IEEE Access
Multidisciplinary : Rapid Review : Open Access Journal

Received January 14, 2019, accepted January 24, 2019, date of publication February 11, 2019, date of current version February 27, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2896913

A Novel Fuzzy-Based Convolutional Neural Network Method to Traffic Flow Prediction With Uncertain Traffic Accident Information

JIYAO AN, (Member, IEEE), LI FU, MENG HU, WEIHONG CHEN, AND JIAWEI ZHAN
College of Computer Science and Electronic Engineering, Hunan University, Changsha 410082, China

Corresponding author: Jiyao An (jt_anbob@hnu.edu.cn)

This work was supported in part by the National Natural Science Foundation of China under Grant 61370097, and in part by the Natural Science Foundation of Hunan Province, China, under Grant 2018JJ2063.

ABSTRACT As a key part of the method of improving traffic capacity, traffic flow prediction is becoming a research hot-spot of traffic science and intelligent technology, in which the accuracy of traffic flow prediction is particularly concerned. In this paper, a novel fuzzy-based convolutional neural network (F-CNN) method is proposed to predict the traffic flow more accurately, in which a fuzzy approach has been applied to represent the traffic accident features when introducing uncertain traffic accidents information into the CNN at the first time. First, for the sake of extracting the spatial–temporal characteristics of the traffic flow data, this paper divides the whole area into small blocks of 32×32 and constructs three trend sequences with inflow and outflow types. Second, uncertain traffic accident information is generated from the real traffic flow data by utilizing a fuzzy inference mechanism. Finally, the F-CNN model is realized to train the internal information of the trend sequence, the uncertain traffic accident information, and the external information. Moreover, pre-training and fine-tuning strategies are efficiently developed to learn the parameters of the F-CNN. At last, the real Beijing taxicab trajectory and the meteorology datasets are employed to show that the proposed method has superior performance compared with the state-of-the-art approaches.

INDEX TERMS Traffic flow prediction, traffic accident information, fuzzy inference system (FIS), convolutional neural network (CNN), fuzzy-based convolutional neural network (F-CNN).

I. INTRODUCTION

With the rapid development of economy and automotive technology, the number of vehicles in cities has greatly increased. As a result, many related traffic problems have arisen, such as traffic jams, traffic accidents, environmental pollution, money and time wastage [1]. In order to solve these serious problems, traffic flow prediction research has become a vital part of the research in the transportation field. Accurate and timely traffic flow prediction not only mitigates related traffic problems, but also provides passengers with useful traffic information [2]. Furthermore, the emergence of big data makes the traffic data exploding, which undoubtedly brings great opportunities for the prediction of traffic flow.

Nowadays, we have entered the era of intelligence, and there are more or less intelligent software assistance in the travel. The Intelligent Transportation System is an integrated transportation management system that integrates electronic information, artificial intelligence, geographic information, global positioning, image analysis, communication engineering and other technologies. It is considered as a novel and effective method to solve traffic problems, especially in it has a unique performance in solving traffic congestion, reducing accidents and traffic pollution [3]. Therefore, as an important part of the research in the field of transportation, traffic flow prediction technology plays an important role in the guidance of intelligent transportation systems [4], which makes the research of optimizing prediction accuracy become especially urgent.

It is well known that traffic flow data has the characteristics of limitations, complexity, and noise interference [5]. In order to deal with these issues, researchers have made many contributions. The limitations of traffic data and noise interference pose a huge challenge to the prediction of traffic

The associate editor coordinating the review of this manuscript and approving it for publication was Yuedong Xu.

flow. First, the uncertain factors that affect the prediction of traffic flow are diverse, and the impact of different factors on the prediction results is different. The singularity of data types leads to limited extraction of factors. Secondly, the noise interference further increase the difficulty of data feature extraction. However, the emergence of fuzzy theory in the 19th century provided a good research direction for the processing and unification of data uncertainty. Moreover, fuzzy systems based on fuzzy theory have been continuously researched, which are widely used in many practical problems such as control engineering, pattern recognition, and artificial intelligence [6]. Compared with the previous deterministic description, fuzzy theory can describe the original data more naturally, furthermore, it has better advantages in dealing with the limitations of data and the robustness of noise [7].

On the other hand, based on the complexity of data, many network models are built to extract data features. However, some existing shallow models have the drawback of over-fitting, which makes the deep learning method widely concerned in feature extraction. As a key model of deep learning, convolutional neural networks have become one of the research hot-spots in many scientific fields, especially in the fields of image processing, pattern recognition, natural language processing and classification. Among them, the achievement in image processing is particularly prominent, which is because the original image can be directly input without complicated preprocessing, avoiding time-consuming reverse error propagation. Based on this idea, it is of great research value to process the traffic data into a map form similar to the image and use the convolutional neural network(CNN) to construct the predictive model.

In our work, we focus on the prediction of traffic flow data with uncertainty and spatial-temporal characteristics, in which the amount of data used for prediction is limited and efficient rather than redundant. In this paper, a fuzzy-based convolutional neural network (F-CNN) method for traffic flow prediction is proposed, which uses fuzzy theory and CNN to improve prediction accuracy. The method use the fuzzy inference system(FIS) to generate uncertain traffic accident information, alleviates the limitations of traffic data. In addition, the layer-by-layer training model of convolutional neural network is used to learn the characteristics of traffic data internal information, traffic accident information and external information, and form the F-CNN prediction model for future traffic flow prediction. Finally, the advantages of the proposed method are compared with other methods from four performance aspects. To the best of our knowledge, the proposed method shows great performance in terms of traffic prediction.

Generally, the contributions of this article are mainly reflected in the following three aspects:

(1)Introducing fuzzy theory into the traffic flow prediction model of convolutional neural networks. The uncertain traffic accident information obtained by the fuzzy inference method is combined with the spatial-temporal feature information in the traffic flow data, and then feature learning is performed using CNN model.

(2)The visualization of F-CNN depth model is realized. we performed the research of fuzzy-based CNN method to traffic flow prediction, which uses fuzzy methods to derive uncertain traffic accident information from the original traffic flow data to effectively handle the uncertainty problem in data.

(3)This method exhibits superior performance compared to existing methods. The superior performance of the proposed method is proved by experiments. The RMSE of the proposed method is reduced to 17.14, which has 15.65% improvements compared with the best method in other methods in the experiment. And the proposed method has good convergence in feature learning. Based on the consideration of uncertain traffic accident information, if this model is used to train and predict the traffic flow in frequent traffic accident areas will expected to show better prediction performance.

The rest of this paper is organized as follows. Section II reviews the related work of traffic flow prediction, while the preparation of this study is introduced in section III. In Section IV, we presents the proposed F-CNN method and algorithm. And with the performance analysis of the F-CNN method is carried out through experiments in Section V. Finally, Section VI summarizes the research of this paper and provides the future research direction of traffic flow prediction.

II. RELATED WORK

As mentioned above, traffic flow data has the characteristics of limitations, complexity, and noise interference, which make traffic flow prediction with obvious uncertainty and complex space-time characteristics. Therefore, establishing a learning model to learn the hidden characteristics of data and taking effective measures to deal with the uncertainty in the data is particularly important in traffic flow prediction. So far, a great deal of research has been done in this field and many models and methods have been adopted, such as: Kalman state space filtering models [8], autoregressive integrated moving average [9], support vector machine model [10], neural network [11], fuzzy logic approach [12], fuzzy-neural systems [13], back propagation neural network model(BPNN) [14], K-nearest neighbor (KNN) model [15], Bayesian network model [16], portfolio models [17], and some deep learning models [18]–[20]. In these prediction methods, the deep learning model has become the focus of many experts and scholars due to the effectiveness of its data feature extraction and the outstanding processing ability of big traffic data [21].

Driven by the flood of traffic data, a challenge arises: Based on the large volume, large variety, large velocity and large veracity characteristics of big flow data [22], can we make full use of the potential knowledge hidden in traffic big data to predict traffic flow? In recent years, deep learning has brought a series of breakthroughs for applications on complex and large data sets (such as images, languages) [18], [23]. Deep learning integrates intrinsic features from multi-layer architecture and classification/regression extraction in an end-to-end manner and has proved to be a promising traffic prediction tool [21]. Wu *et al.* constructed a deep neural network based traffic flow prediction model(DNN-BTF), which considered the weekly/daily cycle and space-time characteristics of traffic flow. In their work, an attention-based model was introduced to automatically learns to determine the importance of past traffic flows, and the spatial characteristics of traffic flows were extracted using convolution neural networks, while the temporal characteristics were extracted using recurrent neural network [21]. According to the unique spatial-temporal characteristics of traffic flow data, Zhang *et al.* [20] designed an end-to-end structural model ST-ResNet, which predicted crowd traffic in each area of the city, the structural model of ST-ResNet is built based on CNN. And recently, Chen *et al.* [5] proposed a fuzzy deep convolutional network (FDCN) structure based on fuzzy theory and deep residual network, which introduced fuzzy representation into deep learning model to reduce the impact of data uncertainty. Wang and Xu [24] reconstructed the traffic time series by using the spatial-temporal correlation of traffic flow, and proposed a city highway traffic flow time series prediction model based on the deep learning framework LSTM-RNN. Du *et al.* [25] constructed a multi-layered integrated hybrid deep learning architecture to predict short-term traffic flow. This architecture used a recurrent neural network to capture the time dependence of traffic flow, while a convolutional neural network to capture local trend features. Zhang and Huang [26] proposed a model that combined a deep belief network and genetic algorithm to predict traffic flow. Since the traffic flow has various characteristics at different times, they optimized the model by using genetic algorithm to found the optimal hyper-parameter of deep belief network at different times. Furthermore, Zhao *et al.* [27] designed a master-slave parallel computing structure for the parallel computing learning model of the proposed DBN learning process. And Kong *et al.* [28] using the deep learning algorithm of the restricted Boltzmann machine (RBM) to construct a long-term polymorphism model of chaotic time series for traffic flow prediction. In their work, phase space reconstruction was applied to identify the data and achieved good results.

