic prediction models, which can better predict urban road traffic volumes in real time. Kumar, Moreira and Chandra (2023) developed a DYGCN-LSTM that models the nonlinear spatial-temporal interactions of data using actual traffic datasets, reducing RMSE errors significantly. In the context of smart cities, Zafar et al. (2022) integrated heterogeneous data obtained from sensors, services and external data sources into spatio-temporal features, and on this basis established an LSTM-GRU hybrid model to predict traffic speed.

M. Muñoz-Organeroo (2023) used different CNNS to extract the spatial patterns of each type of image, and then combined the output of the two CNNS and fed to the final LSTM -based RNN. Meanwhile, Bogaerts et al. (2020) combined the CNN model and LSTM model to analyze the spatio-temporal characteristics of traffic flow, thereby enhancing the accuracy of short-term traffic predictions.

By comparing the prediction effects of unidirectional LSTM and BiLSTM models, Abduljabbar, Dia, and Tsai (2021) proposed that the overall structure of the BiLSTM model is composed of two unidirectional LSTMs stacked in forward and reverse directions. This allows the model to simultaneously consider the periodic characteristics of time series data from different directions of the data. Therefore, when dealing with time series data, the analysis and processing effect is better than that of single LSTM. Ma, Dai and Zhou (2022) developed an enhanced LSTM-BiLSTM approach that builds upon the foundations of both LSTM and BiLSTM models. This refined method surpasses the single model in delivering more accurate and stable prediction outcomes. Zhuang and Cao (2023) devised a hybrid KNN-BiLSTM model that utilizes the KNN algorithm to identify and select traffic flow data exhibiting high spatial correlation, which is subsequently processed through a BiLSTM model for forecasting. The BiLSTM model employs a bidirectional LSTM

architecture, considering information transmission both from the past and into the future, thereby enhancing the precision of traffic flow predictions. Zhu, Boada and Boada (2024) implemented an approach where LSTM was utilized to capture temporal dynamics, and GAT was employed to gather spatial data of the road system. This integrated model was thus better suited for accurately predicting different road traffic scenarios.

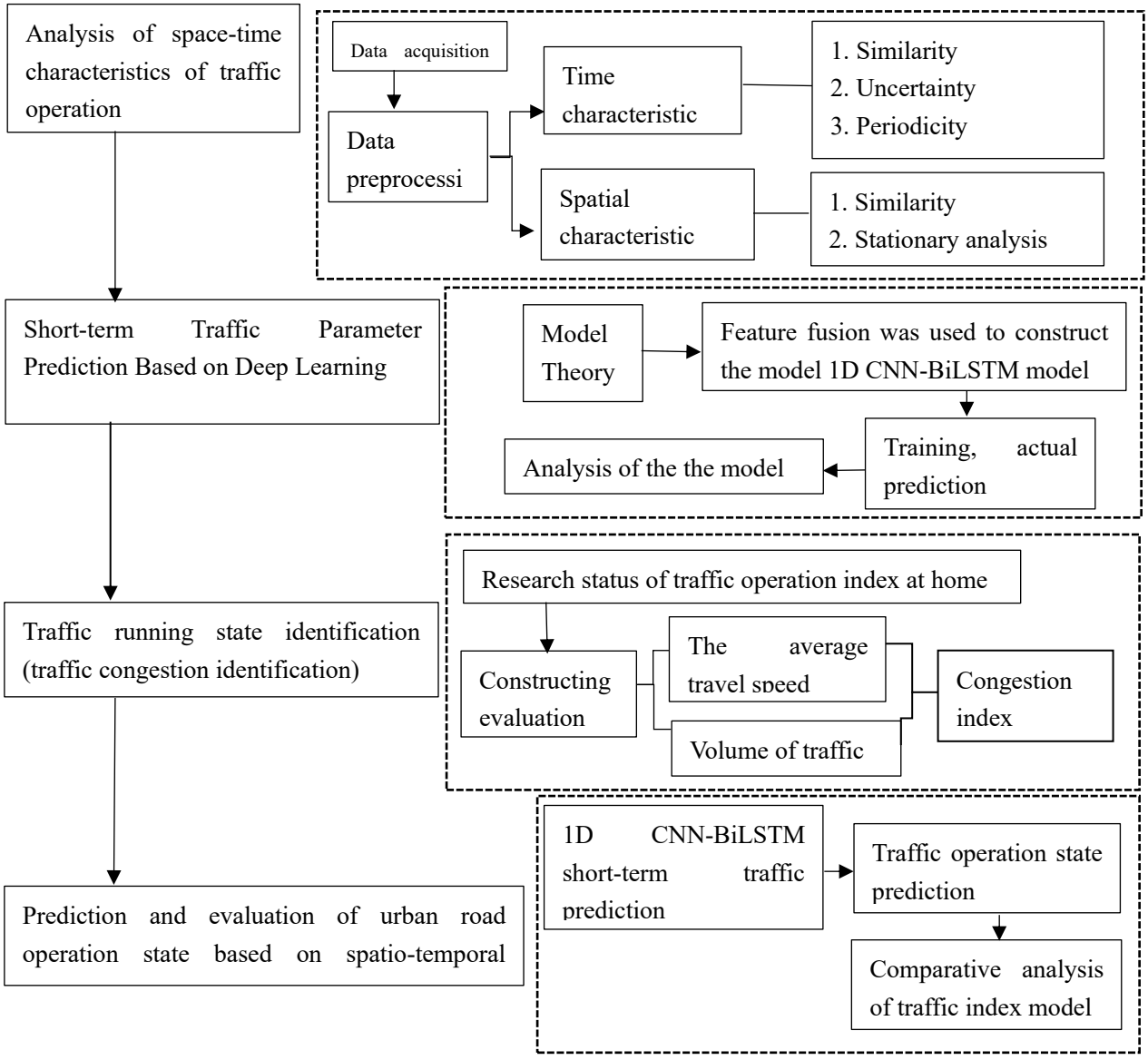
In conclusion, traditional data analysis methods often have specific model structures and numerous assumptions, which cannot meet the needs of analyzing traffic volume data. Therefore, it is essential to investigate and evaluate the accuracy and reliability used for traffic state discrimination and prediction. The conventional shallow prediction model is unable to fulfil the dual requirements of time and space dependence in traffic prediction. Compared with RNN model, CNN model can better analyze the spatial characteristics of urban road system. However, the traditional CNN mainly deals with data with spatial structure, which may result in the loss of important temporal features. This paper will apply a version extended to deal with one-dimensional time series, namely 1D-CNN. Therefore, this study intends to use the CNN-BiLSTM hybrid model. In addition, the average speed and traffic flow are further evaluated as evaluation parameters to assess the congestion state of urban roads (Kumar et al. 2017).

Combined forecasting model based on the above methods: investigate the potential correlation of different traffic parameters (traffic flow and average speed). On this basis, a CNN-BiLSTM short-term traffic prediction model is established to predict short-term traffic parameters and the running state of road traffic.

Chapter 3

Research methods

3.1 Technical route



3.2 Analysis of space-time characteristics of traffic operation

3.2.1 Data acquisition

This paper is based on the monitoring data of urban road gantry (provided by Kunming Planning Institute). Gantry observation is the use of fixed position detectors to calculate the information of passing vehicles, which can detect the speed and flow of vehicles within a certain time interval, and is applied to the analysis and research of traffic operation conditions.

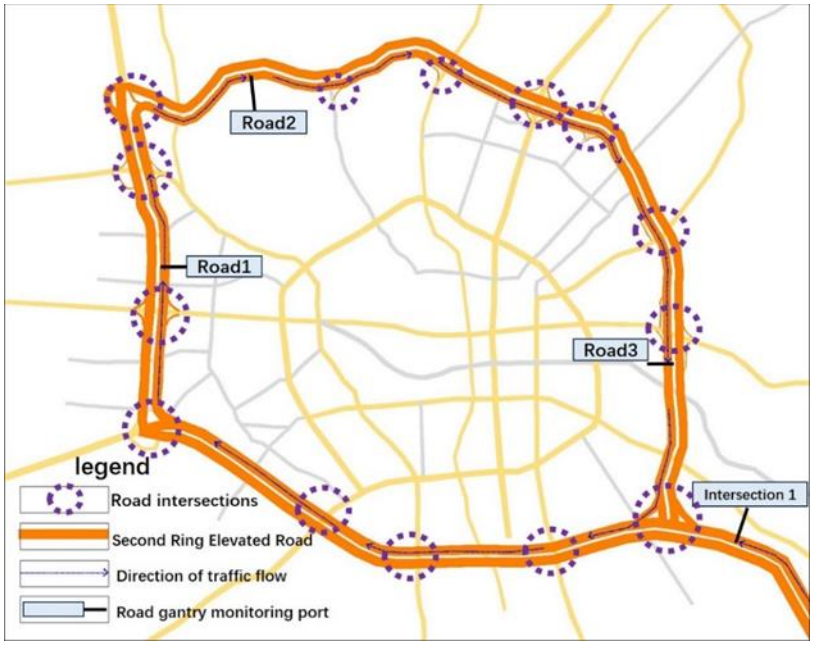


Figure 3.1: Schematic of the study area with gantry location

Taking the second ring viaduct in Kunming city as the research object, the Second Ring Viaduct is used as the expressway system in the central urban area. With a total length of 27km, a total of four gantry detection devices have been installed. In order, enter the south crossing line of the second Ring Road from Intersection1 (east to west research section of K9+900 of the Shantu-Kunming Expressway), driving from east to west. First pass Road1 (upper Mingbo interchange of Line B of the West Second Ring Road to the middle section of Huangtupo Interchange), driving from south to north. Subsequently, after Road2 (North second Ring Line B Jinjingshan Armed Police Gas station section), drive from west

to east. Then, after Road3 (East second ring line B upper Juhua Lijiao 1km gate frame), continue from north to south. From Road1 to Road2, two overpasses pass along the route, and six overpasses pass from Road2 to Road3. See Figure 3.1 for details of the study area and gantry location.

