An Incremental Learning Based Convolutional Neural Network Model for Large-Scale and Short-Term Traffic Flow

Feng Yu, Jinglong Fang, Bin Chen, and Yanli Shao

Abstract—Traffic flow prediction is very important for smooth road conditions in cities and convenient travel for residents. With the explosive growth of traffic flow data size, traditional machine learning algorithms cannot fit large-scale training data effectively and the deep learning algorithms do not work well because of the huge training and update costs, and the prediction accuracy may need to be further improved when an emergency affecting traffic occurs. In this study, an incremental learning based convolutional neural network model, TF-net, is proposed to achieve the efficient and accurate prediction of large-scale and short-term traffic flow. The key idea is to introduce the uncertainty features into the model without increasing the training cost to improve the prediction accuracy. Meanwhile, based on the idea of combining incremental learning with active learning, a certain percentage of typical samples in historical traffic flow data are sampled to fine-tune the prediction model, so as to further improve the prediction accuracy for special situations and ensure the real-time requirement. The experimental results show that the proposed traffic flow prediction model has better performance than the existing methods.

Index Terms—Traffic flow prediction, convolutional neural network, spatio-temporal features processing, incremental learning, active learning.

I. INTRODUCTION

With the rapid development of the world economy and the accelerating urbanization process, the urban road load is increasing due to the significant increase of traffic flow in urban and highways. This leads to a series of problems, such as traffic congestion, which may not only cause people to waste a lot of time on transportation when they go out, but also cause economic losses or even traffic accidents [1]. Traffic flow prediction is a technique for predicting traffic conditions in a region at some point in the future based on its historical traffic flow data and other relevant factors.

The authors are with the School of Computer Science, Hangzhou Dianzi University, Hangzhou, China (e-mail: hdu_yufeng@163.com, fjl@hdu.edu.cn, hertz158123@gmail.com, shaoyanli@hdu.edu.cn).

Especially during rush hour, if the traffic flow is predicted to increase sharply at the next time in a certain region and may cause potential congestion, relevant department can quickly take countermeasures, which will help a lot. On the one hand, the traveler can effectively plan routes based on real-time traffic information to avoid traffic jams and save travel time. On the other hand, the transportation department can take scientific and effective management and guidance measures to keep the roads smooth, thereby improving the resources utilization and achieving an efficient, safe and comfortable traffic environment. Therefore, the traffic flow prediction is an important part of intelligent transportation system (ITS) and has become one of the research hotspots in the field of intelligent transportation.

Traditional traffic flow prediction often uses statistical models and machine learning methods. With the continuous expansion of the scale of traffic flow data, traditional methods are difficult to fit large-scale data sets effectively to obtain accurate prediction results. In recent years, with the rapid development of computer hardware, deep learning has emerged and has been successfully applied in the fields of computer vision, natural language processing, as well as traffic flow prediction. It is increasingly favored by researchers because of its ability to effectively extract features and solve complex problems.

Most of the existing deep learning models for traffic flow prediction only consider temporal and spatial characteristics regardless of uncertainty, although it has a great influence on the traffic flow trend as well. How to introduce the uncertain features to effectively improve the prediction accuracy is an urgent issue in current traffic flow prediction. Furthermore, short-term traffic flow prediction within half an hour or even 15 minutes is an important basis for intelligent traffic control and management, which is extremely practical for solving urban traffic congestion problems. The short-term traffic flow prediction task itself has high real-time requirement and should maintain a certain model update speed to guarantee real-time and high accuracy, which is contradictory with extensive training time requirement of deep learning model.

Based on the above analysis, an incremental learning based convolutional neural network(CNN) model, TF-Net, is proposed to solve the large-scale and short-term traffic flow prediction problem in this study. The rest of this paper is organized as follows. After the introduction of related work, the research motivation and method overview are given in Section III. Section IV and Section V detail the model framework and training and prediction methods respectively.

Section VI describes the related experiments and discussion analysis. The conclusions and future work are finally offered in Section VII.

II. RELATED WORK

Efficient and accurate traffic flow prediction is an important prerequisite and key element for the establishment of ITS, which is of great significance for people's travel route planning and traffic management of transportation department. Nowadays, there have been many studies on traffic flow prediction, and various comprehensive traffic flow prediction models were proposed, which can be roughly classified into three categories: (1) the statistical model based on probability theory and mathematical statistics methods. (2) the data-driven machine learning method. (3) the deep learning method with high predictive performance and model expression ability.

As early as 1977, Javad Abdi proposed autoregressive integrated moving average (ARIMA) to predict short-term highway traffic in a time series forecasting manner [2]. Sheng Jin *et al*. developed a predictive model combining Gaussian mixture model and Kalman filter for online prediction of highway traffic flow value [3]. Gradually, the machine learning methods, such as kNN [4], SVM [5] and Neural Network [6], have replaced the above statistical methods since the traditional statistical models usually have fewer parameters and poorer fitting ability. However, only the temporal characteristics are used in the above methods for prediction, resulting in low prediction accuracy and large prediction deviation. The fast and accurate prediction of short-term traffic flow is still a challenge.

Then, spatial features are deliberately introduced to address low prediction accuracy issue caused by considering only temporal features. Traffic flow prediction is no longer treated as a simple time series prediction problem. Sun *et al*. proposed a new Bayesian network based traffic flow prediction method. The training data which explicitly includes information from adjacent roads are used to analyze current road trends [7]. Based on the temporal and spatial characteristics of traffic flow data, Vlahogianni *et al*. used neural network to predict short-term traffic flow and genetic algorithm for model structure optimization [8]. Most of the above methods are based on the traditional machine learning models, although they have achieved a certain improvement in prediction accuracy, the sample space cannot be well fitted as they are limited by the model expressive ability of large-scale data.