III. PRELIMINARY MEASURES

A. DATA PREPROCESSING

In practice, most traffic-related flow data obtained from various sensors or systems is incomplete and there are always exist defects such as missing, limited, and noisy. However, most researches on traffic flow prediction ignore the effects of this phenomenon and deal with it by simply deleting the missing data. Although this method is simple and easy to operate, it also brings inevitable errors to the final prediction. Firstly, direct deletion of missing data reduces the amount of data used in training and affect model training, especially when the missing data is large. Secondly, deleting missing data will result in a lack of spatial-temporal dependence for other data, because missing data may be spatial-temporal dependence for other data. Therefore, it is necessary to effectively process missing data. Based on the periodic characteristics of traffic flow data, we supplement the missing data with adjacent data of the same trend and define the following proximity alternative method. This method can effectively reduce the impact of missing data on the final result and improve the performance of the prediction.

PROXIMITY ALTERNATIVE METHOD

Since traffic flows are time-cycled, that is, daily or weekly traffic flows have the same trend. Therefore, data missing at a certain time will be associated with the data which has same time of the previous day (or the next day) or the same time of the previous week (or next week). Thus, we define a proximity alternative method that use data from the trend period as the value of the missing data, as follows:

$$x_t^{(i,j)} = \begin{cases} x_{t\pm 48k}^{(i,j)} & (k = 1, \ldots 7) \\ x_{t\pm 336p}^{(i,j)} & (p = 1, \ldots 4) \\ 0 & others \end{cases} \tag{1}$$

where $x_t^{(i,j)}$ indicates that the traffic flow of the region block (i, j) at the time t. Since one day is divided into 48 time slots, 48 in Eq.(1) represent the period of one day, and 336 is the period of one week. And as the existence of data used for supplementing is uncertain, the values of k and p are random, and their values are usually from small to large within the range satisfying Eq.(1). In addition, when determining the substitute value, the value determined by k is prioritized, followed by p. While the value exist, the value is used as supplementary data, otherwise the data of the missing time period is defaulted to 0.

B. FUZZY TRAFFIC ACCIDENT REPRESENTATION

In 1960, the University of California (Berkeley) proposed fuzzy logic. Fuzzy logic is a mathematical method used to simulate the expression and reasoning of human concepts, which can describe the uncertainty and ambiguity of data [29]. With the development of fuzzy logic, more and more researchers have applied fuzzy theory based on the knowledge of human expert experience to intelligent traffic [30]. The key to intelligent transportation is traffic flow, and fuzzy logic is ideal for dealing with uncertainties information that affect traffic flow. Therefore, fuzzy logic is considered to be a promising intelligent method for traffic information modeling [29].

When traffic data occasionally changes greatly at a certain moment, the feature is ignored by the training model due to its contingency, however, in practice, this phenomenon may be caused by a traffic accident. Therefore, it is necessary to consider traffic accident information in the research of traffic flow prediction. Due to traffic accident data is not easy to obtain and its representation is ambiguous, most of the research on traffic flow prediction does not involve traffic accident data or use traffic accident data with a large amount of missing data. Currently, methods for processing uncertainty information include random processing and fuzzy processing. However, the implementation and statistics of random processing are complicated, while fuzzy theory can process fuzzy information in a very simple and convenient form and is used in this paper. As we all know, the occurrence of traffic accidents depends largely on traffic flow. When the traffic flow increase sharply, the probability of accidents will increase greatly. In addition, the traffic flow that the road can withstand will also affect the occurrence of the accident. Based on the idea of Tang *et al.* [31] in the literature used improved fuzzy neural network to predict traffic flow velocity, we define a uncertain traffic accident information to simulate actual traffic accident information. The traffic accident information used in the following refers to this uncertain information. In this study, traffic accident information is determined by relative traffic flow, relative flow change rate, and relative road tolerance. The following definitions of these relative quantities are given based on actual traffic flow data.

Definition 1 (Relative Traffic Flow): The actual traffic data is four-dimensional data, and the introduction of the data will be given later. Each data consists of two parts, inflow and outflow. We define the difference between the inflow and the outflow as the relative traffic flow for a certain time in a certain area.

$$F_t^{(i,j)} = x_t^{in(i,j)} - x_t^{out(i,j)} \tag{2}$$

Definition 2 (Relative Flow Change Rate): The change rate of traffic flow is relative to the previous moment. We divide the relative traffic flow at the previous moment with the relative traffic flow at the current time, and use the result as the relative flow change rate (the first relative flow change rate is initialized to 1; in addition, while the value of relative traffic flow at the current time is 0, then, change it to 1).

$$R_t^{(i,j)} = F_{t-1}^{(i,j)} / F_t^{(i,j)} \tag{3}$$

Definition 3 (Relative Road Tolerance): It is determined by the average of the relative traffic flow at all times on the day of the block.

$$T_d^{(i,j)} = \sum_{t=(d-1)48+1}^{48d} F_t^{(i,j)} / 48 \tag{4}$$

In the above-mentioned Eqs.(2-4), F, R, T indicates the relative traffic flow, the relative flow change rate, and the relative road tolerance, respectively. Among them, t means the time t, while (i, j) represent a small block. In the Eq.(2), the in/out represents the inflow and outflow of the corresponding area block respectively. And for Eq.(4), d represents the number of days. As a result of collecting values of 48 points per day, the tolerance of the day is the average value of the value of the 48 time slots of the day.

Definition 4 (Relative Traffic Accidents): Obtain from FIS with three input variables: relative traffic flow, relative flow change rate, and relative road tolerance, ie $F_t^{(i,j)}, R_t^{(i,j)}, T_t^{(i,j)} \xrightarrow{fuzzy} A_t^{(i,j)}$, the A is indicate the uncertain traffic accidents. The detail process will be introduce later.

C. CONVOLUTIONAL NEURAL NETWORK MODEL

As one of the deep learning models, CNN can not only simulate nonlinear relationships in input data, but has also been revealed to successfully and effectively utilize the temporal character of data [32]. In addition, it has significant tiered capture capabilities for spatial structure information. Therefore, CNN can effectively simulate the spatial dependence between data [33].

Usually, the basic structure of CNN consists of two parts. One is the feature extraction layer, in which the input of each neuron is connected to the local area block of the previous layer, then the complete feature vector can obtained by extracting the features of each local area block. The other is the feature mapping layer. As can be seen from the previous layer, each computing layer of the network will composed of multiple feature maps. In order to ensure that the feature map has a constant displacement, an activation function is added to the convolutional network of the feature mapping structure. In addition, weight sharing method is used in the mapping layer to reduce model parameters.

In this work, we use CNN to extract features in traffic data. First, the data is constructed into two formats: inflow and outflow. Second, use three trends (time trend, day trend, and weekly trend) to build the internal information of the time trend sequence. Then, traffic accident information is generated based on the fuzzy inference system. Finally, the CNN model is used to train internal information, traffic accident information, and external information. The purpose of training data using the network model is to continually adjust the parameters used in the network to achieve model optimization.

The CNN has two stages of forward transmission and reverse transmission. For the forward transmission stage, a 3 * 3 convolution kernel is used to deal with internal information and accident information in the convolution layer. Then add a deviation parameter on the basis of the convolution of the upper feature map, and obtain the output of the layer under the action of the activation function. As shown below, where n is the number of input matrices or the dimension of the last dimension of the tensor. x_i represents the ith input matrix, while w_i represents the ith sub-convolution kernel matrix of the convolution kernel. And $h_{w,b}(x)$ is the value of the corresponding position element of the output matrix corresponding to the convolution kernel W^T.

$$h_{w,b}(x) = f(W^T x) = f(\sum_{i=1}^{n} w_i x_i + b) \tag{5}$$

In the reverse transmission stage, the weights and offsets are adjusted by minimizing the residuals. The residual of the output layer of CNN is different from the calculation method of the residual of the inter-mediate layer. The residual of the output layer is the error between the output value and the actual value, while the residuals of the middle layers are derived from the weighted sum of the residuals of the next layer. The equations are as follows:

$$\delta_i^{(n_l)} = \frac{\partial}{\partial z_i^{(n_l)}} \frac{1}{2} ||y - h_{W,b}(x)||^2 = -(y_i - a_i^{(n_l)}) \cdot f'(z_i^{(n_l)}) \tag{6}$$

$$\delta_i^l = (\sum_{j=1}^{s_{t+1}} W_{ji}^{(l)} \delta_j^{(l+1)}) f'(z_i^{(l)}) \quad (l = 1, 2, \ldots n_l - 1) \tag{7}$$

The residual of the output layer is calculated by the Eq.(6), in which $\delta_i^{(n_l)}$ represent the residual of the corresponding output unit i in the nth layer (output layer). y and f represents the output, activation function. And a corresponds to the h value of each layer, z is the value corresponding to the forward transmission after the activation function. While the residual calculation method of the i-th node of the l layer is as shown in Eq.(7).

D. PROBLEM DESCRIPTION

The problem of traffic flow prediction is actually to predict the traffic flow volume of a certain area at a certain time in the future through historical traffic data. In order to fully extract the spatial-temporal characteristics of traffic flow data, this paper divide the whole area into small areas of 32*32, and each area uses two types of flow data, inflow and outflow, to characterize the spatial characteristics of the data. In addition, since the traffic flow in the adjacent time has the same characteristics, and the trend of traffic flow in the same time of adjacent day (or the same day in adjacent week) is similar, we use the trend sequence to describe these time characteristics of the traffic flow, which is show in follow. Among them, the trend sequence consist of three components: time trend, daily trend and weekly trend.

$$x_{trend} : [x_{t-l_c \cdot c}, x_{t-(l_c-1) \cdot c}, \cdots, x_{t-c}] \tag{8}$$

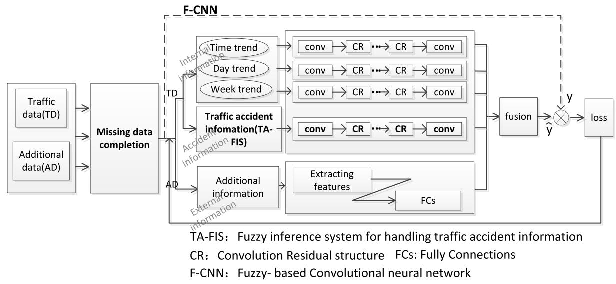
In Eq.(8), *trend* represents time trend, day trend and week trend, respectively. c is the period of each trend, and l_c indicate the length of the trend. Furthermore, we expand uncertain traffic accident information through FIS. The prediction problem to be solved is: according to the observed traffic flow data at the historical moment of the area $\{x_t^{(i,j)} |\ t=0,1,\ldots,n-1\}$, predict the traffic flow at a certain moment in the future $x_n^{(i,j)}$.