3.2.2 Data preprocessing

In this paper, the traffic volume and average speed are selected as the traffic parameters of the traffic flow operation characteristics. After repairing the missing data, the daily statistical time was sorted by time, and then it was made into a time series label. The statistical time data were resampled according to the frequency of 15 minutes, to obtain the vehicle flow of each gallows for 24 hours of the day. Similarly, the average speed of the road section resamples the statistical time data at the frequency of 15min and takes the average speed of the interval. In this process, the missing values are processed and the statistical moments of each day are sorted by time. This was used as a time series label, and finally the average speed of each road section in every 15min interval for 24 hours of the day was obtained.

Handling missing values:

(1) Inpainting based on similar historical trend data

The traffic flow and average speed of the same road have certain periodicity, and the data collected at the same time of the week demonstrate a certain degree of similarity. This similarity feature is used to repair the data, and the formula is as follows (Zhang and Jiao, 2023):

$$t=t^{h} \tag{3.1}$$

(2) repair based on the data of adjacent periods

For individual missing values, the average value of two adjacent periods before and after the missing value can be used, and the formula is as follows:

$$t_k=\frac{t_{k-1}+t_{k+1}}{2} \tag{3.2}$$

In the similarity analysis of adjacent roads, road1, road2 and road3, which are connected clockwise, have individual missing values at 17:30 on October 16. Therefore, this study

uses the average value of the data at 17:15 and 17:45 to repair the data.

(3) repair based on similar road data

For the road sections with similar characteristics such as road type and location, the traffic data at the same time point will show a certain degree of similarity. Hence, data from similar roads can be used to repair the missing values (Li, et al.,2014). In this study, the data of road2 in the period from 6:45 to 15:15 on October 12 was missing, so the data of road3 in the same time, which was strongly related to road2, was used to repair. In fact, these are two connected paths.

3.2.3 Time characteristic analysis

(1) Similarity

Combined with the traffic flow direction of this study is to follow the clockwise direction of the second ring road after entering through Intersection1. Firstly, the traffic flow and average speed of Intersection1 are detected respectively for five consecutive days (10/14-10/18). Data were acquired according to the interval of 15min, and 96 data were acquired every day. The data were consolidated to obtain Figure 3.2 below.

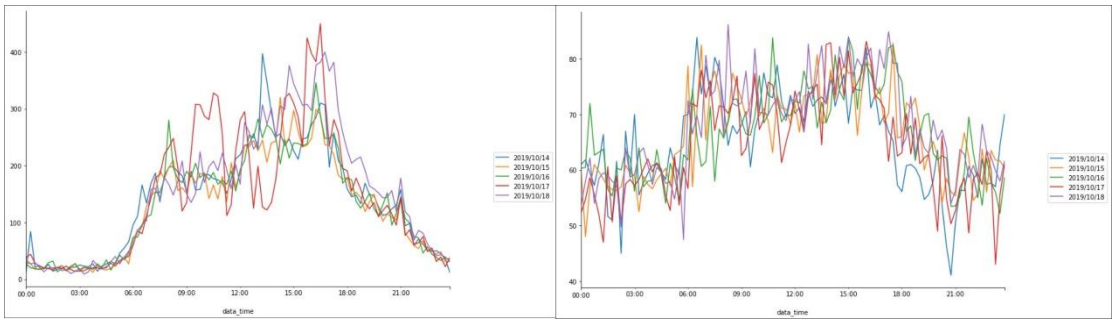


Figure 3.2: Weekday traffic flow variation (left) and average speed variation (right) at Intersection 1

Figure 3.2 (left) indicates that there is a certain degree of similarity in the time change of the traffic flow of the bayou over a five-day period. That is, the overall trend is stable, rising, morning peak, fluctuating rising, evening peak, and falling. Figure 3.2(right) shows that the average speed of five consecutive days also has a certain degree of similarity in time variation. That is, the overall trend is stable, the fluctuation increases, and the

fluctuation decreases.

For the location of the same road, the traffic operation changes show strong regularity. For example, the traffic flow on weekdays generally has two peak periods of morning and evening, while the usual traffic flow is generally smaller than the traffic flow on holidays. Therefore, the Pearson correlation coefficient ϱ is used to compare the similarity of the two-day time series.

The expression is as follows:

$$\rho_{x,y}=\frac{cov(X,Y)}{\sigma_x\sigma_y} \tag{3.3}$$

X and Y are random quantities.

Correspondingly, the formula of its sample correlation coefficient r is:

$$r=\frac{\sum_{i=1}^{n}(X_i-\bar{X})(Y_i-\bar{Y})}{\sqrt{\sum_{i=1}^{n}(X_i-\bar{X})}\sqrt{\sum_{i=1}^{n}(Y_i-\bar{Y})}} \tag{3.4}$$

n is the total number of data, X_i, Y_i are variables when X, Y are in i moment of the actual value. $\bar{X}, \bar{Y}$ respectively respectively is the mean of X and Y data. Note that r is in the range [-1,1]. When the value is 1, it is completely linear. When the value is -1, there is a completely negative linear correlation. A value of 0 means it is irrelevant.

Therefore, the similarity results of daily traffic flow and average speed on working days of bayonet are shown in Tables 3.1 and 3.2.

	2019/10/14	2019/10/15	2019/10/16	2019/10/17	2019/10/18
2019/10/14	1.000	0.930	0.934	0.794	0.911
2019/10/15	0.930	1.000	0.950	0.860	0.932
2019/10/16	0.934	0.950	1.000	0.848	0.919
2019/10/17	0.794	0.860	0.848	1.000	0.849
2019/10/18	0.911	0.932	0.919	0.849	1.000

Table 3.1: Correlation coefficient of traffic flow at Intersection 1

It can be found that the correlation coefficient of traffic flow in five consecutive days is large, except for the correlation coefficient between the 14th and 17th. Others are greater than 0.8 (r>0.8, highly correlated).

	2019/10/14	2019/10/15	2019/10/16	2019/10/17	2019/10/18
2019/10/14	1.000	0.661	0.516	0.687	0.633
2019/10/15	0.661	1.000	0.592	0.716	0.732
2019/10/16	0.516	0.592	1.000	0.662	0.700
2019/10/17	0.687	0.716	0.662	1.000	0.729
2019/10/18	0.633	0.732	0.700	0.729	1.000

Table 3.2: Average vehicle speed correlation coefficient at Intersection 1

Table 3.2 shows that r is moderately correlated between 0.4 and 0.6, and strongly correlated between 0.6 and 0.8. The correlation coefficient of the average speed of five consecutive days is basically between 0.5 and 0.7. That is, between the two degrees of correlation. Therefore, the traffic flow embodies extremely strong similarity in terms of the time dimension, and the average speed embodies similarity features of moderate intensity to strong correlation in terms of the time dimension.

(2) Uncertainty in the temporal characteristics of traffic flow

The traffic flow parameters will have large fluctuations in time, and the use of differential values can reflect the degree and uncertainty of random fluctuations. Due to the large fluctuation of traffic flow data, the difference method is used to ensure the data smooth.

Difference: The change in the function y=y(x) when the independent variable x is changed to x+1.

It should be noted that unnecessary differences will distort the time series model and reduce the prediction accuracy. Therefore, the first-order difference is used here, and the corresponding formula is as follows:

$$\Delta y_x = y_{x+1} - y_x \tag{3.5}$$

In this case, it is the difference between the time series at time t and t-1.

The study on traffic flow operation characteristics of Intersection 1 is continued. Although the coefficients for both October 14 and October 18 were large, they were excluded to reduce the influence of adjacent dates (weekends). Only three days of October 15, October 16, and October 17 were analyzed.

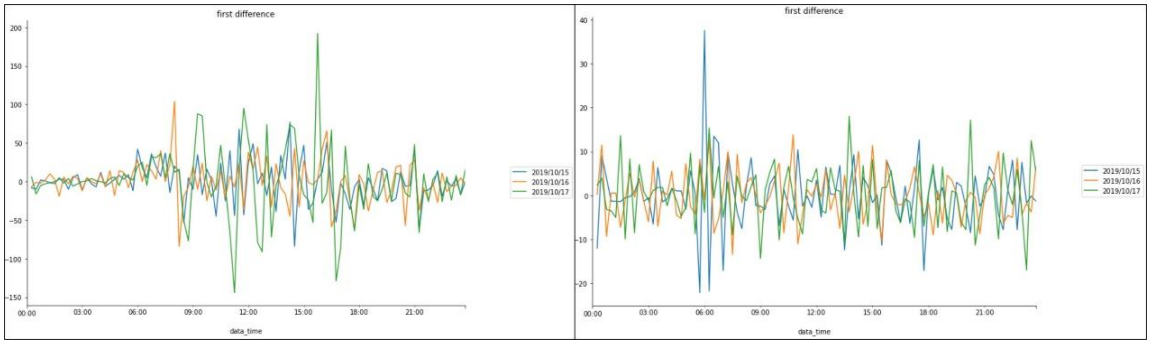


Figure 3.3: First-order difference map of traffic flow (left) and first-order difference map of average speed (right) at Intersection 1

It can be seen from Figures 3.3 that the traffic flow parameters on this road section all have fluctuations between certain ranges as time changes. For traffic flow, the data volatility is particularly pronounced in the time periods of 8:00-9:00, 11:00-13:00 and 16:00-18:00. For the average speed, the data volatility is particularly pronounced between 5:00 and 7:00. These sections are in the morning and evening peak periods of the road section, so there will be large fluctuations compared with other time periods. On the one hand, this is in line with reality. On the other hand, although there is a certain similarity between the two parameters in the time domain, the changes, and fluctuations of the data at each moment are different and random.

In addition, by comparing the two figures, the volatility of October 16 is maintained in a small stationary range. Therefore, October 16 is selected as the time point for the following periodic analysis and spatial analysis.

（3）Periodicity

According to the above, the traffic flow time series of the third working day of a week is used as the analysis data. Similarly, Intersection 1 was taken as the study road section, and the following two graphs were obtained by selecting Wednesday of this road section for five consecutive weeks (16/10/2019-13/11/2019).