After 2010, the deep learning model was frequently applied to traffic flow prediction due to its good complex feature representation ability. Xue *et al*. combined unsupervised clustering algorithm - k-means with Long Short Term Memory Network (LSTM) model for traffic flow prediction to avoid the limitations of single model performance [9]. Tian *et al*. supplemented the missing data in the data set based on the time series information, and added the parameters needed for data completion to the LSTM model for joint training, and achieved good performance [10]. Based on the spatio-temporal characteristics of traffic flow data, Zhang *et al*. designed the end-to-end deep learning model, ST-ResNet. The residual convolutional neural network is used to predict two-dimensional traffic flow data in each region of the city at a certain time [11].

The application of deep learning in traffic flow prediction makes model have strong expressive ability and sample space fitting ability. Many models use it as feature extractor instead of traditional feature dimension reduction and feature extraction algorithms [12]. But only considering the time dependence and spatial dependence cannot further improve the prediction accuracy. Thus, attempts have been made to introduce the uncertainty characteristics contained in the traffic flow for fusion prediction with spatio-temporal features. Tao Ma et al. proposed a Time Space Threshold Vector Error Correction (TS-TVEC) model for short term traffic state prediction to concurrently address multi-seasonality, non-stationarity, temporal and spatial correlations, and dynamics between traffic variables [13]. Zhang *et al*. used weather conditions and holidays as input uncertain characteristics to combine with traffic flow data for prediction to reduce the impact of uncertainty on the prediction model [11]. Chen *et al*. established a FDCN model by introducing fuzzy theory into the CNN to solve the uncertainty of traffic flow data [14]. Although the impact of uncertainty is considered in the above methods, the model should be extensively modified in order to fit the uncertainty of the data, i.e., adding model branches or adding uncertainty analysis modules, which increase the model complexity, training time and prediction cost. The traffic flow prediction model, especially the deep learning model based on large-scale training data, requires lots of computational resources and a large amount of training time. The above methods further amplified this defect. As a result, how to increase the prediction accuracy by introducing the uncertainty of traffic flow data without increasing the training cost is still a huge challenge.

III. RESEARCH MOTIVATION AND METHOD OVERVIEW

A. Research Motivation

Traffic flow prediction is regarded as a key part for "Smart City" construction. As mentioned earlier, large-scale traffic flow data is highly complex, and it is subject to time dependence, spatial dependence and uncertainty. Time dependence means that the traffic flow value at a certain moment depends largely on traffic flow data of current time and previous period close to the moment. Spatial dependence refers to the traffic flow value in a certain area is determined by the location of this area and its adjacent areas. However, in the case of considering only time dependence and spatial dependence, the prediction model has low prediction accuracy and large deviation in certain circumstances. This attributes to that traffic flow is affected by many factors, and the uncertainty of these factors causes irregular fluctuations in traffic flow data value, which impedes prediction accuracy.

Generally, uncertainty factors can be divided into external factors and internal factors, as well as objective factors and subjective factors. External factors include weather changes, whether it is a holiday and whether there are large-scale activities or events in the city. Internal factors include basic road characteristics and vehicle operation in the road network and so on. Clearly, there are great uncertainties in these factors. At this stage, internal factors and subjective factors are difficult to predict. Any machine learning even deep learning models cannot accurately predict why someone is going out at a certain moment or what the driver will do at the next moment, but the influence of some external factors can be predicted to some extent.

In addition, traffic flow prediction is a work with high real-time requirement. However, the deep learning model with high prediction accuracy requires a large amount of training time. It will not be effectively applied to the traffic flow prediction if the above time bottleneck cannot be solved.

Based on the above analysis, there are two serious and urgent problems to be solved in traffic flow prediction: the low prediction accuracy caused by neglecting the uncertainty and the time lag caused by the deep learning model with high training cost. Although in Zhang's work [11], two uncertainty factors, i.e., weather and holiday information, are added to the traffic flow prediction model by adding model branches, this approach increases model complexity and time cost. However, the historical traffic flow database is very large and the scale is gradually growing, which requires a long model training period. The model update time is too long to meet the real-time requirement of short-term traffic flow prediction. Therefore, how to introduce the uncertainty to improve the prediction accuracy without increasing the complexity and training cost of the model is still a hard issue.

Fig. 1. Method overview.

B. Method Overview

In response to the above issues, this paper proposes an incremental learning based CNN model - TF-Net model. The method overview is given in Fig. 1, and the specific steps are as follows:

Step 1: Data preparation. Dividing the urban area to be predicted into two-dimensional (2D) areas as 32×32, and organizing the historical data into 2D data according to the regional location. The spatial information is retained to better fit the spatial dependence in the traffic flow. The 2D traffic flow data in the continuous historical time period is stacked into three-dimensional (3D) data by combining the uncertainty data (i.e., bad weather and holiday data) of each area to fit the time dependence and uncertainties in the model input.

Step 2: Model training. The mean-square error loss function is used to measure the difference between predicted value and measured value during the training process, and the corresponding value is gradually reduced by using the optimization method. The validation set is used to measure model performance when a certain number of training iterations reaches. When the model performance reaches the bottleneck, the training process is terminated according to the early stopping strategy.