IV. THE F-CNN METHOD AND ALGORITHM

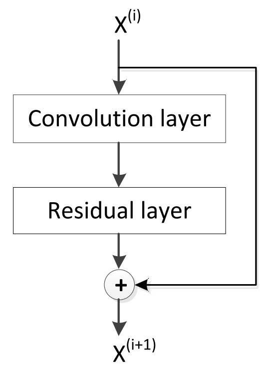
In this section, we present the proposed F-CNN methodologies, including the method implementation and the prediction algorithm.

A. THE PROPOSED F-CNN METHOD

Traffic flow is a key factor in traffic flow prediction, making the research on its characteristics essential. Meanwhile, due to the spatial-temporal characteristics and uncertainty of traffic flow data, it is of great significance to introduce fuzzy theory into CNN. Among them, fuzzy theory can effectively represent the uncertainty of data, while CNN can train the spatial-temporal features of high-dimensional data.



(a)



(b)

FIGURE 1. The process architecture of F-CNN. (a) The F-CNN model. (b) CR model.

As shown in Figure 1 is the process architecture of the F-CNN. In Figure 1(a), the original data is divided into traffic data and additional data after missing data completion. Then, use traffic flow data construct internal information according to Eq.(8). Besides, through the TA-FIS to generate the uncertain traffic accident information, and training with the CNN model. The CNN model consists of multi-layer convolution layer and convolution residual (CR) structures, in which the CR structure consists of a convolution layer and a residual layer, as shown in Figure 1(b). In addition, the additional data is subjected to feature extraction and full connections process to obtain additional factor features. Finally, the above three path extraction features are combined to obtain the final predicted value. Meanwhile, the loss function $L(\theta)$ is used to find the optimal model parameters.

The entire model involves three types of information, namely, internal information of traffic flow data, uncertain traffic accident information and external factor information. The internal information of the traffic flow data is constructed into a trend sequence according to the introduction of the previous section. Then, a fuzzy inference system is used to generate uncertain traffic accident information. The external factors information is formed by the weather, temperature, wind speed and time point in the obtained meteorological data.

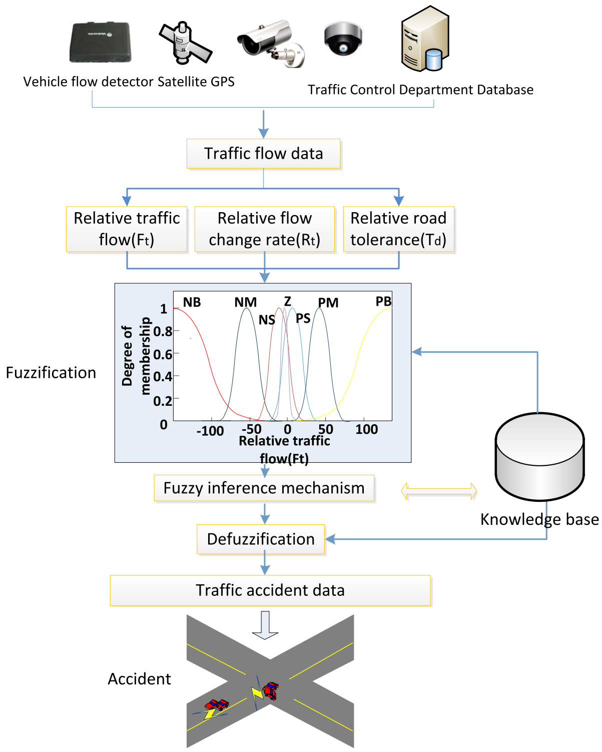


FIGURE 2. The FIS architecture for traffic accident information.

1) TRAFFIC ACCIDENT INFORMATION USING TA-FIS

As discussed in previous sections, traffic accident data is determined by TA-FIS. The TA-FIS consists of four parts: fuzzification, rule base, fuzzy reasoning and defuzzification, as shown in Figure 2. Among them, the fuzzification process includes transforming variables into a fuzzy representation and determining the membership function. According to the definition and introduction in the preliminary measures, construct the input part of relative traffic flow, relative flow change rate and relative road tolerance, and the output part of uncertain traffic accident information, and the fuzzification are shown in Table 1. Furthermore, S-type membership function, Gaussian membership function, triangular membership function and ladder membership function are commonly used membership functions. Since Gaussian membership function can well describe the input and output data, we use it in this study. As shown in Eq.(9), it is clearly to find that the Gaussian membership function is determined by σ and μ parameters, and parameter σ usually takes a positive value while the parameter μ is used to determine the center of the curve. In this study, these parameters are determined based on the range of actual data and expert experience, and are continuously adjusted through experiments.

$$f(x, \sigma, \mu) = e^{\frac{(x-\mu)^2}{2\sigma^2}} \tag{9}$$

After fuzzy processing, the input and output are transformed into fuzzy-quantity representations. Then, based on the expert experience, we establish the control relationships

TABLE 1. Fuzzy set representation of variables.

variables		fuzzy sets
input	relative traffic flow	NB NM NS Z PS PM PB
	relative flow change rate	NB NM NS Z Z1 PS PM PB
	relative road tolerance	S M B
output	uncertain traffic accident	VS SM S M B BM VB

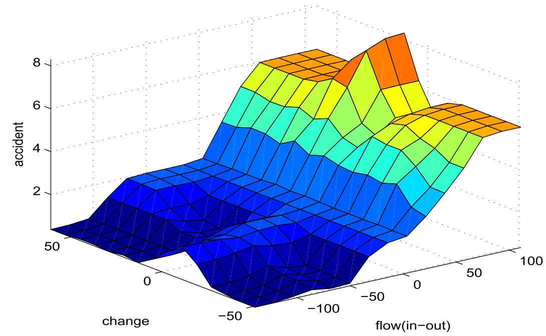


FIGURE 3. 3D map of input and output relationships.

between the input fuzzy quantity and the output fuzzy quantity. Generally, the fuzzy rule are established by a sentence similar to "IF THEN". In this study, the input fuzzy quantity consists of three parts: relative traffic flow, relative flow change rate, and relative road tolerance. And the output fuzzy quantity represents the uncertain traffic accident. There are some of the fuzzy rules used in this study.

$$\begin{cases} IF\ \ F\ \text{is}\ NB\ \text{and}\ R\ \text{is}\ NB\ \text{and}\ T\ \text{is}\ M,\ \text{then}\ A\ \text{is}\ VS \\ IF\ \ F\ \text{is}\ NM\ \text{and}\ R\ \text{is}\ NS\ \text{and}\ T\ \text{is}\ B,\ \text{then}\ A\ \text{is}\ SM \\ IF\ \ F\ \text{is}\ NS\ \text{and}\ R\ \text{is}\ PM\ \text{and}\ T\ \text{is}\ S,\ \text{then}\ A\ \text{is}\ S \\ IF\ \ F\ \text{is}\ Z\ \text{and}\ R\ \text{is}\ NM\ \text{and}\ T\ \text{is}\ B,\ \text{then}\ A\ \text{is}\ M \\ IF\ \ F\ \text{is}\ PS\ \text{and}\ R\ \text{is}\ PS\ \text{and}\ T\ \text{is}\ S,\ \text{then}\ A\ \text{is}\ B \\ IF\ \ F\ \text{is}\ PM\ \text{and}\ R\ \text{is}\ PB\ \text{and}\ T\ \text{is}\ M,\ \text{then}\ A\ \text{is}\ BM \\ IF\ \ F\ \text{is}\ PB\ \text{and}\ R\ \text{is}\ NB\ \text{and}\ T\ \text{is}\ B,\ \text{then}\ A\ \text{is}\ VB \end{cases}$$

Then, fuzzy reasoning is performed based on the fuzzy inference rule base and the current situation of the system to obtain the fuzzy value of the output. This output value appears as a fuzzy subset on the output domain, however, what is needed in practical applications is not such a fuzzy representation. Therefore, the fuzzy subset needs to be defuzzified and converted into precise control quantities, which are then applied to the actual project. Defuzzification has the largest membership averaging method, weighted average method, center of gravity method and central averaging method. Based on practical considerations and comparisons of related methods after experiment, we use the method of center of gravity, which uses the center of gravity of the region surrounded by the membership function curve and abscissa as the output value of the final fuzzy inference.

With the establishment of fuzzy rule, the 3D coordinate map which through the fuzzy decision matrix operation and defuzzified can be obtained in the surface observer window, as shown in Figure 3. Each coordinate axis in the figure represents a fuzzy variable. Due to the limitation of the dimension, only two inputs can be displayed in the three-dimensional coordinate graph. Therefore, the x and y axis in this study respectively represents the input fuzzy variable F,R, and the z-axis represents the output fuzzy variable A, and the coordinate range is the domain of the fuzzy variable. It can be seen that the three-dimensional coordinate map is equivalent to a fuzzy control query table. It can intuitively verify whether the establishment of the fuzzy rule is reasonable. When the control expectation value is not near the center of the fuzzy inference output conclusion (for example, more than 20%), it is required re-adjust the membership function and control rules until the requirements are met. In addition, various bumps on the surface show the effectiveness of introducing traffic accident information into traffic flow prediction.

2) EXTERNAL FACTORS INFORMATION

In actual situations, people's travel may be affected by external factors such as weather condition, temperature and weekdays. For example, when the weather is poor, the number of cars on the road will be less. While the weather is good, the amount of travel will increase accordingly. In addition, as the temperature decreases, the travel time of people will also be biased. It can be seen that external factors have impact on traffic flow.

External factors in this study include the weather (WE), wind speed (WD), temperature (TE), and weekday/weekend information. Wind speed and temperature are actual values that do not need to be processed. The weather is made up of a 17-bit binary form, each bit representing a weather condition, we uses a integer values in 1-17 instead it. In addition, based on the time slot information, workday/weekend information is obtained, which is represented by 8-bit binary data, with the first 7 digits representing the day of the week and the last digit representing the weekdays or weekends.