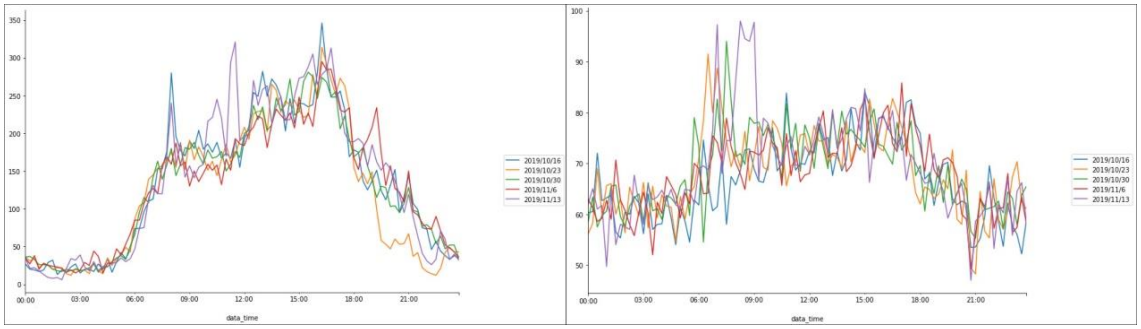


Figure 3.4: Periodic traffic flow variation (left) and periodic average vehicle speed variation (right) at Intersection 1

It is observed that the time-varying trend of traffic flow time series variation phenomenon is roughly the same even for the same working day in different weeks. The traffic flow parameters show relatively strong periodic characteristics with the change of time. As a result, the similarity coefficients of the same working day in different weeks are shown in Table 3.3

	2019/10/16	2019/10/23	2019/10/30	2019/11/6	2019/11/13
2019/10/16	1.000	0.931	0.954	0.931	0.928
2019/10/23	0.931	1.000	0.942	0.905	0.897
2019/10/30	0.954	0.942	1.000	0.938	0.934
2019/11/6	0.931	0.905	0.938	1.000	0.915
2019/11/13	0.928	0.897	0.934	0.915	1.000

	2019/10/16	2019/10/23	2019/10/30	2019/11/6	2019/11/13
2019/10/16	1.000	0.551	0.551	0.689	0.533
2019/10/23	0.551	1.000	0.586	0.551	0.557
2019/10/30	0.551	0.586	1.000	0.721	0.650
2019/11/6	0.689	0.551	0.721	1.000	0.592
2019/11/13	0.533	0.557	0.650	0.592	1.000

Table 3.3: Correlation coefficient of periodic traffic flow (left) and correlation coefficient of periodic average vehicle speed (right) at Intersection 1

As can be seen from the above table, both the traffic volume and the average speed are corresponding to the results obtained by the preorder similarity analysis: the correlation coefficients of the traffic flow are all greater than 0.8, indicating that the traffic flow shows the periodic characteristics of strong correlation with time. However, the correlation coefficient of the average speed is still between 0.5 and 0.8, which indicates a moderate or strong correlation in the period.

3.2.4 Spatial characteristic analysis

(1) Similarity of adjacent road segments for spatial characteristics of traffic flow

Because the urban road network is a complex system, the influence of its internal structure will form a dynamic transfer, so the traffic flow parameters in adjacent Spaces have similar characteristics of traffic flow state, which is the transfer characteristics of traffic operation

in space. Aiming at the traffic analysis of the second ring road in the city, four sections from entering the second ring road in Guandu District to following the whole second ring road around the second ring road are selected for analysis. This includes an intersection and three road segments, Intersection1, road1, road2, and road3. The traffic data of the four adjacent road sections on October 16, 2019 were obtained, and the time series of the traffic parameters of the four roads were made in Figure 3.5, and the respective correlation analysis was carried out (see Table 3.4 for details).

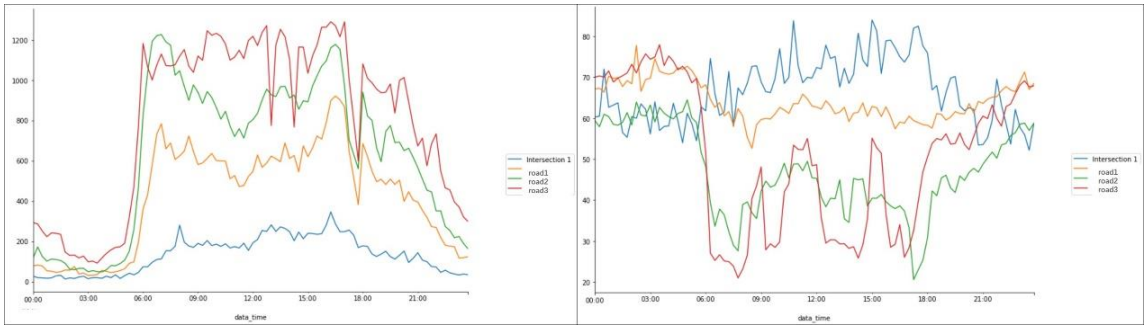


Figure 3.5: Time series of traffic flow (left) and average speed (right) in four adjacent sections of the second Ring Road

	Intersection	section 2	section 3	section 4
intersection1	1.000	0.894	0.837	0.854
road1	0.894	1.000	0.955	0.926
road2	0.837	0.955	1.000	0.934
road3	0.854	0.926	0.934	1.000

	Intersectio	section 2	section 3	section 4
intersection1	1.000	-0.583	-0.660	-0.601
road1	-0.583	1.000	0.836	0.749
road2	-0.660	0.836	1.000	0.833
road3	-0.601	0.749	0.833	1.000

Table 3.4: Correlation analysis of traffic flow (left) and correlation analysis of average speed (right) in four adjacent sections of the Second Ring Road

It can be seen from Figure 3.5 (left) that the time series diagram of traffic flow of adjacent sections basically tends to the same trend and has strong similarity. Traffic increased sharply from 5:00 to 8:00, changed slightly from 8:00 to 16:00 in a certain range, decreased sharply from 16:00 to 18:00, and decreased gently from this time.

As can be seen from Figure 3.5 (right), the average speed also has distinct peak hours and almost the same interval segments. However, there is a negative correlation in the spatial similarity. Road1, road2, the space of road3 strong similarity to each other related or relevant, but negatively related to the Intersection1 space. This is determined by the road function and location of the four observation points in the road network. Intersection1 is an observation point on the urban expressway and located near the urban entrance and

exit. After the traffic flow passes through Intersection1, there are multiple traffic flows (see Figure 3.1 for details). Section1, section2 and section3 are all observation points on the urban ring road. This indicates that roads in adjacent Spaces will affect the degree of spatial similarity of traffic flow due to their different functions.

Through the analysis, the three roads as adjacent sections appear increasing or decreasing trend in the same period, and the traffic characteristic values (traffic flow and average speed) of the three roads analyzed by the correlation coefficient have strong correlation. This indicates that the traffic variation trend of the three adjacent sections throughout the day is similar, and the urban traffic operation has a near-area spatial similarity transfer rule.

(2) Spatial stationarity analysis of traffic flow

Traffic parameters will fluctuate in a certain period, but there will be a certain stationarity between the fluctuation and the expected value. The spatial stationarity of traffic flow reflects the running condition of road traffic and the stationarity of the change law in a certain space. Due to the different function, type and spatial location of urban roads, the traffic parameters and fluctuation range of different roads will be different. Therefore, the spatial stationarity analysis of traffic flow is a measure of the dispersion of traffic data, which is used to measure the difference between the traffic time series value and the average value. Standard deviation is the arithmetic square root of the variance and can be understood as how much the data fluctuates around the mean.

Four observation points Intersection1, road1, road2 and road3 were selected to analyze their distribution characteristics and dispersion degree from two aspects of traffic flow and average speed.

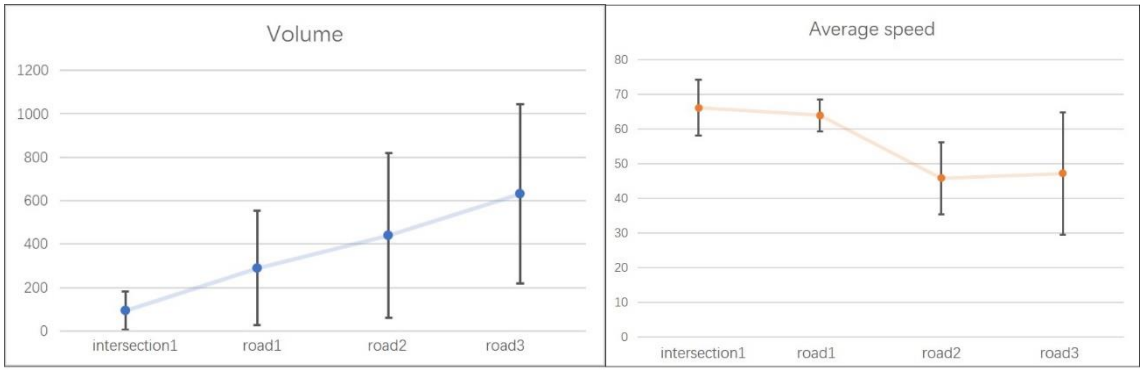


Figure 3.6: Spatial stationarity analysis of traffic flow (left) and average speed (right) in four adjacent sections of the second Ring road

In summary, through the analysis of the traffic flow time series variation trend and correlation analysis of the same road section, it can be concluded that the traffic flow and average speed of the same road section have similarity, periodicity, and uncertainty in the time dimension. In the spatial dimension, through the analysis of the traffic flow time series change trend of adjacent sections, correlation analysis, stationarity analysis and so on. It can be concluded that the spatial correlation degree of adjacent roads is high, and that the traffic parameters exhibit similar transfer characteristics in the near area. However, the traffic parameters are unstable. The above research provides perfect data feature information for the subsequent traffic parameter prediction and traffic state evaluation.