Step 3: Model incremental update and prediction. Generally, it is very time consuming and computationally expensive to reconstruct the prediction model from scratch by directly increasing the data set. Paradoxically, the short-term traffic flow prediction has higher real-time requirement, especially when special traffic conditions like traffic accident or foggy weather occur. In order to improve the training efficiency and prediction accuracy under this situation, the idea of combining active learning with incremental learning [15] was introduced to fine-tune the prediction model to improve the predictive performance by combining typical samples and atypical samples in the same situation in the historical database with current traffic flow data. This is consistent with the real-time requirement of short-term traffic flow prediction since only small-scale data is used for fine-tuning.
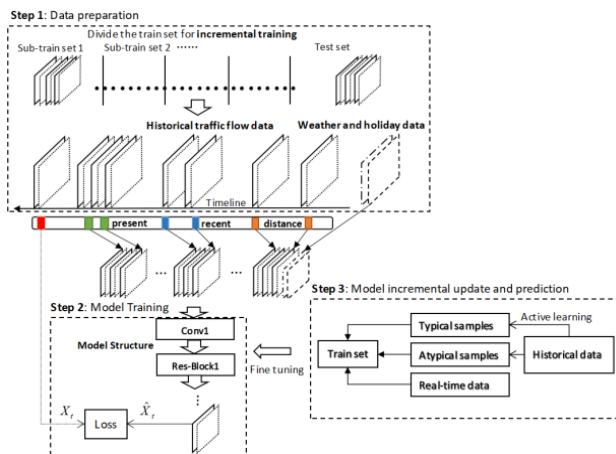
IV. TF-Net Model

A. Problem Formulation

The goal of this study is to predict the large-scale traffic flow of a certain region at a certain point in the future quickly and accurately. To a certain extent, the traffic flow prediction problem can be abstracted as a time series prediction problem. However, the traditional time series prediction models including statistical models like ARIMA model and deep learning models like LSTM both cannot effectively deal with the spatial characteristics in traffic flow data. Therefore, the CNN based TF-Net model is proposed to simultaneously fit the spatio-temporal characteristics of traffic flow data. The reasons are as follows:

(1) CNN has been proved to have excellent performance on large-scale data in many fields due to its large parameter size of the deep network.

(2) The model input is 3D data including historical traffic flow data, weather conditions and holiday information. Meanwhile, CNN has powerful feature extraction capability and less training cost [16]-[18] thanks to its convolution layer setting which can extract spatio-temporal features from traffic flow data and parameter sharing strategy.

(3) Compared to traditional neural network, CNN can achieve better performance on a smaller parameter scale.

In addition to the spatio-temporal characteristics, the uncertainty factors are also used comprehensively to improve the prediction accuracy in this study. Based on the above analysis, here, the whole city is divided into $I \times J$ grids, and the whole traffic flow data can be defined as tensor form: $X \in R^{T \times K \times I \times J}$, where T is time, K is the flow type (input flow or output flow), I is longitude and J is a dimension. Moreover, $X_T \in R^{K \times I \times J}$ represents traffic flow data at time T, $Y_T \in R^{I \times J}$ denotes the corresponding weather conditions., while $Z_T \in R^{I \times J}$ indicates whether the day is a holiday or not. The traffic flow prediction problem can be defined as:

Problem 1. Traffic flow prediction is to predict the future traffic flow data X_n based on the given observed traffic flow data sequences and uncertainties including holidays and weather conditions: $\{X_t, Y_n, Z_n | t = 1,2,3,\ldots\ldots n-1\}$.

B. CNN Based TF-Net Model

1) Model input

As described in Section III.B, the entire two-dimensional (2D) urban area is divided into 1024 sub-areas according to its geospatial factors. Wherein the spatial information is preserved to fit the corresponding spatial dependence. The 2D traffic flow of the latest n consecutive time points in the continuous historical time period is stacked to form $32 \times 32 \times 2n$ 3D data. As shown in Fig. 1, the solid and dashes line portion represent the input and output data, respectively. This 3D data is the historical traffic flow data portion of the input data.

Furthermore, the weather conditions and holiday information are deliberately introduced into the model to improve its prediction accuracy. Similar to the urban traffic flow data, the weather information is organized as 32×32 2D data as well. The weather conditions are divided into 17 categories, and the value from small (0) to large (16) indicates that the severity of the impact of weather conditions on traffic. For example, 13~16 represent heavy snow, heavy fog, sandstorms, and amnesty. Therefore, each numerical point in the above 2D data represents a weather condition of corresponding predicted region. The holiday information is relatively simple. If the predicted time point is a holiday, it is a 2D matrix of all 1, otherwise it is a matrix of all 0. Unlike Zhang's work [11], instead of using additional model branch to specifically deal with weather information and holiday information, the above-mentioned uncertainty features are artificially extended to 2D information and stack with traffic flow data. That is, historical traffic flow data ($32 \times 32 \times 2n$) is stacked with weather conditions and holiday information (shape is $32 \times 32 \times 2$) as the model input ($32 \times 32 \times (2n+2)$). This can reduce model complexity and training cost, which is also a way to introduce non-two-dimensional features into CNN [19].