3) F-CNN METHOD

Recent years, with the rapid development of the information industry, big data has been widely used. Big data is characterized by large size, variety, and fast speed, which pose challenges to deep learning models [22]. Meanwhile, the depth calculation model has been proved to be effective for hierarchical analysis of large data and representation learning in tensor space. Therefore, it is necessary to use tensor to represent the complex big data of traffic data. In our work, the entire area is divided into 32*32 small area blocks by the form of a grid, the position of each area block is represented by (i, j), and the data of inflow and outflow are used in each small block. So, the traffic data of a small area at any time can be represented by the tensor $x \epsilon R^{2*i*j}$.

As can be seen from the above, the internal information of the traffic flow data, the uncertain traffic accident information and the external factor information constitutes the complete input data of the CNN model.

Then, in the stage of CNN, as show in Figure 1, multi-layer CR structures combines convolution layer with residual layer constructed the core part of CNN model. After the fully connections of the external factors information, the convolution result is merged with the result of external information. In addition, the loss function $L(\theta)$ is used to continuously optimize the parameters of the model to construct the entire F-CNN model.

The key to the F-CNN model are the TA-FIS and the establishment of CR structures, in which the TA-FIS is used for generating uncertain traffic accident information, while CR structure aims to better extract the characteristics of the internal information and accident information. Because in the convolution layer of the CNN, small changes in the input can cause large differences in the output, but this difference is not what we would like to see. Therefore, in order to solve this problem of enabling the model to simulate more subtle changes, an activation function is used to add nonlinear factors. It is a well known fact that most activation functions have the problem that the saturation state network weight cannot be updated. However, the residual function can largely solve the gradient dissipation caused by the above problem. So, this study uses a CR structure which combines convolution and residual. The activation function and convolution function used in the structure are as follows.

$$f(a) = max(0, a) \tag{10}$$

$$x_{trend}^{(i)} = f(w_{trend}^{(i)} * x_{trend}^{(i-1)} + c_{trend}^{(i)}) \tag{11}$$

In the Eqs.(10-11), f is the activation function, a is the input value for each layer, and w and c are the learning parameters in the i layer. (i) indicates the current layer, while $(i-1)$ indicates the layer in front of the current layer. *trend* represents the three components of the time trend, daily trend, and the weekly trend which introduce in the previous section. Under the action of the residual layer, the processing of a CR structure is as Eq.(12). The whole CNN model consists of multi-layer CR structures.

$$x_{trend}^{(i)} = x_{trend}^{(i-1)} + \delta(x_{trend}^{(i-1)}, \theta_{trend}^{(i-1)}) \tag{12}$$

where δ is the residual function and $\theta_{trend}^{(i-1)})$ represents all parameters of the previous layer of the CR structure, including w and b. The model is optimized by continuously adjusting the model parameters to optimize the residuals.

B. THE DESIGNED ALGORITHM

Algorithm 1 outlines the main points of the prediction algorithm of F-CNN. We first perform the missing data completion on the original dataset. Then, we construct an instance set consisting of internal information of the trend sequence, uncertain traffic accident information, external information, and current traffic flow information. Next, we use the F-CNN model to train the instance set in batches until the model stop criterion is reached, find the optimal model parameters. Finally, use the optimal model to predict the test set data. The criteria for stopping in the F-CNN model is the early stopping strategy. The parameters of the optimal model is determined by minimizing the target mean square error $L(\theta)$, as shown in Eq.(13), where θ is the learning parameters, and Y_t and $\hat{Y}_t$ are the real flow matrix and the predicted flow matrix, respectively.

$$L(\theta) = \|Y_t - \hat{Y}_t\|_2^2 \tag{13}$$

Algorithm 1 Prediction Algorithm of F-CNN

Input:

Traffic data(in-out): $\{x_t^{(i,j)} | \text{ t=0,1,...,n-1}\}$;

Meteorology data: WE, TE, WD;

Output:

The F-CNN model $M(W, b)$;

D=∅;

Interpolate missing data $x_t^{(i,j)}$ with Eq.(1);

for x_t in all timestamps t **do**

  According to section 3-4, construct x_{trend}, traffic accident info AD and external info ED;

  put an instance($\{x_{trend}, AD, ED\}$,x_t) into D;

end for

Initialize learning parameters θ in F-CNN;

repeat

  randomly select a batch of instances D_i from D;

  deal with F-CNN model using Eqs.(10-12) and Eqs.(5-7);

  find θ by minimizing the objective (13) with D_i;

until reach model stop criteria

return $M(W, b)$;

The process of the entire algorithm can be divided into the following specific steps:

Step 1: Data preprocessing

(1) According to the existing traffic flow data, the corresponding time stamp is set with the daily data sampling value of 48. When the data sampling value of one day is less than 48, the missing data is processed using Eq.(1).

(2) Linearly transform the original data using Eq.(14) to map the data to the range of [−1, 1].

$$x_i^* = \left(\frac{x_i - x_{\min}}{x_{\max} - x_{\min}}\right) * 2 - 1 \tag{14}$$

where x_i represents the original data, $x_{\max}$ and $x_{\min}$ represent the maximum and minimum values in the original data, respectively, and x_i^* is the processed value.

Step 2: Parameter initialization of the F-CNN model.

The parameters of the F-CNN model mainly include the connection weights and deviations between the layers. The connection weights between the layers are initialized using randomization and all deviation values are initialize to 0.

Step 3: The construction of three kinds information.

The feature data sequence construction includes sequence construction of internal information, accident information, and external information.

(1) The internal information sequence is constructed according to the Eq.(8) introduced above.

(2) The accident information is constructed using the TA-FIS system of the chapter IV-A1.

(3) External information is processed according to the chapter IV-A2.

Step 4: F-CNN Model forward processing

The F-CNN model consists of a series of consecutive convolutional blocks and residual blocks, which are connected to each other using the edge of the weighted value, and the deviation value is added.

(1) After processing a convolutional layer in a convolutional block, an activation function is used to add a nonlinear factor. The implementation is as follows:

$$y_{w,b}^{conv}(x) = f(W^T x) = f(\sum_{i=1}^{n} w_i x_i + b) \tag{15}$$

where W^T is the weight matrix of the convolutional layer in the convolutional block, x and $y_{w,b}^{conv}(x)$ are the input matrix and output matrix of the convolutional layer in the convolutional block, b is the deviation value, and f is the activation function shown in Eq.(10).

(2) The residual block is used to solve the problem of gradient dissipation caused by the inability to update the saturation network weight. Processing is as follows:

$$y_{rei}^{(i)} = x^{(i-1)} + \delta(x^{(i-1)}, \theta^{(i-1)}) \tag{16}$$

where $x^{(i-1)}$ and $y_{rei}^{(i)}$ are the input and output of the current residual block, respectively, and δ is the residual of the current residual block. The residual is calculated using Eq.(17), where W_1, W_2, b_1, and b_2 are the weight and deviation values in the first and second layers of the residual block, σ is the ReLU nonlinear activation function.

$$\delta(x, \theta) = W_2\sigma(W_1 x + b_1) + b_2 \tag{17}$$

(3)With a fusion operation to fuse the output of trend sequences, accident information and external information.The implementation is as follows:

$$y_{fu} = w_{tr}^T y_{tr}^T + w_{ua}^T y_{ua}^T + y_{ex}^T \tag{18}$$

where w_{tr}^T and w_{ua}^T are the weights corresponding to the trend feature and the accident feature. y_{tr}^T, y_{ua}^T and y_{ex}^T are the output values of the previous layer corresponding to the three types of features, and y_{fu} is the result of the fusion.

The fused values are then nonlinearly transformed using the following formula.

$$\widehat{y_t} = \frac{e^y - e^{-y}}{e^y + e^{-y}} \tag{19}$$

where y is the value before the above transformation. And $\widehat{y_t}$ is the value after the transformation, that is, the predicted value of the final t period.

Step 5: F-CNN model reverse fine-tuning

The reverse fine-tuning of the model is mainly based on the loss function and the stochastic gradient descent(SGD) method to update the model parameters.

(1) The loss function is determined by the mean square error, and the calculation method is as Eq.(13).

(2) There is a stochastic gradient descent value from each of the output module to the input module. The gradient of the parameter set (W^l, b^l) from each layer to the next layer is calculated according to the loss function as follows:

$$g(\theta_{ij}^{(l)}) = \frac{\partial L}{\partial \theta_{ij}^{(l)}} = \frac{\partial L}{\partial y_{ij}^{(l+1)}} \frac{\partial y_{ij}^{(l+1)}}{\partial x_{ij}^{(l+1)}} \frac{\partial x_{ij}^{(l+1)}}{\partial \theta_{ij}^{(l)}} \tag{20}$$

where θ contains parameters w and b, L is the loss function value corresponding to the current layer, (l) and $(l + 1)$ represent the current layer and the next layer, respectively, and x_{ij} and y_{ij} represent the input and output values of the corresponding position, respectively.

(3) According to the SGD method, the parameters are updated in each iteration by:

$$\theta_{ij}^{(l)} = \theta_{ij}^{(l)} - \eta g(\theta_{ij}^{(l)}) = \theta_{ij}^{(l)} - \eta \frac{\partial L}{\partial \theta_{ij}^{(l)}} \tag{21}$$

where η is the learning rate, based on actual expert experience, the learning rate is set to 0.0002;

Step 6: Train the F-CNN model and save the relevant parameters of the current optimal model until the early stopping strategy is met.

Step 7: Test the test set data using the trained optimal F-CNN model.

V. EXPERIMENTS

This section firstly introduces the experimental preparation of the proposed method, including experimental data and experimental platform. Secondly, the performance of F-CNN method is compared with other methods in similar application scenarios, the proposed method shows superior effectiveness; finally, predictive data are used to provide a simple mapping of traffic congestion.

A. EXPERIMENTAL PREPARATION

This experiment use taxi GPS data and meteorological data in Beijing, and data covers data from November 2015 to April 2016. This area is divided into 32*32 small areas in the form of grids. Meanwhile, the data collected every 30 minutes is used, therefore, for each small area, there will be 24*2 time points in one day. At every time point, there are flow-in and flow-out two types traffic data. Among the data, taxi GPS data includes in/out traffic data, and the meteorology data includes data of four kinds of information such as weather, temperature, wind speed and date. In all data, the last month's data is used as test data, while the other data is used as training data.