Chapter 4
Model building

4.1 1D CNN

Convolutional neural network is a multi-layer supervised learning network that is used to process grid-like data, such as time series data and image data. The CNN structure mainly consists of convolutional layer, pooling layer, and fully connected layer, as shown in Figure4.1.

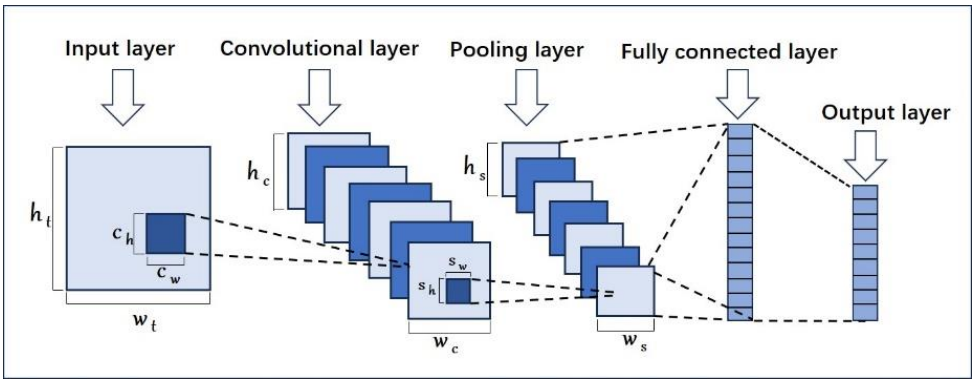


Figure 4.1: CNN Architecture

For this paper, the data volume is small and the model is simple, so a variant of convolutional neural network, 1D CNN, is used. Because the convolution kernel of one-dimensional CNN will only be convoluted along the time step order, it is convenient to construct a hybrid model that can process two parameters of three adjacent paths at the same time to capture spatial features. The principle of 1D CNN is similar to the traditional CNN, except for the difference in data processing. The convolution kernel of a 1D CNN will only perform convolution along the time step order: the input data is one-dimensional, such as time series data, and the convolution operation can capture the temporal characteristics in the sequence data.

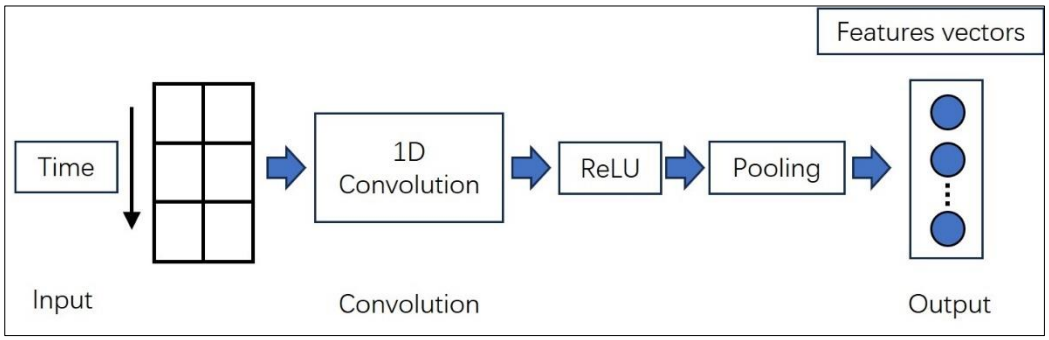


Figure 4.2: Diagram of the structure of a 1D CNN

According to the figure above, 1D-CNN is mainly composed of the following parts:
(1) Input layer: receives 1D sequence data as input to the model.

(2) Convolutional layer: Uses a series of trainable convolutional kernels to slide over the input data and extract features. The convolution operation is able to effectively extract local information to capture the local patterns of the input sequence.

(3) Activation function: nonlinear transformation is applied to the output of the convolutional layer to enhance the expression ability of the model. Commonly used activation functions are ReLu and sigmoid (Qin et al., 2023).

(4) Pooling layer: Reduces the amount of computation and improves the generalization ability of the model by reducing the dimensionality of the convolutional layer output.

The main part of this is convolutions and pooling. To learn more about 1D convolutions, activation functions, and pooling.

1) The essence of convolution is the effect of inversion, sliding (corresponding to multiplication), and superposition

The mathematical expression (in continuous form) is:

$$f * g = \int_{-\infty}^{\infty} f(\tau) g(t-\tau) d\tau \tag{4.1}$$

In practice, the formula of one-dimensional convolution is as follows:

$$Length\ of\ the\ convolution\ result = data\ length + Convolutional\ kernel\ length\ -1 \quad (4.2)$$

2) The $Relu$ activation function has low computational complexity, fast convergence speed and is in the x>0 region. Coupled with no vanishing gradient, is a commonly used function. Corresponding formula:

$$Relu = max(0, x) \quad (4.3)$$

The $sigmoid$ activation function compresates the range of values to (0,1), which fits well with the probability distribution characteristics and can be used in the output layer of probabilistic prediction. The formula is as follows.

$$sigmoid(x) = \frac{1}{1+exp(-x)} \quad (4.4)$$

3) Pooling is to reduce the size of data. There are mainly Max pooling and average pooling. One of the most used is Max pooling.

One-dimensional Max pooling process: in the convolution results, the maximum value of adjacent results is extracted with a step size of 2, which reduces the features, leads to a reduction of parameters, and then simplifies the complexity of the convolution network calculation.

If the data length is 10 and the convolution result is the same length as the data length, the pooling result is of length 5.

4.2 BiLSTM

BiLSTM is an extended form of LSTM, so it is important to learn LSTM first. The BiLSTM structure includes two unidirectional LSTMS, stacked up and down (Zhuang and Cao, 2022). In addition, LSTM improves the defects of RNN that gradient disappears or cannot be relied on for a long time due to expansion (Erdem, 2020). The hidden layer of LSTM structure is equipped with linear self-cycling storage units, which can improve the information retention ability between distant nodes. The gating link of LSTM is divided into three parts: forgetting, input and output, and a state unit is introduced to coordinate the operation of the whole network.

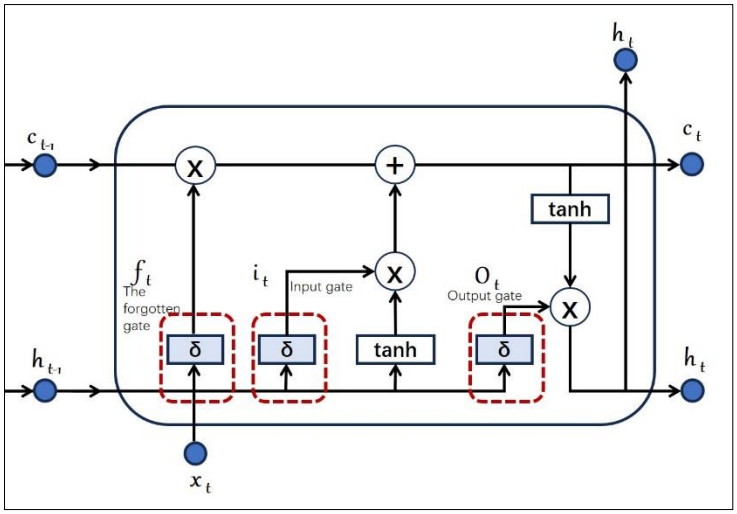


Figure 4.3: LSTM Architecture

(1) Forget gate: The information with low correlation between input and output is forgotten to facilitate the memory of subsequent new information.

$$f_t=\sigma\left(\mathrm{W_f} * [h_{t-1}, x_t] + b_f\right) \tag{4.5}$$

Where σ is the sigmoid function, the weight matrix $\mathrm{W_f}$ and bias $\mathrm{b_f}$, h_{t-1} is the output state at time t-1, and x_t is the input at time t.

(2) Input gate: it selectively absorbs the useful information extracted from the forget gate and the information received at the current time.

Input gate formula:

$$i_t=\sigma(\mathrm{W_i} * [h_{t-1}, x_t] + b_i) \tag{4.6}$$

The formula for the candidate cell state $\tilde{C}_t$ is:

$$\tilde{C}_t=\tanh(\mathrm{W_C} * [h_{t-1}, x_t] + b_C) \tag{4.7}$$

Update the memory, the new cell state formula:

$$C_t=f_t*C_{t-1}+i_t*\widetilde{C_t} \tag{4.8}$$

The activation function tanh, f_t*C_{t-1} is the retained information at time t-1, and $i_t*\widetilde{C_t}$ is the remembered information at time t.

(3) Output gate: The updated information is selected from the cells of the input gate and output.

The formula for O_t is:

$$O_t=\sigma(W_O * [h_{t-1}, x_t] + b_O) \tag{4.9}$$

The formula for h_t is:

$$h_t=O_t*\tanh(C_t) \tag{4.10}$$

The final output is computed by both the output gate and the current memory information.

BiLSTM contains both forward and backward LSTM unit structures, and connects and influences each other through a shared weight matrix, which can better extract the bidirectional change features of time series. It can avoid gradient disappearance and explosion while solving the long-term dependence problem. The BiLSTM structure is shown in Figure 4.4.

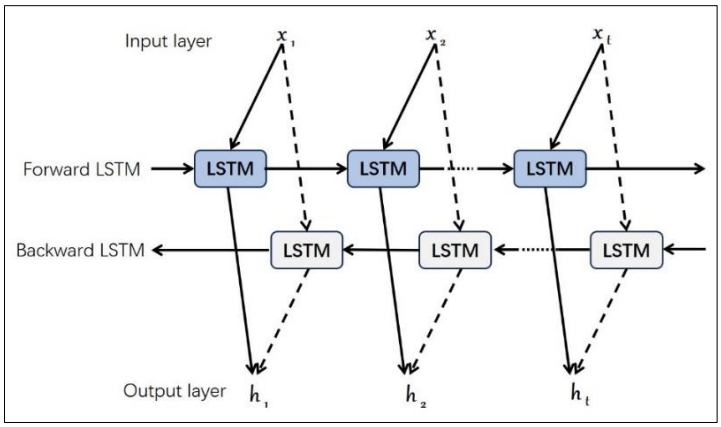


Figure 4.4: BiLSTM architecture

The forward LSTM structure calculation process in BiLSTM network is similar to the single LSTM calculation process. The hidden layer state of BiLSTM network is obtained by combining the forward hidden layer state and the backward hidden layer state, and its calculation formula is as follows:

$$\overrightarrow{h_t}=LSTM\left(x_t, \overrightarrow{h_{t-1}}\right) \tag{4.11}$$

$$\overleftarrow{h_t}= LSTM\left(x_t, \overleftarrow{h_{t-1}}\right) \tag{4.12}$$

The output of a BiLSTM usually consists of the concatenation of hidden states from both directions:

$$h_t=\alpha\overrightarrow{h_t}+\beta\overleftarrow{h_t} \tag{4.13}$$

Where: x_t, $\overrightarrow{h_t}$, $\overleftarrow{h_t}$ are the input, forward LSTM hidden layer output and backward LSTM

hidden layer output at time t, respectively LSTM is the LSTM unit. Both α and β are constant coefficients, which are the weights of $\overrightarrow{h_t}$, $\overleftarrow{h_t}$.