2) Model architecture

As mentioned above, the CNN based TF-Net model is proposed to achieve the efficient and accurate prediction of traffic flow due to its powerful feature extraction and processing capabilities, and the model architecture is shown

in Fig. 2. In the mode input, instead of separating historical traffic flow data, weather conditions, holiday information and other uncertain features in different model branches, they are organized as the 3D input using the same CNN for processing. In this way, the convolution operation of each convolution kernel can simultaneously consider the influence of spatio-temporal features of a certain region as well as external uncertainty features, i.e.,weather conditions and holiday information. The advantages are: (1) reduce the redundancy and parameter scale and improve the training speed of the model; (2) fully learn and fit the correlation between spatio-temporal features and uncertainty features. However, the ST-ResNet model [11] treats each feature with different model branches, and fuses the processed data to complete the prediction. This method fully considers the particularity between various features but ignores the correlation between different features. After all, holiday information and weather condition are directly affecting traffic condition.
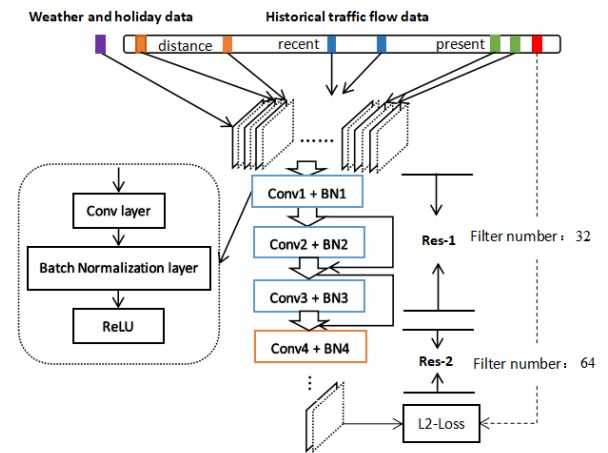

Fig. 2. The convolutional neural network structure.

In the model training process, batch normalization operation is used in each convolutional layer in order to mitigate the gradient disappearance and over-fitting problems [20]. As shown in the left half of Fig. 2, batch normalization operation is used between the convolution operation and the activation function to normalize the output of next layer. Such the input can be converted into a standard normal distribution with a mean of 0 and a variance of 1, which can effectively increase the gradient value in the backpropagation, enhance the information flow and speed up the training convergence. In this study, the entire operational flow within the dashed box is called a convolutional layer operation, which includes convolution operations, batch normalization operations, and activation operations. Note that the convolution kernel size used in this model is 3×3, which is the minimum size that can take into account information around a certain location, reducing the computational cost without degrading the performance and presentation capability [21].

In addition, since the performance of deep learning model is largely determined by the depth of network structure [22], the deep residual structure is used to mitigate the gradient disappearance in the deep convolution layer caused by the excessive depth of the network model [23]. A depth residual

unit is defined as follows:

$$X^{(k+1)} = X^{(k)} + f(X^{(k)}) \tag{1}$$

where $X^{(k)}$ and $X^{(k+1)}$ represent the input data and output data of the k-th layer convolutional layer, and f represents the data processed by the k-th layer convolutional layer. As is shown in Fig. 3, the residual structure can be divided into two residual unit models according to the number of convolution kernels (32 and 64).
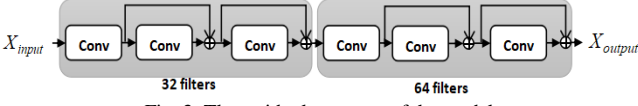

Fig. 3. The residual structure of the model.

The loss function is defined as follows:

$$L(Y, f(X)) = \sum_{i=1}^{n} (Y - f(X))^2 \tag{2}$$

where X is the model input, Y and $f(X)$ represent the measured value and the predicted value, respectively.

In addition, it is necessary to introduce a regularization term in the loss function to reduce over-fitting. Therefore, L2 regularization is specifically used to prevent the model from getting into a serious over-fitting state.

$$L(Y, f(X)) = \sum_{i=1}^{n} (Y - f(X))^2 + \alpha \Omega(\theta) \tag{3}$$

where $\Omega(\theta)$ represents all the parameters in the model and α represents L2 regularization parameters, which is set to 0.005 in this study.

For traffic flow prediction problems, common machine learning models and statistical learning method are difficult to efficiently learn the spatio-temporal features in 3D data. However, convolution operations in the proposed model can effectively deal with the spatial characteristics of 2D traffic flow data, and the stacking of traffic flow data at different time points (forming 3D data from 2D data) enables the model to complete the time fitting. In addition, the weather conditions and holiday information are stacked with the historical traffic flow data to introduce some uncertain factors into the model for learning to improve the prediction performance. Here, the increase in training cost and complexity of the model can be neglected. Therefore, the CNN based TF-net model proposed in this study can effectively fit the time dependence, spatial dependence and uncertainty from traffic flow data to get higher prediction performance.

V. INCREMENTAL LEARNING BASED MODEL TRAINING, UPDATE AND PREDICTION

A. Incremental Learning Based Model Training and Update

At this stage, the training cost of the deep learning model depends on the model architecture and training data size, and

it often requires a large amount of training time and a certain scale of computing resources. As a result, it is difficult to meet the real-time requirement directly to the short-term traffic flow prediction task. It is an important and challenging task to maintain the prediction performance while shortening the training time. Generally, the training cost is expensive if the newly data is directly added to the historical data set as new training data to train the model from scratch. This study aims to propose a traffic flow prediction model that takes into account historical traffic flow data, newly arrived data and the related external features simultaneously to improve the model update speed and the prediction performance in special cases while meeting real-time requirement.

Based on the above analysis, the proposed TF-Net model uses incremental learning rather than batch learning to realize the model training. The historical traffic flow data set is divided into multiple sub-train sets and one test set, and the model is iteratively trained in the sub-train sets. Each sub-train set is further divided into train set and validation set. The train set is used for training and the validation set is used to verify current model performance. When the model performance of multiple iterations on the current sub-train set is not improved continuously, the model is transferred to the next sub-train set for training. In terms of model training, Adam optimization algorithm with dynamic learning rate is applied to adjust the model parameters. Firstly, compared with the gradient descent algorithm, Adam algorithm can dynamically adjust the learning rate with the increase of the number of iterations and accelerate the model convergence process. Secondly, forgetting problem generally exists in the incremental training process. In the iterative training process, the model will intermittently forget the knowledge obtained from the previous sub-train set, resulting in the performance degradation. Studies have shown that the dynamic adjustment of the learning rate setting can slow down the performance degradation caused by the forgetting problem [24].