The use of old data in Beijing from November 2015 to April 2016 can well describe and reflect the traffic conditions in Beijing at that time, and can reflect the practicality of the proposed method in the actual scene. Moreover, in the current regional traffic flow prediction, the excellent article of Zhang *et al.* [20] also used this data, which on the other hand reflects the value of this data. And as far as we know, this data is the latest and covers the long-term regional traffic flow data. Although the data used may have defects, such as inaccuracy of GPS data caused by external factors. We can improve the accuracy of GPS data by extracting and processing data from different devices. By enhancing the location characteristics of different information, An *et al.* [34] proposed a fuzzy weighted location mechanism based on fuzzy Kalman filter method.

Our preparatory work includes two aspects: data preparation and experimental platform preparation. For data preparation, in our model, we use the proximity alternative method to process the original data and use the max-min normalization method to reduce the data range to [−1,1]. At the time of evaluation, the data is re-expand back to the actual normal value. In addition, on the basis of traffic data, generates uncertain traffic accident information using TA-FIS. The format of the traffic accident information is in accordance with the format of traffic data, and the range is also between [−1,1]. For external factor data, use max-min normalization method to convert the weather, temperature and wind-speed into the data between [0, 1].

For the experimental platform preparation, the experiment in this study was implemented under the Windows 7 64-bit operating system. The memory and storage capacity of the computer are 8GB and 932GB respectively. The type of the CPU used is Intel(R) Core(TM) i5-4590 CPU @3.30GHz, and the configuration of the graphics card is NVIDIA GeForce GT 705.

The experimental software used in the experiment was JetBrains PyCharm Community Edition 5.0.3, Python 3.6 and Matlab. Among them, PyCharm is an integrated development environment of Python. The experiments in this paper were implemented on PyCharm, and our network was built using theano and Keras libraries on the PyCharm experimental platform. Based on the limitations of the experimental platform graphics, theano's GPU acceleration mechanism was used in the study. In the convolution layers of our network, 64-bit and 2-bit filters are used, the filter size is 3*3, and the number of CR structural layers is 12. Due to the requirements of the actual equipment, our batch_size is set to 8, during the training model, 90% of the data in the training set is used to train the model, and another 10% is used as the verification set. The use of validation sets can detect problems in models or parameters in time, such as verifying differences in models on the set, infinity of results, increase or slow growth of the map. At the same time, using the validation set, you can verify the generalization capabilities of the model. In addition, the results of this study are shown using the python visualization library matplotlib. Based on Matlab's complete and effective fuzzy control module system, this paper mainly uses Matlab to generate uncertain traffic accident information of fuzzy control module.

B. COMPARATIVE ANALYSIS

We compare our method with the following methods.

- SARIMA: Seasonal auto-regressive integrated moving average (SARIMA) model is a variant of the auto-regressive integrated moving average (ARIMA) model, which is a model suitable for describing and predicting time series. On the basis of the ARIMA, the periodic characteristics in the traffic flow data are considered, and the seasonal parameters are added in the model and the seasonal difference processing is performed to solve the instability problem in the traffic flow data.
- VAR: Vector Auto-Regressive (VAR) is an advanced space-time model that is used for multi-variable time series. Each variable is a linear function that has lagged in the past and a lag of other variables in the past. And it can capture the pair relationship between all streams.
- DeepST (Zhang *et al.* 2016): It is a spatio-temporal data prediction model based on deep neural network (DNN), which predicts population traffic and focuses on different time dependencies and external factors.
- ST-ResNet: a method based on deep-learning, which use a residual neural network to model the spatial-temporal characteristics of data and integrates some external factors. ST-ResNet learning aggregates the outputs of three residual neural networks dynamically based on data, and assigns different weights to different branches and regions.
- F-CNN series: F-CNN methods is the method proposed in this paper. It use the CNN to extract the feature in internal information, uncertain traffic accident information and external information. The three methods of the F-CNN used here have different control quantities for uncertain traffic accident information(the quantities of fuzzy control output variable). Among them, the output of F-CNN1(uncertain traffic accident information) is represented by 5 quantities : very small(VS), small(S), medium(M), big(B), very big(VB). While F-CNN2 is represented by 7 quantities : very small(VS), medium small(MS), small(S), medium(M), big(B), medium big(MB), and very big(VB), and F-CNN3 is represented by 9 quantities : pretty small(VVS), very small(VS), medium small(MS), small(S), medium(M), big(B), medium big(MB), very big(VB) and pretty big(VVB).

1) THE COMPARISON OF RMSE, MSE AND MAE

It is well known that the evaluation criteria for traffic flow prediction are mean square error (MSE), mean absolute error (MAE), and root mean square error (RMSE). The corresponding calculation method is as follows. This paper uses these three criterions to evaluate the proposed prediction method. As can be seen from Eq.(23), the MSE amplifies the error, making its actual reference poor. Based on this consideration, in the experiment, we focus on evaluating our method with RMSE and MAE.

$$RMSE = \sqrt{\frac{1}{n}\sum_{i}^{n}(y_i - \widehat{y_i})^2} \tag{22}$$

TABLE 2. Comparison between different methods.

Model	RMSE	MSE	MAE
SARIMA	30.91	956.73	18.33
VAR	26.31	691.90	15.22
DeepST	20.91	436.93	12.17
ST-ResNet	20.32	413.06	12.58
F-CNN1 (with 5 quantities FIS)	17.59	309.60	9.97
F-CNN2 (with 7 quantities FIS)	17.53	307.56	9.93
F-CNN3 (with 9 quantities FIS)	17.14	293.91	9.96

$$MSE = \frac{1}{n}\sum_{i}^{n}(y_i - \widehat{y_i})^2 \tag{23}$$

$$MAE = \frac{1}{n}\sum_{i}^{n}|(y_i - \widehat{y_i})| \tag{24}$$

In the Eqs.(22-24), y_i and $\widehat{y_i}$ are the actual value and predict value of traffic flow, respectively. And n is the number of all predicted values.

Based on taxi GPS data and the corresponding weather data from November 2015 to April 2016 in Beijing, we first give the performances of the four other methods on the RMSE, MSE, and MAE, as shown above the dotted line of the Table 2. In addition, three F-CNN methods we used are given below the dotted line. It can be seen from the table that the proposed method outperforms other methods in the three evaluation performances. In particular, the RMSE error rate of the F-CNN3 method can be reduced to 17.14, which greatly improves the accuracy of prediction, and the RMSE performance is improved 15.65% compared to the best performance in the other methods.

Furthermore, comparing the three F-CNN methods we used, it can find that as the output fuzzy control increase, the predicted performance increase gradually. Therefore, when the output control amount is further increased, the prediction performance of the model is expected to continue to improve. However, as the output control amount increases, the time complexity of calculating the data will also increase. In subsequent research, parallel computing technology or other methods can be used to solve this problem. The application of parallel computing in deep learning big data environment has been introduced in the article "Overview of Cognitive Computing Based on Deep Learning" of Chen *et al.* [35]. And Zhao *et al.* [27] designed a master-slave parallel computing structure for the parallel computing learning model of the proposed DBN learning process.

2) IMPROVEMENT ANALYSIS OF RMSE AND MAE

The performance improvement can clearly reflect the value and significance of the proposed method. Based on the first prediction method ARIMA given in Table 1, the performance improvement of the latter method relative to the previous optimal method on the two evaluation criteria of RMSE and MAE is given in turn. The improvement of the evaluation is performed by the Eq.(25).

$$E(m) = \frac{R_{op}^{E} - R_{m}^{E}}{R_{op}^{E}} \tag{25}$$

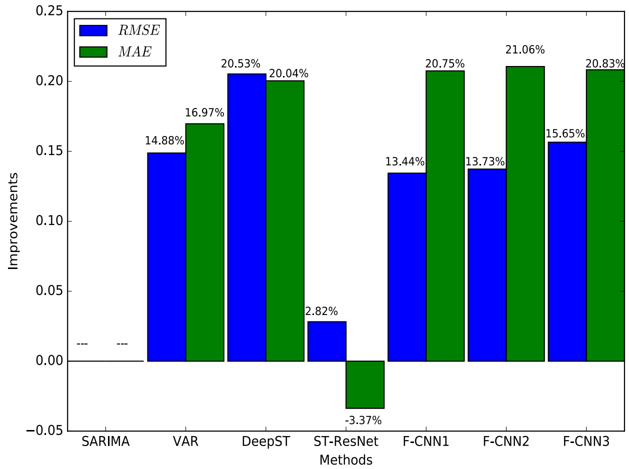


FIGURE 4. Improvements of different methods on RMSE and MAE.

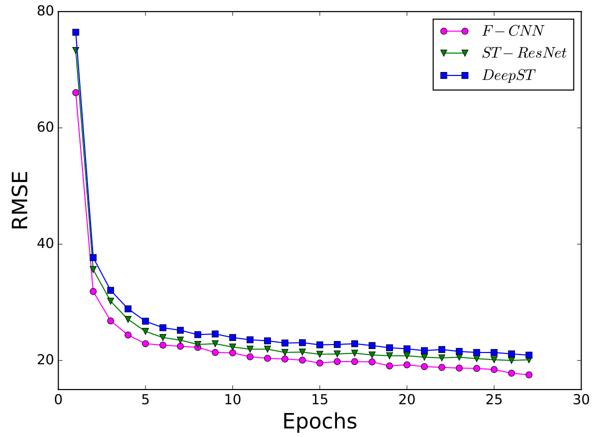


FIGURE 5. The RMSE results of three methods in different epochs.