Using BiLSTM to capture temporal features can capture more complete information. According to the personal understanding, in a one-way LSTM, the n-th neuron can only extract its own information and the n-1 information (including n-1, the information left by n-1 previous neurons), which is related to n before. A two-way loop runs the other way around to capture more information. Algorithmic logic: The past influences the future, the future also has an effect on the past.

4.3 1D CNN+BiLSTM mode

Based on the antecedent analysis, the traffic parameters of a road network are related to its historical traffic parameter data and affected by the traffic parameters of its adjacent locations. The spatial characteristics of the local road network play a significant role in determining the traffic flow. In the constructed combination model of short-term traffic flow prediction, the CNN module in CNN-BiLSTM network adopts a single-layer one-dimensional convolutional layer and pooling layer structure (Ranjan et al., 2020). The spatial characteristics of short-term traffic parameters are captured through the convolutional layer, and the feature information is sent to the pooling layer for dimension reduction, and then transformed through the fully connected layer and output to the BiLSTM network. BiLSTM trains and learns the received bidirectional time series features, extracts the spatial features to capture the temporal characteristics between vectors, and backpropagates according to the calculated network output, and the final traffic flow prediction results are output by the fully connected layer.

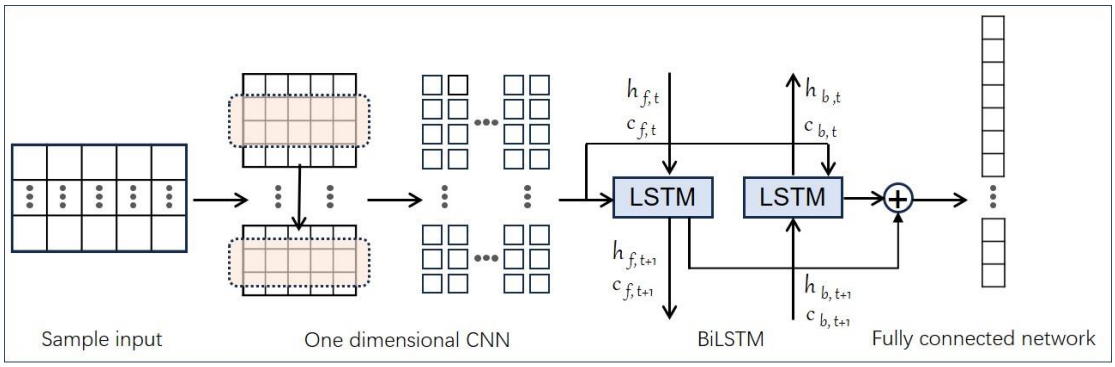


Figure 4.5: 1D CNN+BiILSTM model architecture

4.3.1 Complete structure and functional setup of CNN+BILSTM model

(1) Data preprocessing

The spatial characteristics of traffic flow are reflected in the "upstream and downstream" relationship in its spatial location. Generally, the traffic flow at one bayonet is often affected by the traffic flow at its adjacent bayonet. Indeed, for the second Ring viaduct expressway in Kunming City, due to the potential connection between the two parameters, the two parameters are used as the input of the model at the same time instead of running independently. In the clockwise direction, from the second ring West Road to north from south, entering the second ring North Road from west to east, and entering the second ring East Road from north to south, the data of the two parameters that have been processed for missing values in the three sections of the Second Ring West Road (road1), the second ring North Road (road2) and the second ring East Road (road3) are merged and then preprocessed. Specifically, the merged data with dimension 6 is normalized. Here, Min-Max Normalization is used to linearly map the data to the [0, 1] interval using the following formula:

$$x^*=\frac{x-min}{max-min} \tag{4.14}$$

x and x^* are the values before and after the transformation, and max and min are the maximum and minimum values of the samples, respectively.

Then split the processed time series data set 4:1. The first 80% of the dataset (the first 42 days from October 8 - November 19) was used as the training set, and the last 20% (the 11 days from November 20 - November 30) was used as the test set.

(2) 1D CNN Module

The data features of traffic flow and average speed of three connected roads were extracted by convolution kernels through the convolution layer to capture the spatial characteristics of traffic parameters (road1, road2, road3 are adjacent roads that affect each other). The Max pooling layer is used to control the convergence of the neural network. Meanwhile, the activation function ReLU is used to fit the nonlinear relationship between the two parameters and generate the feature vector (Qin et al., 2023).

In the 1D convolutional layer of the CNN module, the convolution step size is set to 1, the

fill value is 3, and the size of the convolution kernel is 3*3. The pooling method of Max pooling is used in the pooling layer structure, the pooling kernel size is set to 2*2, the step size is 2, and the fill value is 0.

MaxPooling1D layer is used as the pooling layer

The input is L_{In}, and the output is calculated:

$$L_{out}=\left[\frac{L_{In}+2*padding-dilation*(kernel\_size-1)-1}{stride}+1\right] \tag{4.15}$$

Finally, the CNN sequence is changed.

(3) BILSTM module

The neural network layer in LSTM is 2 and the hidden layer is 64.

(4) Fully connected layer

The sigmoid function is used to transform the data into the output on (0,1).

(5) Model training

Batchsize is set to 32, the Adam optimization function is used, the learning rate is set to 0.001, and the Epoch is 20.

The Adam optimization algorithm is helpful to accelerate the convergence of the model and update the parameters smoothly. The learning rate of each parameter is adaptively adjusted by exponentially weighted moving average and adjusting Adam according to the first-order matrix and second-order matrix estimates of its gradients.

The learning rate is usually set to a small positive number, one of 0.001, 0.0001, 0.00001, or 0.000001. The initial learning rate can be chosen first, and if the loss is decreasing too fast, there is a requirement to decrease the learning rate. Otherwise, the learning rate is appropriately increased.

Epoch represents the number of complete traversals of the dataset during model training. Through multiple "epochs", the model gradually adapts to the training data, improving the performance, and the generalization ability of the model is monitored by validating the performance.

The loss function is a function that measures the difference or error between the predicted results of the model and the actual labels during training. It is the core of the optimization algorithm, which adjusts the parameters of the model by minimizing the loss function so that the model can better fit the training data. To facilitate further calculation of error loss calculation, the loss function is chosen directly as the evaluation index. That is, MSE is both the loss function and the evaluation metric. (Details are in the result analysis section).

4.3.2 The complete code idea of the model running process

(1) Data processing:

1) Merge the speed and flow of the three roads

2) Normalize the data (min-max method)

3) Through the sliding window method, the data is split into features and labels, that is, the speed and flow of the first five time points (08/10/2019 00:00-1:00) are used as the model input, and the speed and flow of the next time point are used as the label.

4) The data is split into training and test sets. In order of time, the first 80% of the data is used as the training set. The last 20% is used as the test set.

(2) Model structure:

The model consists of three parts: (1) 1D CNN; (2) BiLSTM (3) Fully connected layer (output layer)

1) 1D convolutional Neural Network:

① Convolutional layer: input channel is 6; The output channel is 6, the convolution kernel is 3, the padding is 3, and the step size is 1

② The ReLu activation function

③ Max pooling layer: convolution kernel is 2, step size is 2, padding is 0

2) Bidirectional LSTM layer neural network:

① The input dimension is 6, the hidden layer dimension is 64, and the neural network layer is 2

3) Fully connected layer

The input dimension is 128 and the output dimension is 6

(3) Training process:

```
Epoch [1/20], Train Loss: 0.0300 , Test Loss: 0.0088
Epoch [2/20], Train Loss: 0.0081 , Test Loss: 0.0069
Epoch [3/20], Train Loss: 0.0063 , Test Loss: 0.0047
Epoch [4/20], Train Loss: 0.0059 , Test Loss: 0.0047
Epoch [5/20], Train Loss: 0.0056 , Test Loss: 0.0043
Epoch [6/20], Train Loss: 0.0053 , Test Loss: 0.0041
Epoch [7/20], Train Loss: 0.0051 , Test Loss: 0.0044
Epoch [8/20], Train Loss: 0.0051 , Test Loss: 0.0040
Epoch [9/20], Train Loss: 0.0050 , Test Loss: 0.0040
Epoch [10/20], Train Loss: 0.0049 , Test Loss: 0.0039
Epoch [11/20], Train Loss: 0.0049 , Test Loss: 0.0041
Epoch [12/20], Train Loss: 0.0048 , Test Loss: 0.0038
Epoch [13/20], Train Loss: 0.0048 , Test Loss: 0.0037
Epoch [14/20], Train Loss: 0.0047 , Test Loss: 0.0043
Epoch [15/20], Train Loss: 0.0046 , Test Loss: 0.0036
Epoch [16/20], Train Loss: 0.0047 , Test Loss: 0.0041
Epoch [17/20], Train Loss: 0.0045 , Test Loss: 0.0035
Epoch [18/20], Train Loss: 0.0045 , Test Loss: 0.0035
Epoch [19/20], Train Loss: 0.0044 , Test Loss: 0.0038
Epoch [20/20], Train Loss: 0.0044 , Test Loss: 0.0035
```

Figure 4.6: Results of 20 Epochs of training

Chapter 5
Analysis of experimental result

5.1 Model evaluation

The mean squared error is the most used regression loss function and is calculated as follows:

$$F_{loss}=\mathrm{MSE}=\frac{1}{n}\sum_{i=1}^{n}(y_i-\tilde{y}_i)^2 \tag{5.1}$$

n is the number of samples, y_i is the true value, and $\tilde{y}_i$ is the predicted value.