For model update, the newly generated data can be updated as a new sub-train set by using the above incremental learning method. Since the data set is much smaller in scale than the data set fused with the original training data, the update time of the model will be greatly shortened, which is consistent with the real time requirement of short term traffic flow prediction task as well.

B. Model Prediction in Special Cases

As mentioned before, the efficient and accurate prediction of short-term traffic flow is of great significance for ITS construction, especially in special cases. For example, in the case of heavy fog, rain, snow or major gatherings in the city occur (this study takes holidays as an example), the overall trend and numerical value of traffic flow data are quite different from the usual ones. The predicted value obtained by using the original prediction model has a large deviation from the measured value. Meanwhile, the model must be updated with a certain speed due to the real-time requirement of short-term prediction. Therefore, the combination

approach of incremental learning and active learning is adopted to perform the model fine-tuning to improve the prediction accuracy under the premise of ensuring the real-time requirement.

Different data samples have different contributions to the learning model. If a part of the most valuable data can be selected as train set, it is possible to obtain an efficient and accurate prediction model based on only a small amount of sample data, while reducing the time and space cost of model training. It is necessary to reduce the size of train set used for model fine-tuning to satisfy the real-time requirement. Therefore, the idea of active learning is adopted. Active learning method queries the most useful unlabeled samples through a certain algorithm and then marks them by experts. These labeled samples are used as train set to improve the model performance. Extracting the most useful samples from large data set is a desirable way to reduce the training cost of the model without reducing its performance.

Taking the traffic flow prediction in the holidays as an example, as shown in Fig. 4, considering that the data sample collected in the holidays are the most valuable, these samples are regarded as typical samples, and the rest are atypical samples. Generally, the whole training data set contains holiday information, but these data samples in the holidays account for a small proportion in the entire data set, so the prediction accuracy in this case is relatively low. To address this issue, some typical samples and atypical samples chosen from the historical database by random sampling are used to form historical sub-train set. At the same time, the newly generated traffic flow data during current holidays are merged with historical sub-train set to form a new sub-train set to fine-tuning the model through the incremental learning.
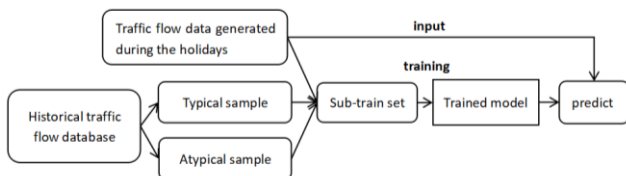

Fig. 4. Active learning in the prediction process during the holidays.

The typical sample in the above sub-train set accounts for a large proportion, and the updated model can obtain more information than original model so as to show better performance. Note that the time-delay caused by the fine-tuning model does not affect the real-time requirement since the use of active learning and incremental learning reduce the size of train set.

VI. EXPERIMENTS AND DISCUSSION ANALYSIS

A. Experimental Settings

We implemented our proposed model in Python with Anaconda3, Pycharm and Tensorflow library. The experiments are run on a 64-b Win10 platform with an Intel Core i7-7700 3.60 GHz CPU and 16 GB memory.

The data set used in this study is Beijing taxicab GPS trajectory data for several consecutive time periods, i.e., 1

July 2013 to 30 October 2013, 1 March 2014 to 30 June 2014, 1 March 2015 to 30 June 2015 and 1 November 2015 to 10 April 2016. The dataset divides the entire Beijing urban area into a grid of 32×32, collecting multiple attributes (i.e., input traffic, output traffic, weather conditions, etc.) of grid data every half hour, and obtaining 48 samples per day. There may be some missing data in the dataset. Here, the data with less than 48 samples sampled in one day is regarded as the missing data, and such data is discarded in the actual data extraction.

To demonstrate the performance of the proposed approach, three groups of experiments are conducted as follows. The first is about overall performance analysis. The second is incremental learning performance analysis, which is done by comparing the prediction accuracy and time cost of batch training and incremental training. And the last one is about real-time prediction performance analysis in special cases. During the model training process, the samples of the last 4 weeks are used as the test set, and the rest are used as the train set. At the same time, 10% of the samples were selected as the validation set in the train set, which was also used as the performance metric of the early stopping strategy. Moreover, three indicators including the mean absolute error (MAE), the mean relative error (MRE) and the root mean square error (RMSE) are used to execute performance evaluation.

B. Experiments and Performance Analysis

1) Model performance analysis

In the first experiment, comparison of several statistical models, machine learning models, deep learning models and the proposed models is made on the same data set. The comparison models that were effective in early and recent traffic flow prediction problems, includes AR, ARIMA, kNN-5, kNN-10, SVM, LSTM, DeepST [25] and ST-ResNet [11]. The comparison result of the above models and the proposed TF-Net model is shown in Table I.

TABLE I: COMPARISON RESULTS OF MODELS

Model	MRE	MAE	RMSE
AR	0.582	30.82	28.24
ARIMA	/	/	22.78
kNN-5	0.283	12.43	17.26
kNN-10	0.276	12.02	16.78
SVM	0.675	18.81	23.43
LSTM	0.554	18.20	23.41
DeepST	/	/	18.18
ST-ResNet	/	/	16.69
TF-Net_Historical_data_only	0.252	9.67	16.84
TF-Net_holiday_weather	0.246	9.47	16.42
TF-Net_week_hour	0.248	9.35	15.68

Note: If the model is normalized or standardized before the training, the MAE and RMSE have no reference value and are not listed.