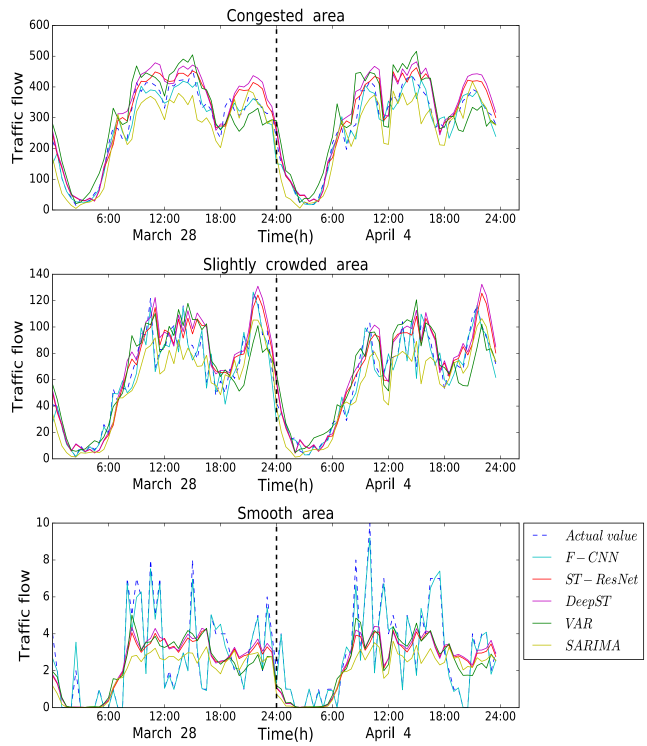


FIGURE 6. The traffic flow for different-type area on Monday.

In Eq.(25), $E(m)$ represents the performance improvement of the method m under the evaluation criterion E, and R_{op}^{E} and R_{m}^{E} represent the results of the current optimal method(op) and the current method(m) according to the evaluation criterion E, respectively. As shown in the Figure 4, the label on the bar is the percentage of performance improvement of the corresponding method compared to the previous best method. It can be seen that the RMSE and MAE performance of each method is generally improved compared to the previous best method, and only the individual MAE performance shows a downward trend, which may be due to a large deviation between a small number of predicted and actual values. Although the proposed method is not the best in terms of performance improvement, it also shows quite a good improvement. Among them, RMSE and MAE performance increased by 15.65% and 20.83%, respectively.

3) CONVERGENCE ANALYSIS

The proposed F-CNN method, ST-ResNet method and DeepST method are deep learning methods. In order to obtain more effective prediction results, there are optimization adjustments in the model training process. In the experiment, we recorded the changes in the predicted performance of the above three models at different epochs. One epoch means that one learning process of all training samples is completed, and all parameters are updated once after each epoch is completed. The 27 epochs of training F-CNN used an early stop strategy. The curve of the RMSE with respect to the number of epochs is provided in Figure 5. As can be seen from the curve in Figure 5, as the number of training epochs increases, the curve becomes flat, which embodies the convergence of the deep learning model. In addition, the F-CCN model converges faster and produces less RMSE than the other two models, that is, the F-CCN model can better learn the data.

4) THE COMPARISON OF PREDICTION EFFECTIVENESS

The comparison of predicted values with real values clearly shows the superior learning ability of the proposed method compared to the existing methods. Here, we compare the predicted results of the proposed method with other methods. Since the whole block of prediction is partitioned into small blocks of 32*32, three different regions are selected as regional representations of three different traffic conditions to compare and analyze the prediction results. In the following, three regions are represented by congested area, slightly crowded area, and smooth area, respectively. For each type of area, the predicted results from Monday to Sunday in the last half of the month are clearly shown in Figures 6-12. Among them, each graph contains two types of curves, where the dotted line represents the actual value, and the solid line represents the predicted value of the different methods.

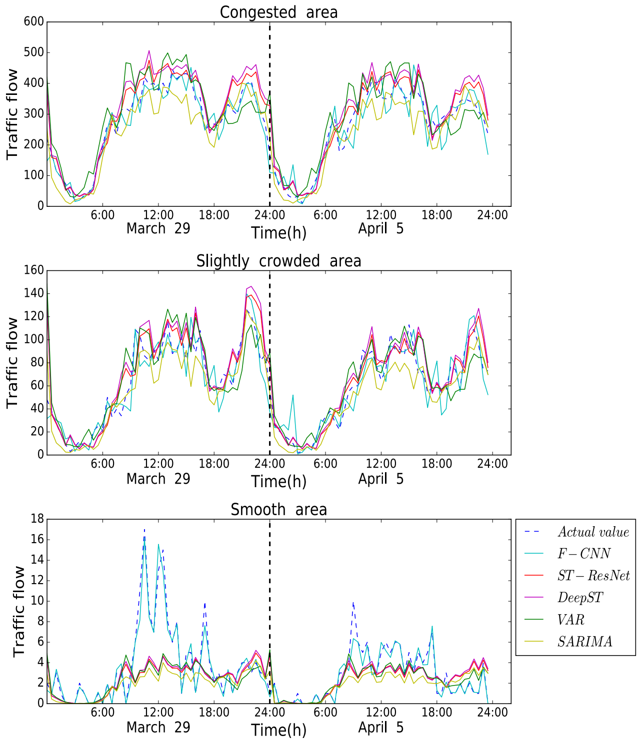


FIGURE 7. The traffic flow for different-type area on Tuesday.

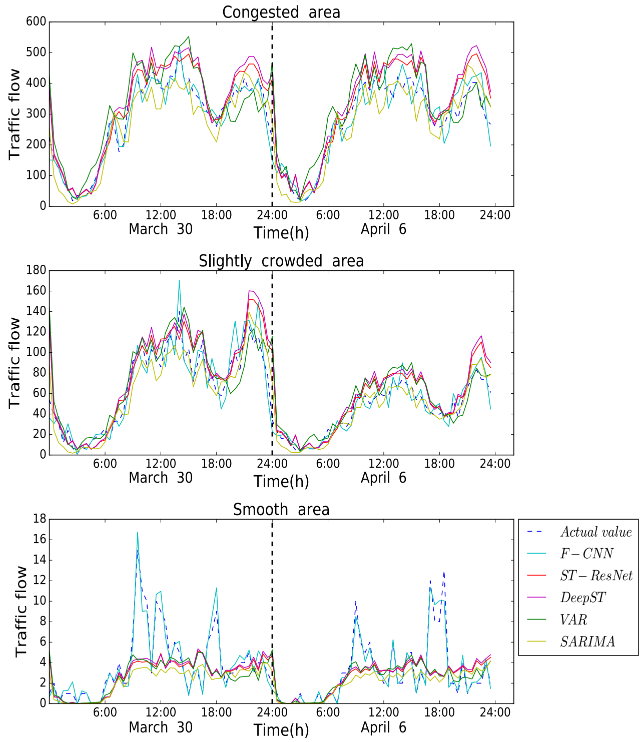


FIGURE 8. The traffic flow for different-type area on Wednesday.

Take the results of Monday shown in Figure 6 as example, it shows the prediction results for two days on Monday in the half-month, the specific date can be seen from the annotations in the figure. In Figure 6, the curves of different colors are used to represent the prediction results of different prediction methods, where blue dashed represents the actual traffic flow values, cyan, deep red, magenta, green, and khaki solid curves represents the predicted results of the F-CNN, ST-ResNet, DeepST, VAR, and SARIMA methods, respectively.

It can be clearly seen from Figure 6 that the proposed F-CNN prediction method has more effective prediction results than other methods. The predicted results of the F-CNN method are very close to the true value of the traffic flow, which can well describe the current traffic conditions. In addition, the ST-ResNet and DeepST methods are better than the VAR and SARIMA methods. Moreover, the traffic flow trends in different regions are generally consistent.

In the form of Figure 6, the prediction results from Tuesday to Sunday are as follows:

From Figures 6-12, we can find that the overall trend of the congested area and the slightly crowded area in a day is similar: the traffic flow will gradually decrease from 24 pm to around 5 pm, then gradually rise to the highest value in the morning, and fluctuated constantly, reaching a low point in the day at around 6 pm, and then entering the peak of the evening. For the smooth area, the flow rate is basically in a gentle fluctuation after the first peak period of the day. During the week, the traffic trends from Monday to Friday are similar, while the traffic trends on Saturday and Sunday are slightly different, which is main performance in

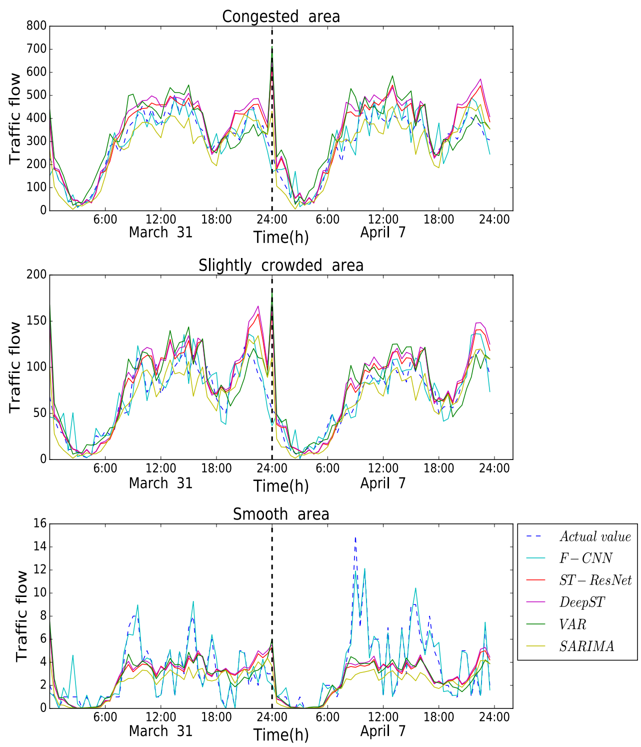


FIGURE 9. The traffic flow for different-type area on Thursday.

that the traffic on the day after peaking around 9 am is in a gentle fluctuation, and has not obvious low point in the afternoon around 6 o'clock. From Table 2, the proposed method used for traffic flow prediction has better result than the

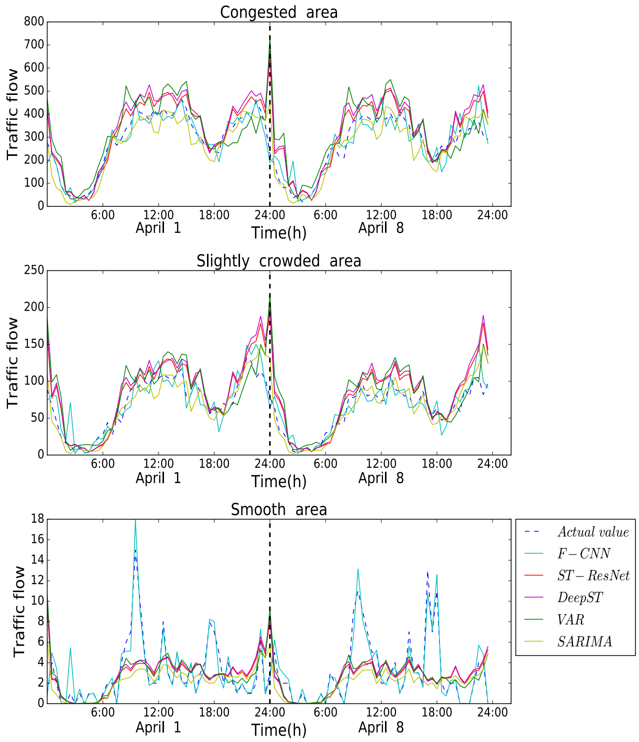


FIGURE 10. The traffic flow for different-type area on Friday.