The higher the value of the above formula, the worse the prediction effect.

Training process:

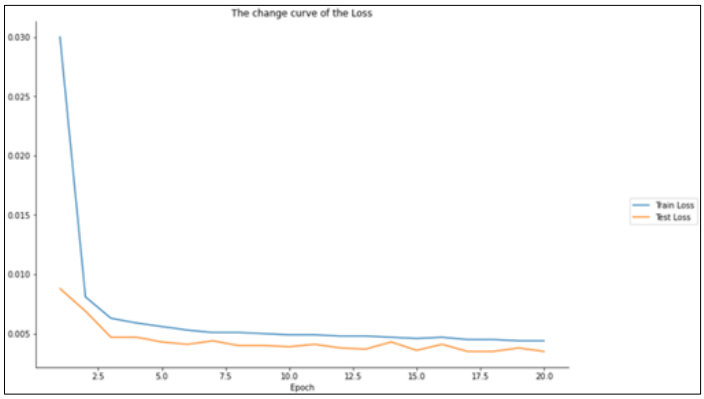


Figure 5.1: Plot of the training results

In the obtained results, both Train loss and Test loss are gradually decreasing as a whole, and train loss shows a sharp decrease from Epoch1 to Epoch2. Then around Epoch6, both tend to be flat, which means that the model has basically finished learning. This result indicates that the model is learning and gradually converging. In addition, comparing Train loss and Test loss, the difference between them decreases sharply and stabilizes at around 0.001. Therefore, this trained model does not overfit and gradient explosion, which falls within a reasonable range. Finally, the MSE (or Test Loss) on the test set is small, ranging from 0.0088 to 0.0035, with an average error value of 0.004385.

5.2 Prediction results of short-term traffic parameter model

Figure 5.2 shows the overall training and testing trend graphs of road1, road2, and road3 respectively (from top to bottom, road1, road2, and road3 respectively; the left graphs are all traffic, and the right graphs are all average speed).

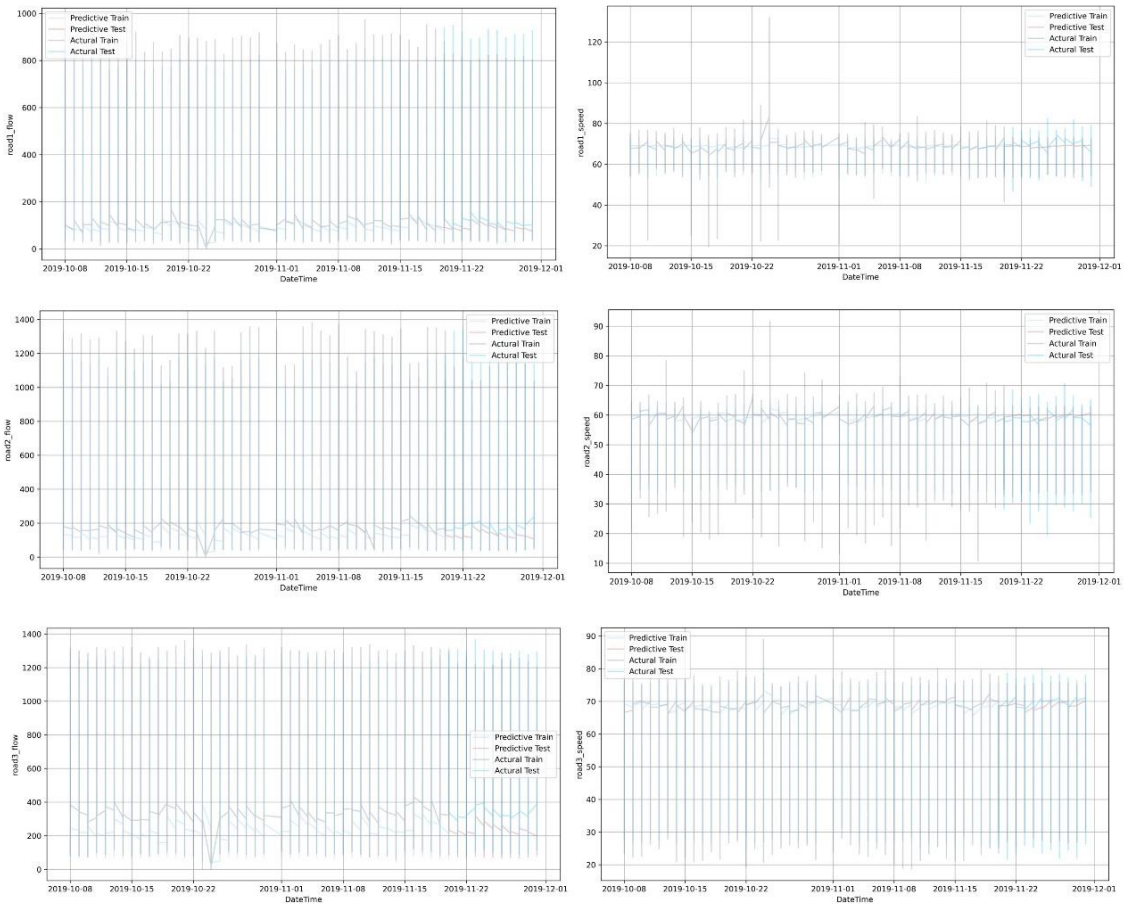


Figure 5.2: Overall training and testing trend plots for road1, road2, and road3 for 53 days

Looking at the actual prediction and real data changes of the test set as a whole, most of the predicted change trends are similar to the actual ones, and are positively correlated. The gap between the predicted average speed and the real average speed is small, and the predicted traffic is generally smaller than the real traffic, especially road3. Since the dataset is small, the test data is used as validation data. The last day of the test set (November 30th) is analyzed to evaluate the performance and generalization ability of the model.

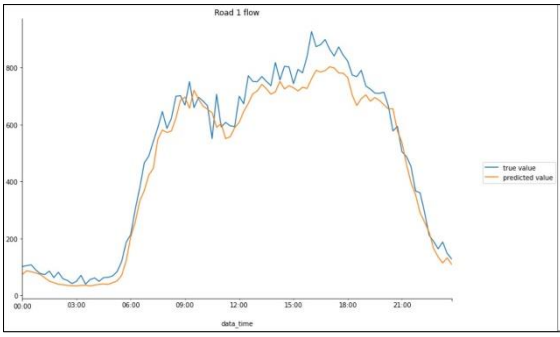


Figure 5.3: road1 predicted traffic change trend diagram

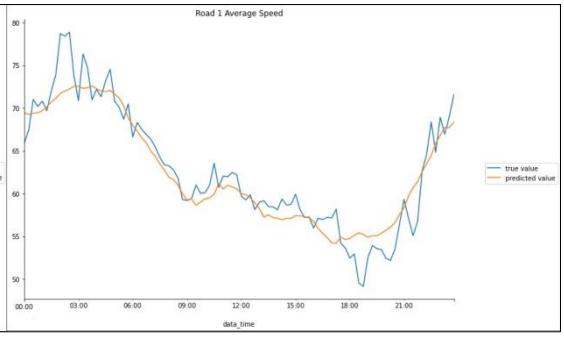


Figure 5.4: road1 predicted average speed change trend graph

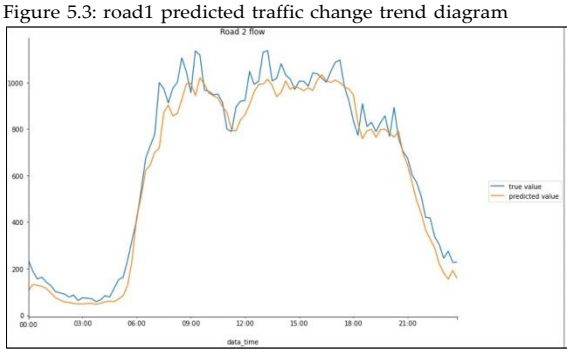


Figure 5.5: road2 predicted traffic change trend diagram

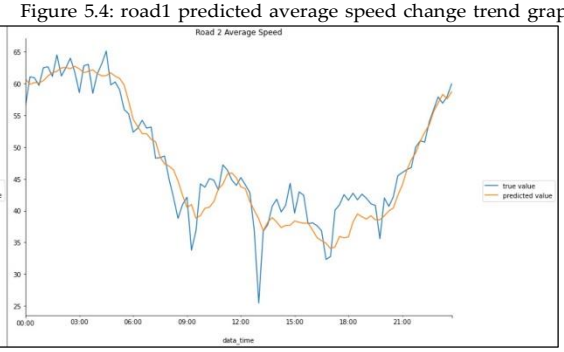


Figure 5.6: road2 predicted average speed change trend graph

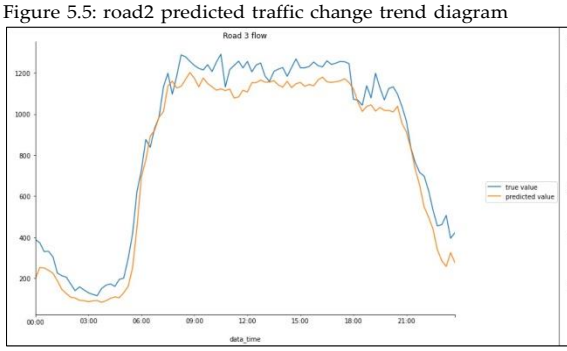


Figure 5.7: road3 predicted traffic change trend diagram

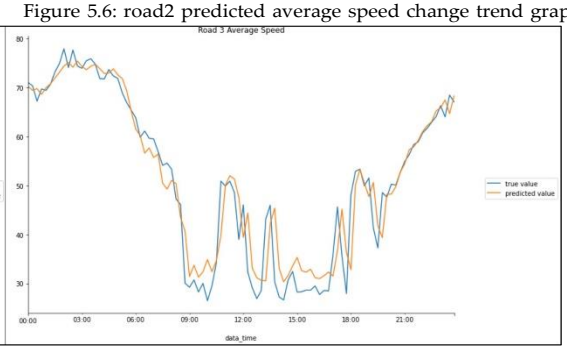


Figure 5.8: road3predicted average speed change trend graph

By comparing the predicted values of the three roads with the real values, it is found that the average speed generally predicts the change trend and has a small difference with the real data (Note: the starting value of the Y-axis of the average speed on the right side is different to facilitate the observation of the change). The trend of traffic flow is also predicted but generally lower than the true value (which will affect the subsequent determination of traffic congestion). These conclusions correspond to those obtained above based on the complete data set.