In Table I, the last three models proposed in this study only differ in the feature selection of model input. The first one only considers the historical traffic flow data, the second one adds the two uncertainties including holiday and weather. The last one adds the two more precise time characteristics including the position of the day in a week and the hour information in one day. We choose the same data set and the same division method of train set and test set, which means the experimental results of the above models are comparable.

As can be seen from Table I, TF-Net model is in the leading position of all models in all three evaluation indicators. While continuously adding new features, the model performance is gradually improving. TF-Net model has a significant improvement on RMSE, which is commonly used in regression problems. Compared with ST-ResNet, the value is improved by nearly 1.

Consider a random region in the city, Fig. 5 and Fig. 6 show the prediction performance of the TF-Net model on the test set for input and output traffic flow. The x-axis in the figure represents the time step, and the y-axis represents the traffic flow value. It can be inferred that the input and output traffic flow at a certain moment in the region are roughly equal in value. TF-Net model can predict the trend of traffic flow very well and the predicted value curve and the observed value curve of the traffic flow coincide highly.
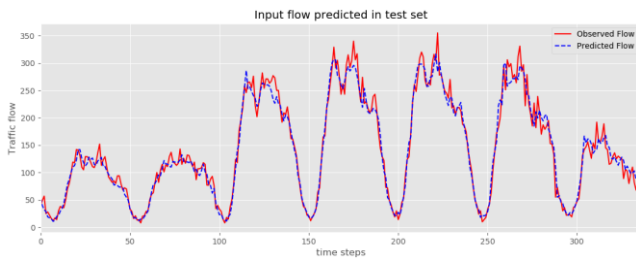
Fig. 5. Model prediction performance in input flow.
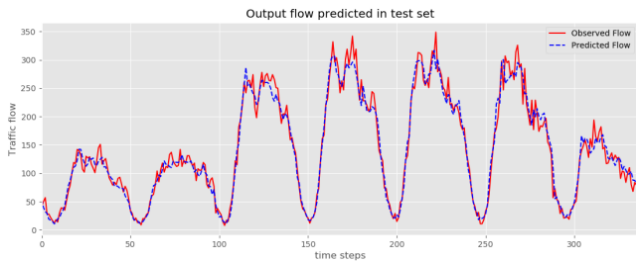
Fig. 6. Model prediction performance in output flow.

2) Incremental learning performance analysis

As described in Section V.A, in order to shorten the model update time when the training data size is increasing continuously, an incremental learning method is introduced to train the model. In this experiment, we first observed the change of the model performance during incremental training period to verify its effectiveness, and then compared the prediction accuracy and training cost between the traditional batch training method and the incremental learning method.

TABLE II: PERFORMANCE OF INCREMENTAL TRAINING

Rounds	MRE	MAE	RMSE
1	0.286	10.990	19.450
2	0.274	10.758	19.017
3	0.265	10.293	18.193
4	0.269	10.163	18.160
5	0.258	9.895	16.945

Table II show the accuracy variation during incremental training. The numbers in rounds represent which sub-train set is currently being trained. It can be seen that the incremental learning method is effective overall since the model prediction accuracy is increasing continuously with

the increase in the number of training rounds of the sub-train set throughout the training process.

Next, we focus on the comparison of prediction accuracy and time cost between the models obtained by the incremental training method and traditional batch training method, as shown in Fig. 7 and Fig. 8. The former is slightly lower in the prediction accuracy than the latter. This is caused by the widespread forgetting problem as mentioned earlier and the inability to obtain complete information on the entire train set. Generally, only one sub-train set can be used for each training, the remaining sub-train sets are invisible to the model, so the model will intermittently forget the previously learned content, causing fluctuations and declines in the performance in the continuous iterative training. But overall, the decline in performance is relatively minor, and it is acceptable considering the improvement in training efficiency. Moreover, the experiment verifies that under the strict real-time requirement, the model performance can be steadily improved after multiple consecutive incremental training, and there is no serious forgetting phenomenon.

On the other hand, the training and update time of incremental training method is significantly less than traditional batch training method. It can be inferred from Fig. 8 the update time cost ratio is about 1:5. Therefore, even though the prediction accuracy is slighted decreased, the training time and update time cost are greatly reduced by using the incremental learning method, which can effectively guarantee the real-time requirement in short-term traffic flow prediction task.
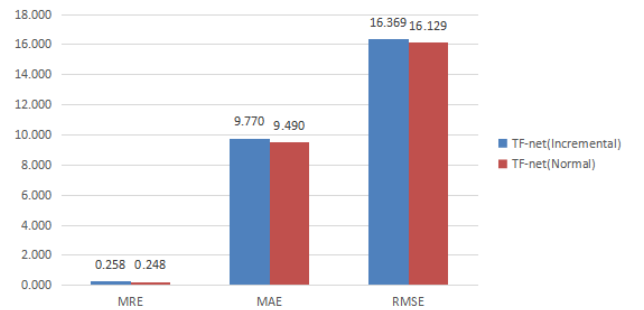

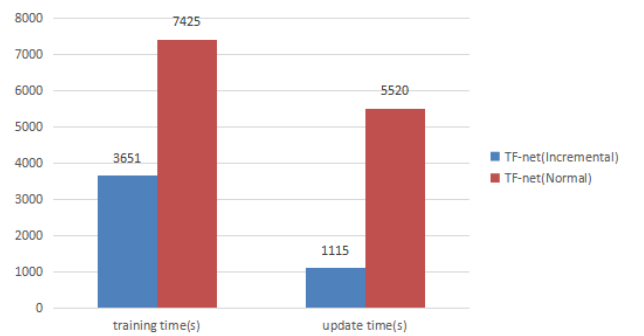Fig. 7. Model prediction performance comparison.