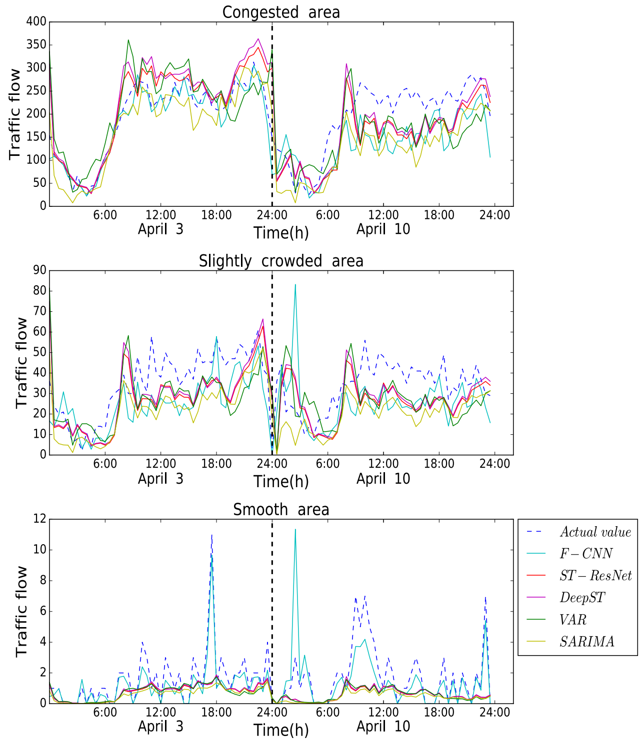


FIGURE 12. The traffic flow for different-type area on Sunday.

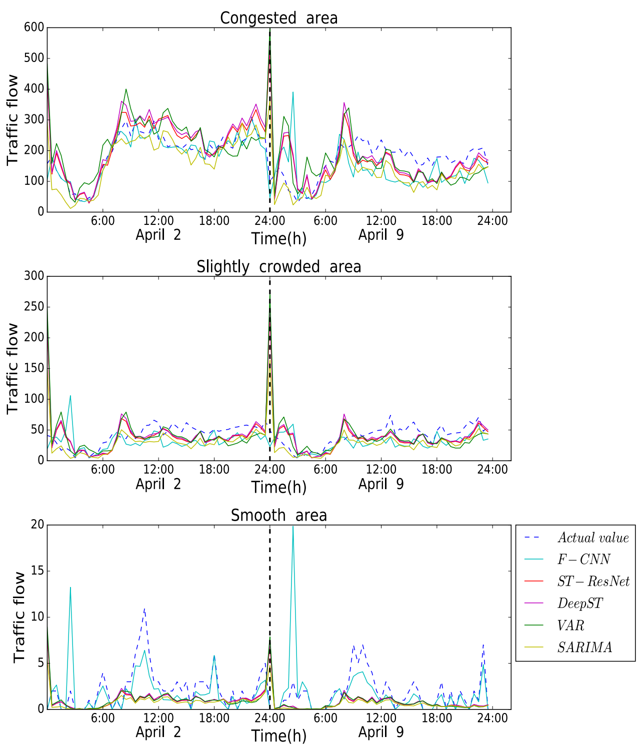


FIGURE 11. The traffic flow for different-type area on Saturday.

state-of-the-art methods. It proves that the fuzzy-based on CNN method with uncertain traffic accident information is a very successful method to be used in traffic flow prediction.

C. DESCRIPTION OF ACTUAL TRAFFIC CONDITIONS

The actual traffic flow value can't directly let us understand the traffic conditions in the current area. In order to better describe the traffic conditions, we use the anti-fuzzification method to convert the traffic flow values into fuzzy quantities which is easy to understand. Here we divide the traffic situation into 8 levels: pretty smooth(VVS), very smooth(VS), smooth(S),medium(M), little crowed(LC), crowded(C), very crowded (VC) and pretty crowded(VVC), etc. The use of fuzzy quantities VVS, VS, S, M, LC, C, VC and VVC can improve the describe precision of traffic conditions.

According to the predicted data which normalized to $[-1, 1]$, the Gaussian membership function is used for processing. After the fuzzy inference, the fuzzy set of 8 levels of each data belongs can be get, then through the anti-fuzzification method median decision to determine the level of each traffic values.

Use the fuzzy processing method similar to Section 4.1.1 to obtain the fuzzy set of 8 levels of each data. Assuming that the fuzzy set is as follows:

$$X = \{x_1, x_2, \cdots x_8\} \tag{26}$$

where X and x_i are the fuzzy set and the membership of the level i of the traffic flow belongs to, respectively. The median decision method refers to taking the element x_k which divides the area surrounded by the membership function curve and the abscissa into two identical parts in the domain as the final decision element. The formula is as follows, in which T is the

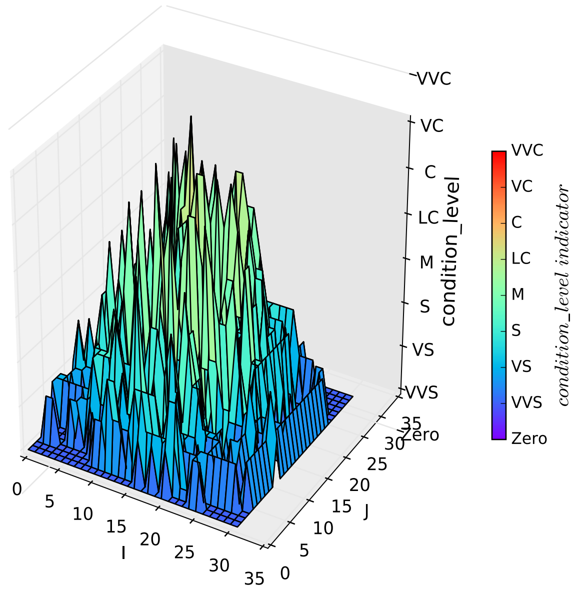


FIGURE 13. **Traffic condition_level at 12:00 in all areas.**

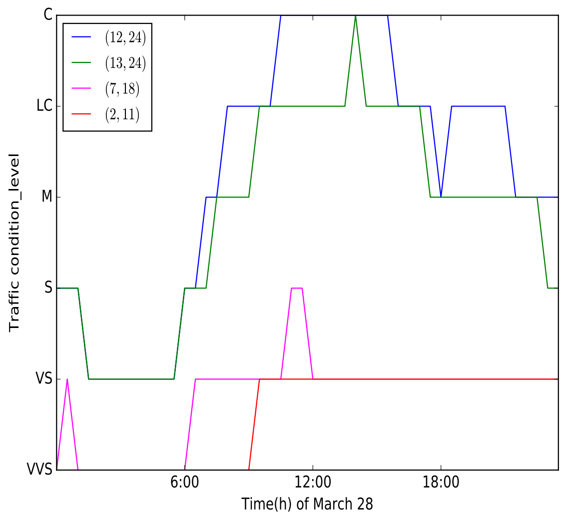


FIGURE 14. **Traffic condition_level for one day in different area.**

final fuzzy control quantity.

$$\sum_{i=1}^{k} T(x_i) = \sum_{i=k+1}^{8} T(x_i) \tag{27}$$

According to the processing in above, the predicted traffic flow value is converted into an eight-level fuzzy amount that is easy to understand. Using this kind of fuzzy quantities to describe the traffic situation allows the traveler to intuitively understand the traffic conditions in the current area and make reasonable decisions about travel. Then the following two kinds traffic flow prediction description pictures are obtained. Figure 13 shows the predicted level of traffic situation for the entire Beijing area at 12:00 on March 28, and Figure 14 shows the predicted level of traffic situation for four different areas on March 28 for the whole day. It can be seen from Figure 13 that the traffic situation level of different regions at 12 noon are different, and the congestion in the central area is significantly higher than that in the suburban area. From Figure 14, we can find that the traffic congestion in the two central areas (12, 24) and (13, 24) is significantly higher than the other two suburban areas. Furthermore, the traffic volume in the central area will be in an increasing phase from 6 am to 12 noon, and will peak at around 12 noon, while there is little change in the suburbs during the day.

VI. CONCLUSION

This paper studied the prediction of traffic flow and developed a F-CNN traffic flow prediction method. To the best of our knowledge, most existing traffic flow prediction studies focus on optimizing models based on the extraction of spatial-temporal features, while ignoring the impact of other influencing factors on traffic flow. Our research proposed a F-CNN approach that uses CNN and TA-FIS. Based on the join of external factors such as meteorology, we also incorporate uncertain traffic accident factors, which obtained through the FIS processing in the basis of original data. Finally, the F-CNN was evaluated on the Beijing taxi dataset and compared the performance with other methods. The experimental results shows that the F-CNN method has better advantages than the other existing methods. Although the results are more optimistic, traffic prediction based on deep learning still requires more future research. In particular, the more effective deep architecture combined with traffic flow theory for urban transport networks. In the future, it will make sense to consider more influencing factors in traffic flow prediction and to use more efficient deep models. Furthermore, we will further study to implements the rules adaptively ability in TA-FIS.

REFERENCES

[1] D. Chen, "Research on traffic flow prediction in the big data environment based on the improved RBF neural network," *IEEE Trans. Ind. Informat.*, vol. 13, no. 4, pp. 2000–2008, Aug. 2017.