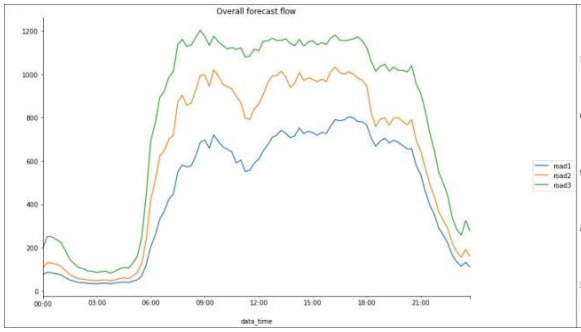


Figure 5.9: Overall predicted traffic trend graph

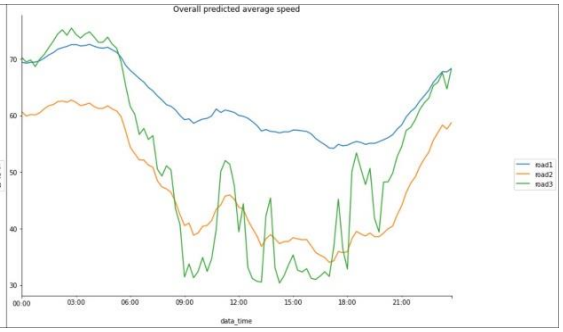


Figure 5.10: Overall predicted average speed trend graph

The overall true image is close to the prediction, and here the overall predicted image is directly analyzed. Several conclusions of preorder analysis are verified :(1) There is periodicity. The change trend of the two graphs is similar to that of the three roads obtained by spatial analysis in the previous sequence, which is consistent to a certain extent. (2) The flow correlation of adjacent road sections is extremely strong, and the average speed correlation is between strong correlation and extremely strong correlation. This also indicates that it is effective to consider the influence between roads and between two parameters when constructing the model.

In addition, it can be seen from the left figure and the grid diagram of preorder analysis that road1 is the entrance to the city from the southwest direction, so the flow is the largest. Then it is passed to road2 and road3 clockwise, so the flow is reduced in turn. By observing the right figure and the grid diagram of preorder analysis, there are many high-speed intersections around road2 and road3, which shows that the average speed of road2 and road3 will decrease sharply. Moreover, due to the existence of multiple peaks, the peak sizes of the three roads are different, but all indicate that there may be a peak. Therefore, it is of practical significance to analyze the congestion.

5.3 Running state analysis: short-term traffic congestion prediction

Identifying the running situation of expressways is very important to improve the level of urban traffic management and traffic service quality. Traffic congestion index is an objective indicator to characterize the degree of traffic congestion (He et al., 2016).

$$Congestion\ index = \left(\frac{Actual\ traffic\ flow}{Road\ capacity}\right) * \left(\frac{Average\ speed}{Free\ flow\ speed}\right) \quad (5.2)$$

(1) Road capacity: the maximum number of vehicles that the road at the section can withstand. The capacity of each lane of domestic expressways is 1350 vehicles.

(2) Actual traffic volume: the number of vehicles passing through a section of the road in a specific time period. This is the traffic flow per unit time (15min).

(3) Average speed: the average speed of the vehicle in the process of driving.

Level of operation	Unimpeded	Basic smooth	Light congestion	Moderate congestion	Heavy congestion
Expressway	V>65	50<V≤65	35<V≤50	20<V≤35	V≤20
Main road	V>40	30< V≤40	20<V≤30	15<V≤20	V≤15
Secondary arterial road, branch road	V>35	25< V≤35	15<V≤25	10<V≤15	V≤10
Note: V represents the average travel speed of the road segment					

Table 5.1: Classification of road section traffic operation levels

(4) Free-flow speed: the average speed of vehicles in the absence of congestion. The adopted congestion threshold indicators for the average vehicle speed of different classes of roads are shown in Table 5.1 above. Road1, 2 and 3 are all urban expressways, so the free flow speed adopted is 50 km/h.

The congestion index has a range of values between 0 and 1, where 0 indicates no congestion and 1 indicates extreme congestion. According to the actual situation of our country, it is generally considered that congestion index greater than 0.6 belongs to the range of congestion.

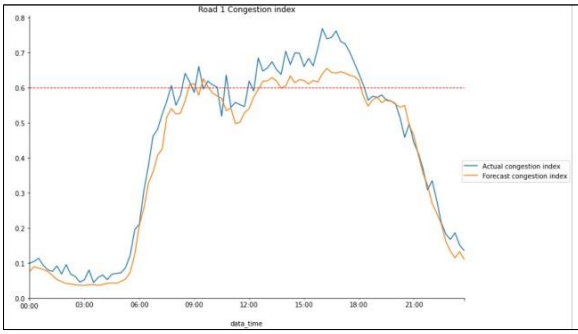


Figure 5.11: Map of road1 congestion index

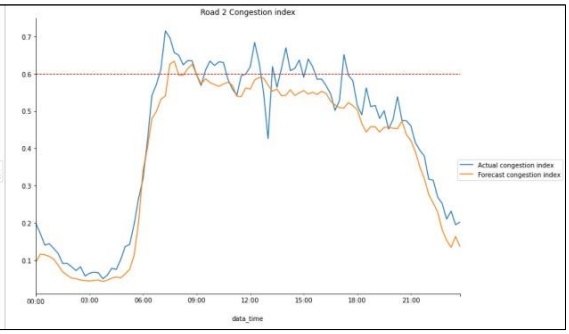


Figure 5.12: Map of road2 congestion index

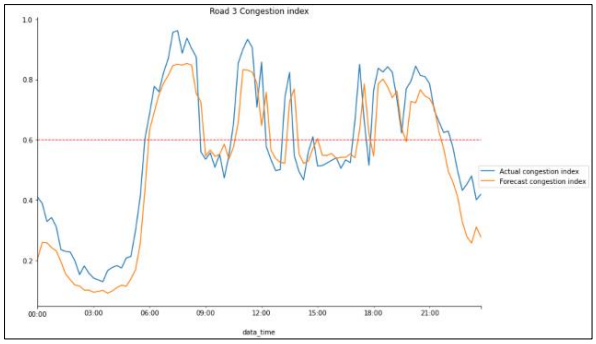


Figure 5.13: Map of road3 congestion index

The evaluation index accuracy of the classification model is calculated:

$$Accury=\frac{(TP+TN)}{(TP+TN+FP+FN)} \tag{5.3}$$

TP is True Positive, TN is True Negative, FN is False Negative, and FP is False Negative. Four cases form the confusion matrix.

Except for the small congestion at 10:45, the other congestion conditions of Road 1 are successfully predicted. In terms of output congestion pattern probability, the accuracy of the model is 87.5%. Road2 only predicted the first congestion situation successfully. The following four jams were not predicted because the congestion index was lower than the target. This primarily occurs due to the traffic predictions being typically lower than the actual traffic counts, which leads to a reduced congestion index. In terms of output congestion pattern probability, the accuracy of the model is 78.1%. The five congestion situations of Road3 are predicted accurately, and the corresponding model output congestion accuracy is 85.4%.

There is a slight lag in the three road predictions, and all of them have a small congestion index. According to the running table, for Road 1 which is unblocked for a long time and Road 3 which is slightly congested for a long time, the congestion index changes dramatically. That is, most of the congestion can be predicted. However, for Road 2 which is normally unblocked, the congestion index is often around the critical value of 0.6. Many congestion indices slightly higher than 0.6 will also be slightly lower than 0.6 due to low traffic, which makes it impossible to effectively predict congestion.

Therefore, based on the set traffic capacity, free flow speed and the corresponding

congestion index formula, congestion prediction is not effective for highways with different operating states. This is also one of the shortcomings in the study.

Based on the above results, we get that the prediction of the whole model is better for the road sections with variously state (unblocked or lightly congested).

Chapter 6
Conclusion

Starting from the spatio-temporal distribution mechanism of urban traffic parameters, this paper analyzes the characteristics and limitations of the current short-term traffic parameter prediction model. Based on the improved neural network model, it uses the full-time data set of the real traffic environment to test, and adjusts the model parameters to form a combination model of short-term traffic parameter prediction. The main conclusions include the following three aspects:

(1) A CNN-BiLSTM short-term traffic parameter prediction model is constructed.
According to the spatio-temporal characteristics of traffic parameters, the one-dimensional CNN model can effectively extract spatial data features, and improve the comprehensively of spatial information by applying the potential connection between the two parameters. Furthermore, LSTM model can effectively extract time series information. In order to consider bidirectional timing information simultaneously, the BiLSTM model runs the LSTM in reverse. Consequently, the CNN-BiLSTM short-term traffic parameter prediction model can fully capture the spatio-temporal information representation of road traffic in the actual situation.

(2) The effectiveness of the short-term traffic parameter prediction model is verified.
By using the traffic data of three consecutive sections of the Second Ring Road in Kunming City, the training set and test set were constructed based on the 53-day data set and the results were verified. This enables the simultaneous use of two intrinsically related parameters for short-term traffic prediction. The results show that the model is learning and gradually converging, and there is no overfitting and gradient explosion.