Fig. 8. Model training cost comparison.

3) Real-time performance analysis in special case

Due to the limitation of the data set, the traffic flow samples during the holiday are taken as an example to analyze the prediction accuracy and time cost when special

events affect traffic conditions occur. The traffic flow situation in a certain area of the city during holidays and non-holidays is shown in Fig. 9. The traffic flow data in the holiday is similar with usual traffic flow data on the trend, but the value is relatively smaller than usual. Therefore, it can be used as a special case that affects traffic condition.
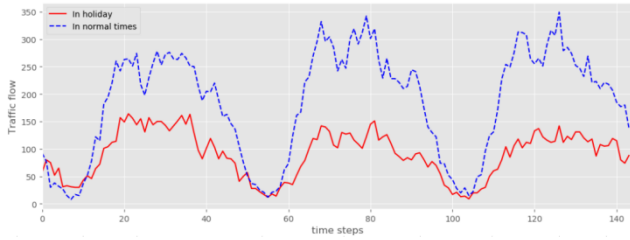

Fig. 9. Traffic flow during holiday and non-holiday.

In this experiment, assuming that the experimental scene is at a certain point in a continuous holiday state, and the traffic flow condition of next holiday period is to be predicted. As described in Section V.B, after fine-tuning the model with sub-train set fused with recent data, typical samples and atypical samples, the prediction accuracy is significantly improved, as shown in Fig. 10, especially in the MRE performance metric, and the other two indicators also slightly decreased.
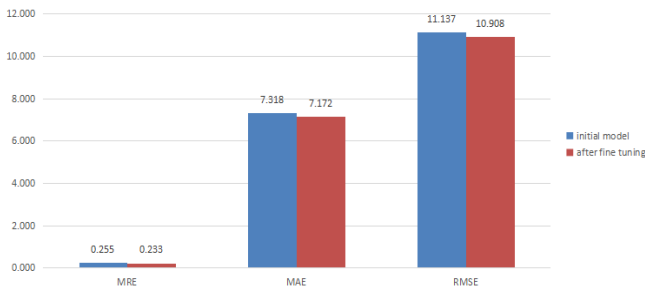

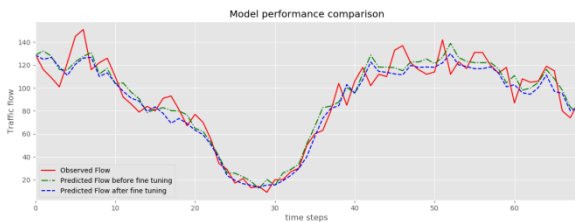Fig. 10. Performance comparison of the model before and after fine tuning.

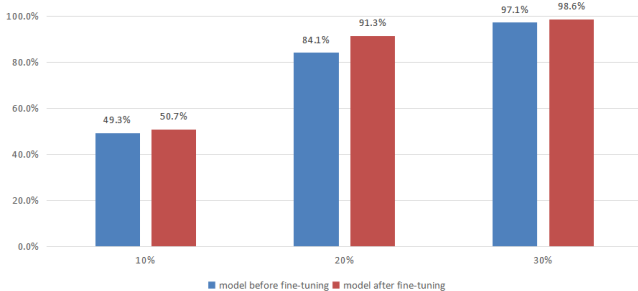Fig. 11. Model performance comparison before and after fine-tuning.

Fig. 12. Comparison result of relative prediction error ratio before and after fine-tuning.

Intuitively, it can be seen from Fig. 11 that after fine-tuning, the prediction effect is better. In Fig. 12, approximately 91.3% of the fine-tuned model prediction values have a relative error value of less than 20%, while it is

84.1% before fine-tuning. So the use of small-scale train set sampled by active learning method to fine-tune the model has significantly improved the prediction performance under special circumstances. Moreover, the average update time can be controlled within 10 minutes. Therefore, this model update strategy can generally meet the real-time requirement of traffic flow prediction task.

VII. Conclusions and Future Work

In this study, an incremental learning based convolutional neural network model, TF-Net model, is proposed to achieve the fast and accurate prediction of large-scale and short-term traffic flow. The main contributions are summarized as follows:

(1) TF-Net model is designed to introduce the external environmental features to further improve the model prediction accuracy without increasing the model complexity and training cost.

(2) During the training process, the traditional batch learning method is replaced by incremental learning in some occasions with high real-time requirement to reduce the update cost and improve the update speed of the model, as well as ensure the real-time requirement.

(3) Based on the idea of combining incremental learning with active learning, the uncertainty factors are considered by using a certain percentage of typical samples and atypical samples in historical traffic flow data. In this way, the model ability to express typical samples can be amplified to obtain a new model quickly with higher prediction accuracy.

The experimental results demonstrate the excellent prediction accuracy and efficiency for large-scale and short-term traffic flow prediction. However, there are still some open problems, such as no deep adjustment to the CNN structure and limited by data set, only a few uncertainties are introduced. In the future work, it is worthwhile to introduce the recent hot AutoML method [26] and more external features into the model.

Conflict of Interest

The authors declare no conflict of interest.

Author Contributions

Feng Yu, Jinglong Fang conducted the research; Feng Yu and Yanli Shao conducted the experiments, analyzed the data and wrote the paper; Jinglong Fang and Bin Chen reviewed the paper; and all authors had approved the final version.