[2] X. Ling, X. Feng, Z. Chen, Y. Xu, and H. Zheng, "Short-term traffic flow prediction with optimized Multi-kernel Support Vector Machine," in *Proc. IEEE Congr. Evol. Comput.*, Jun. 2017, pp. 294–300.

[3] Z. Chunmei, X. Xiaoli, and Y. Changpeng, "The research of method of short-term traffic flow forecast based on GA-BP neural network and chaos theory," in *Proc. 2nd Int. Conf. Inf. Sci. Eng.*, Dec. 2010, pp. 1617–1620.

[4] H. Tan, Y. Wu, B. Shen, P. J. Jin, and B. Ran, "Short-term traffic prediction based on dynamic tensor completion," *IEEE Trans. Intell. Transp. Syst.*, vol. 17, no. 8, pp. 2123–2133, Aug. 2016.

[5] W. Chen *et al.*, "A novel fuzzy deep-learning approach to traffic flow prediction with uncertain spatial–temporal data features," *Future Gener. Comput. Syst.*, vol. 89, pp. 78–88, Dec. 2018.

[6] J. An, X. Liu, and G. Wen, "Stability analysis of delayed takagi-sugeno fuzzy systems: A new integral inequality approach," *J. Nonlinear Sci. Appl*, vol. 10, no. 4, pp. 1941–1959, 2017.

[7] Y. Wang and Y. Chen, "A comparison of mamdani and sugeno fuzzy inference systems for traffic flow prediction," *J. Comput.*, vol. 9, no. 1, pp. 12–21, 2014.

[8] B. Ait-El-Fquih and I. Hoteit, "Fast Kalman-like filtering for large-dimensional linear and Gaussian state-space models," *IEEE Trans. Signal Process.*, vol. 63, no. 21, pp. 5853–5867, Nov. 2015.

[9] W. Yu, J. Su, and W. Zhang, "Research on short-term traffic flow prediction based on wavelet de-noising preprocessing," in *Proc. 9th Int. Conf. Natural Comput.*, Jul. 2014, pp. 252–256.

[10] Y. Zhang and Y. Liu, "Data imputation using least squares support vector machines in urban arterial streets," *IEEE Signal Process. Lett.*, vol. 16, no. 5, pp. 414–417, May 2009.

[11] Q. Ye, W. Y. Szeto, and S. C. Wong, "Short-term traffic speed forecasting based on data recorded at irregular intervals," *IEEE Trans. Intell. Transp. Syst.*, vol. 13, no. 4, pp. 1727–1737, Dec. 2012.

[12] A. Stathopoulos, M. G. Karlaftis, and L. Dimitriou, "Fuzzy rule-based system approach to combining traffic count forecasts," *Transp. Res. Rec., J. Transp. Res. Board*, vol. 13, no. 2183, pp. 120–128, 2010.

[13] H. Yin, S. C. Wong, J. Xu, and C. K. Wong, "Urban traffic flow prediction using a fuzzy-neural approach," *Transp. Res. C, Emerg. Technol.*, vol. 10, no. 2, pp. 85–98, 2002.

[14] C. Li, X. Ying, H. Zhang, and X. L. Yan, "Dynamic division about traffic control sub-area based on back propagation neural network," in *Proc. 2nd Int. Conf. Intell. Human-Mach. Syst. Cybern.*, Aug. 2010, pp. 22–25.

[15] H. A. Fayed and A. F. Atiya, "A novel template reduction approach for the k-nearest neighbor method," *IEEE Trans. Neural Netw.*, vol. 20, no. 5, pp. 890–896, May 2009.

[16] S. Sun, C. Zhang, and G. Yu, "A bayesian network approach to traffic flow forecasting," *IEEE Transactions on Intelligent Transportation Systems*, vol. 7, no. 1, pp. 124–132, Mar. 2006.

[17] A. Cheng, X. Jiang, Y. Li, C. Zhang, and H. Zhu, "Multiple sources and multiple measures based traffic flow prediction using the chaos theory and support vector regression method," *Phys. A Stat. Mech. Appl.*, vol. 466, pp. 422–434, Jan. 2016.

[18] G. Hinton *et al.*, "Deep neural networks for acoustic modeling in speech recognition: The shared views of four research groups," *IEEE Signal Process. Mag.*, vol. 29, no. 6, pp. 82–97, Nov. 2012.

[19] H.-F. Yang, T. S. Dillon, and Y.-P. P. Chen, "Optimized structure of the traffic flow forecasting model with a deep learning approach," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 28, no. 10, pp. 2371–2381, Oct. 2017.

[20] J. Zhang, Y. Zheng, and D. Qi, "Deep spatio-temporal residual networks for citywide crowd flows prediction," in *Proc. AAAI*, 2017, pp. 1655–1661.

[21] Y. Wu, H. Tan, L. Qin, B. Ran, and Z. Jiang, "A hybrid deep learning based traffic flow prediction method and its understanding," *Transp. Res. C, Emerg. Technol.*, vol. 90, pp. 166–180, May 2018.

[22] Q. Zhang, T. L. Yang, Z. Chen, and P. Li, "A survey on deep learning for big data," *Inf. Fusion*, vol. 42, pp. 146–157, Jul. 2018.

[23] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Proc. Int. Conf. Neural Inf. Process. Syst.*, 2012, pp. 1097–1105.

[24] X.-X. Wang and L.-H. Xu, "Short-term traffic flow prediction based on deep learning," *J. Transp. Syst. Eng. Inf. Technol.*, vol. 18, no. 1, pp. 81–88, 2018.

[25] S. Du, T. Li, X. Gong, Y. Yang, and S. J. Horng, "Traffic flow forecasting based on hybrid deep learning framework," in *Proc. 12th Int. Conf. Intell. Syst. Knowl. Eng.*, Nov. 2017, pp. 1–6.

[26] Y. Zhang and G. Huang, "A traffic flow prediction model based on deep belief network and genetic algorithm," *IET Intell. Transport Syst.*, vol. 12, no. 6, pp. 533–541, 2018.

[27] L. Zhao, Y. Zhou, H. Lu, and H. Fujita, "Parallel computing method of deep belief networks and its application to traffic flow prediction," *Knowl.-Based Syst.*, vol. 163, pp. 972–987, Jan. 2019.

[28] F. Kong, J. Li, B. Jiang, and H. Song, "Short-term traffic flow prediction in smart multimedia system for Internet of vehicles based on deep belief network," *Future Gener. Comput. Syst.*, vol. 93, pp. 460–472, Apr. 2019.

[29] N. Sharma and S. Sahu, "Review of traffic signal control based on fuzzy logic," *international Journal of Computer Applications*, vol. 145, no. 13, pp. 18–22, 2016.

[30] H. Hellendoorn and R. Baudrexl, "Fuzzy neural traffic control and forecasting," in *Proc. 4th IEEE Int. Conf. Fuzzy Syst.*, vol. 4, Mar. 1995, pp. 2187–2194.

[31] J. Tang, F. Liu, Y. Zou, W. Zhang, and Y. Wang, "An improved fuzzy neural network for traffic speed prediction considering periodic characteristic," *IEEE Trans. Intell. Transp. Syst.*, vol. 18, no. 9, pp. 2340–2350, Sep. 2017.

[32] A. Mozo, B. Ordozgoiti, and S. Gómez-Canaval, "Forecasting short-term data center network traffic load with convolutional neural networks," *Plos One*, vol. 13, no. 2, 2018, Art. no. e0191939.

[33] C. Zhang, H. Zhang, D. Yuan, and M. Zhang, "Citywide cellular traffic prediction based on densely connected convolutional neural networks," *IEEE Commun. Lett.*, vol. 22, no. 8, pp. 1656–1659, Aug. 2018.

[34] J. An, Y. Yu, and J. Tang, "A novel fuzzy approach to urban vehicle location scheme based on fuzzy Kalman filtering," in *Proc. 19th Int. Conf. Artif. Intell. (ICAI)*, Las Vegas, NV, USA, Jul. 2017, pp. 60–66.

[35] W. Chen, J. An, R. Li, and W. Li, "Review on deep-learning-based cognitive computing," *Acta Automat. Sinica*, vol. 43, no. 11, pp. 1886–1897, 2017.

JIYAO AN received the M.Sc. degree in mathematics from Xiangtan University, China, and the Ph.D. degree in mechanical engineering from Hunan University, Changsha, China, in 1998 and 2012, respectively. He was a Visiting Scholar with the Department of Applied Mathematics, University of Waterloo, ON, Canada, from 2013 to 2014. Since 2000, he has been with the College of Computer Science and Electronic Engineering, Hunan University, where he is currently a Full Professor. He has published more than 80 papers in international and domestic journals and refereed conference papers. His research interests include automotive cyber-physical systems, fuzzy systems, intelligent systems, computational intelligence, and big data analysis. He is a member of the IEEE and ACM and a Senior Member of CCF. He is an Active Reviewer for international journals.

LI FU received the B.S. degree from the Wenhua College, Huazhong University of Science and Technology, China, in 2016. She is currently pursuing the M.S. degree in computer technology with Hunan University, China. She is also a member of the Key Laboratory for Embedded and Network Computing of Hunan Province, China. Her major research interests include fuzzy systems and intelligent traffic flow prediction technology. She is a Student Member of the CCF.

MENG HU received the B.S. degree from the Hubei University of Science and Technology, China, in 2016. She is currently pursuing the M.S. degree in software engineering with Hunan University, China. She is also a member of the Key Laboratory for Embedded and Network Computing of Hunan Province, China. Her major research interests include fuzzy systems and intelligent data fusion technology. She is a Student Member of the CAAI and CCF.

WEIHONG CHEN received the master's degree from Hunan University, in 2006, where she is currently pursuing the Ph.D. degree with the College of Computer Science and Electronic Engineering. She is also a Professor with Hunan City University. Her research interests include cyber-physical systems, distributed computing, and machine learning.

JIAWEI ZHAN received the B.S. degree from the Xi'an University of Finance and Economics of Science and Technology, China, in 2017. He is currently pursuing the master's degree in software engineering with Hunan University, China. He is also a member of the Key Laboratory for Embedded and Network Computing of Hunan Province, China. His main research interests include fuzzy systems and data fusion technology.