(3) CNN-BiLSTM model has certain applicability for traffic operation state prediction.
It is very important to identify the running situation of expressway to improve the level of urban traffic management and the quality of traffic service. Congestion index is used as an

objective index to characterize the degree of traffic congestion, and the prediction data of two parameters are used to calculate the congestion index. The findings demonstrate that the accuracy of the CNN-BiLSTM model in predicting congestion probability is approximately 78% to 87%, which is indicative of its considerable applicability.

Deficiencies of this paper:

(1) In this study, the traffic parameters of three consecutive sections of the Second Ring Road are predicted. These three adjacent sections have strong traffic correlation on the urban road network. Due to the limitation of the data set, the study still lacks an LSTM model that considers the upstream and downstream information of traffic flow from the perspective of the overall road network, and the prediction results will be more accurate.

(2) Since there is no simulation training for a single CNN or BiLSTM model, it is impossible to explain whether the prediction performance of CNN-BiLSTM model combined with spatio-temporal features is good without reference loss value.

(3) The Road2 model is unable to accurately predict congestion levels. In other words, the model's ability to predict congestion is limited, and only those roads with clearly discernible changes in operating state are accurately predicted.

Bibliography

Bogaerts, T. *et al.* (2020) 'A graph CNN-LSTM neural network for short and long-term traffic forecasting based on trajectory data', *Transportation Research Part C: Emerging Technologies*, 112, pp. 62–77. Available at: https://doi.org/10.1016/j.trc.2020.01.010. (Accessed: 25 October 2023).

Dai, G., Ma, C. and Xu, X. (2019) 'Short-term traffic flow prediction method for urban road sections based on space-time analysis and GRU', *IEEE Access*, 7, pp. 143025–143035. Available at: https://doi.org/10.1109/ACCESS.2019.2941280. (Accessed: 21 September 2023).

Deretić, N. *et al.* (2022) 'SARIMA Modelling Approach for Forecasting of Traffic Accidents', *Sustainability (Switzerland)*, 14(8). Available at: https://doi.org/10.3390/su14084403. (Accessed: 5 December 2023).

Doğan, E. (2020) 'Analysis of the relationship between LSTM network traffic flow prediction performance and statistical characteristics of standard and nonstandard data', *Journal of Forecasting*, 39(8), pp. 1213–1228. Available at: https://doi.org/10.1002/for.2683. (Accessed: 21 September 2023).

Du, S. *et al.* (2020) 'A hybrid method for traffic flow forecasting using multimodal deep learning', *International Journal of Computational Intelligence Systems*, 13(1), pp. 85–97. Available at: https://doi.org/10.2991/ijcis.d.200120.001. (Accessed: 15 October 2023).

Fang, W. *et al.* (2023) 'Δfree-LSTM: An error distribution free deep learning for short-term traffic flow forecasting', *Neurocomputing*, 526, pp. 180–190. Available at: https://doi.org/10.1016/j.neucom.2023.01.009. (Accessed: 23 October 2023).

Fouladgar, M. *et al.* (2017) 'Scalable deep traffic flow neural networks for urban traffic congestion prediction', *2017 International Joint Conference on Neural Networks (IJCNN)*, Anchorage, AK, USA, 2017, pp. 2251-2258, doi: 10.1109/IJCNN.2017.7966128. (Accessed: 2 October 2023).

Ghanim, M.S., Muley, D. and Kharbeche, M. (2022) 'ANN-Based traffic volume prediction models in response to COVID-19 imposed measures', *Sustainable Cities and Society*, 81. Available at: https://doi.org/10.1016/j.scs.2022.103830. (Accessed: 15 October 2023).

He, F. *et al.* (2016) 'A Traffic Congestion Assessment Method for Urban Road Networks Based on Speed Performance Index', in *Procedia Engineering*. Elsevier Ltd, pp. 425–433. Available at: https://doi.org/10.1016/j.proeng.2016.01.277. (Accessed: 2 November 2023).

Jonathan Mackenzie, John F. Roddick, and Rocco Zito . (2019) 'An Evaluation of HTM and

LSTM for Short-Term Arterial Traffic Flow Prediction', *IEEE Transactions on Intelligent Transportation Systems*, 20(5), pp. 1847-1857. Available at: https://doi.org/10.1109/TITS.2018.2843349. (Accessed: 20 January 2024).

Kumar, R., Mendes Moreira, J. and Chandra, J. (2023) 'DyGCN-LSTM: A dynamic GCN-LSTM based encoder-decoder framework for multistep traffic prediction', *Applied Intelligence*, 53(21), pp. 25388–25411. Available at: https://doi.org/10.1007/s10489-023-04871-3. (Accessed: 17 October 2023).

Kumar, R., Panwar, R. and Chaurasiya, V.K. (2023) 'Urban traffic forecasting using attention based model with GCN and GRU', *Multimedia Tools and Applications* [Preprint]. Available at: https://doi.org/10.1007/s11042-023-17248-y. (Accessed: 21 October 2023).

Kumar, S.V. *et al.* (2017) 'Integration of exponential smoothing with state space formulation for bus travel time and arrival time prediction', *Transport*, 32(4), pp. 358–367. Available at: https://doi.org/10.3846/16484142.2015.1100676. (Accessed: 21 November 2023).

Li, L. *et al.* (2014) 'Traffic prediction, data compression, abnormal data detection and missing data imputation: An integrated study based on the decomposition of traffic time series', in *2014 17th IEEE International Conference on Intelligent Transportation Systems, ITSC 2014*. Institute of Electrical and Electronics Engineers Inc., pp. 282–289. Available at: https://doi.org/10.1109/ITSC.2014.6957705. (Accessed: 2 December 2023).

Ma, C., Dai, G. and Zhou, J. (2022) 'Short-Term Traffic Flow Prediction for Urban RoadSections Based on Time Series Analysis and LSTM_BILSTM Method', *IEEE Transactions on Intelligent Transportation Systems*, 23(6), pp. 5615-5624. Available at: https://doi.org/10.1016/j.physa.2019.03.007. (Accessed: 18 March 2024).

Medina-Salgado, B. *et al.* (2022) 'Urban traffic flow prediction techniques: A review', *Sustainable Computing: Informatics and Systems*, 35. Available at: https://doi.org/10.1016/j.suscom.2022.100739. (Accessed: 10 November 2023).

Mohammadzadeh, M., Choupani, A.A. and Afshar, F. (2023) 'The short-term prediction of daily traffic volume for rural roads using shallow and deep learning networks: ANN and LSTM', *Journal of Supercomputing*, 79(15), pp. 17475–17494. Available at: https://doi.org/10.1007/s11227-023-05333-w. (Accessed: 3 December 2023).

Muñoz-Organero, M. (2023) 'Using Traffic Sensors in Smart Cities to Enhance a Spatio-Temporal Deep Learning Model for COVID-19 Forecasting', *Mathematics*, 11(18). Available at: https://doi.org/10.3390/math11183904. (Accessed: 6 November 2023).

Qin, P. *et al.* (2023) 'A CNN-LSTM Car-Following Model Considering Generalization Ability', *Sensors*, 23(2). Available at: https://doi.org/10.3390/s23020660. (Accessed: 20 November 2023).

Ranjan, N. *et al.* (2020) 'City-wide traffic congestion prediction based on CNN, LSTM and transpose CNN', *IEEE Access*, 8, pp. 81606–81620. Available at: https://doi.org/10.1109/ACCESS.2020.2991462. (Accessed: 5 December 2023).

Rusul L. Abduljabbar , Hussein Dia, and Pei-Wei Tsai. (2022) 'Unidirectional and Bidirectional LSTM Models for Short-Term Traffic Prediction', *Journal of Advanced Transportation*, 3, pp1-15. Available at: https://doi.org/10.1155/2021/5589075. (Accessed: 2 April 2024).

T. Alghamdi, K. Elgazzar, M. Bayoumi, T. Sharaf and S. Shah. (2019). 'Forecasting Traffic Congestion Using ARIMA Modeling', *2019 15th International Wireless Communications & Mobile Computing Conference (IWCMC)*, Tangier, Morocco, 2019, pp. 1227-1232, doi: 10.1109/IWCMC.2019.8766698. (Accessed: 24 November 2023).

Wu, Y. *et al.* (2018) 'A hybrid deep learning based traffic flow prediction method and its understanding', *Transportation Research Part C: Emerging Technologies*, 90, pp. 166–180. Available at: https://doi.org/10.1016/j.trc.2018.03.001. (Accessed: 14 October 2023).

Zafar, N. *et al.* (2022) 'Applying Hybrid Lstm-Gru Model Based on Heterogeneous Data Sources for Traffic Speed Prediction in Urban Areas', *Sensors*, 22(9). Available at: https://doi.org/10.3390/s22093348. (Accessed: 20 October 2023).

Zhang, Z. and Jiao, X. (2023) 'A spatio-temporal grammar graph attention network with adaptive edge information for traffic flow prediction', *Applied Intelligence* [Preprint]. Available at: https://doi.org/10.1007/s10489-023-05020-6. (Accessed: 1 December 2023).

Zhu, T., Boada, M.J.L. and Boada, B.L. (2024) 'Adaptive Graph Attention and Long Short-Term Memory-Based Networks for Traffic Prediction', *Mathematics*, 12(255), pp2-18. Available at: https://doi.org/10.3390/ math12020255. (Accessed: 20 April 2024).

Zhuang, W. and Cao, Y. (2022) 'Short-Term Traffic Flow Prediction Based on CNN-BILSTM with Multicomponent Information', *Applied Sciences*, 12, pp2-15. Available at: https://doi.org/ 10.3390/app12178714. (Accessed: 12 March 2024).

Zhuang, W. and Cao, Y. (2023) 'Short-Term Traffic Flow Prediction Based on a K-Nearest Neighbor and Bidirectional Long Short-Term Memory Model', *Applied Sciences (Switzerland)*, 13(4). Available at: https://doi.org/10.3390/app13042681. (Accessed: 21 October 2023).