Acknowledgment

of China (No.2015C33003).

REFERENCES

[1] G. D. L. Torre, P. Rad, and K. K. R. Choo, "Driverless vehicle security: Challenges and future research opportunities," *Future Generation Computer Systems*, 2018.

[2] J. Abdi, "Short-term traffic flow forecasting: Parametric and nonparametric approaches via emotional temporal difference learning," *Neural Computing & Applications*, vol. 23, no. 1, pp. 141-159, 2013.

[3] S. Jin, D. H. Wang, C. Xu, and D. F. Ma, "Short-term traffic safety forecasting using Gaussian mixture model and Kalman filter," *Journal of Zhejiang University Science A*, vol. 14.4, pp. 231-243, 2013.

[4] Y. Lv, Y. Duan, W. Kang, Z. Li, and F. Y. Wang, "Traffic flow prediction with big data: A deep learning approach," *IEEE Transactions on Intelligent Transportation Systems*, vol. 16, no. 2, pp. 1-9, 2014.

[5] M. Castro-Neto, Y. S. Jeong, M. K. Jeong, and L. D. Han, "Online-SVR for short-term traffic flow prediction under typical and atypical traffic conditions," *Expert Systems with Applications*, vol. 36, no. 3, pp. 6164-6173, 2009.

[6] X. M. Jiang and A. Hojjat, "Dynamic wavelet neural network model for traffic flow forecasting," *Journal of Transportation Engineering*, vol. 131, no. 10, pp. 771-779, 2005.

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

[8] E. I. Vlahogianni, M. G. Karlaftis, and J. C. Golias, "Optimized and meta-optimized neural networks for short-term traffic flow prediction: A genetic approach," *Transportation Research Part C: Emerging Technologies*, vol. 13, no. 3, pp. 211-234, 2005.

[9] Z. L. Xue and X. Yang, "Multi long-short term memory models for short term traffic flow prediction," *IEICE Transactions on Information and Systems*, vol. 101, no. 12, pp. 3272-3275, 2018.

[10] Y Tian, K Zhang, J Li, X Lin, and B Yang, "LSTM-based traffic flow prediction with missing data," *Neurocomputing*, vol. 318, pp. 297-305, 2018.

[11] J. Zhang, Y. Zheng, and D. Qi, *Deep Spatio-Temporal Residual Networks for Citywide Crowd Flows Prediction*, 2016.

[12] Y. Peng and B.-L. Lu, "Discriminative extreme learning machine with supervised sparsity preserving for image classification," *Neurocomputing*, vol. 261, pp. 242-252, 2017.

[13] T. Ma, Z. Zhou, and A. Baher, "Nonlinear multivariate time-space threshold vector error correction model for short term traffic state prediction," *Transportation Research Part B: Methodological*, vol. 76, pp. 27-47, 2015.

[14] C. Weihong, A. Jiyao, L. Renfa, F. Li, X. Guoqi *et al.*, "A novel fuzzy deep-learning approach to traffic flow prediction with uncertain spatial-temporal data features," *Future Generation Computer Systems*, 2018.

[15] L. Viktor, B. Hammer, and H. Wersing, "Incremental on-line learning: A review and comparison of state of the art algorithms," *Neurocomputing*, vol. 275, pp. 1261-1274, 2018.

[16] F. Qin, N. Gao, Y. Peng, Z. Wu, and A. Grudtsin, "Fine-grained leukocyte classification with deep residual learning for microscopic images," *Computer Methods and Programs in Biomedicine*, vol. 162, pp. 243-252, 2018.

[17] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, *et al.*, "Going deeper with convolutions," in Proc. the IEEE Conference on Computer Vision and Pattern Recognition, 2015.

[18] K. Alex, I. Sutskever, and G. E. Hinton. "Imagenet classification with deep convolutional neural networks," *Advances in Neural Information Processing Systems*, 2012.

[19] D. Silver, J. Schrittwieser, K. Simonyan, I. Antonoglou, A. Huang *et al.*, "Mastering the game of Go without human knowledge," *Nature*, vol. 550, no. 7676, p. 354, 2017.

[20] I. Sergey and C. Szegedy., "Batch normalization: Accelerating deep network training by reducing internal covariate shift," arXiv preprint arXiv: 1502.03167, 2015.

[21] S. Karen and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," arXiv preprint arXiv: 1409.1556, 2014.

[22] F.-W. Qin, J. Bai, and W.-Q. Yuan, "Research on intelligent fault diagnosis of mechanical equipment based on sparse deep neural networks," *Journal of Vibroengineering*, 19, no. 4, 2017.

[23] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. the IEEE Conference on Computer Vision and Pattern Recognition*, 2016.

[24] L. Vincenzo and D. Maltoni, "Comparing incremental learning strategies for convolutional neural networks," in *Proc. IAPR Workshop on Artificial Neural Networks in Pattern Recognition*, Springer, Cham, 2016.

[25] J. Zhang, Y. Zheng, D. Qi, R. Li, and X. Yi, *DNN-Based Prediction Model for Spatial-Temporal Data*, 2016.

[26] H. L. Zhang, S. Kiranyaz, and M. Gabbouj, "Finding better topologies for deep convolutional neural networks by evolution," arXiv preprint arXiv:1809.03242, 2018

Feng Yu is currently working toward the B.E. degree at School of Computer Science and Technology, Hangzhou Dianzi University, Hangzhou, China. His research interests include intelligent transportation systems and deep learning and its application.

Jinglong Fang received his Ph.D degrees in Zhejiang University of Technology, China. Now he is a professor and deputy head of School of Computer Science at Hangzhou Dianzi University, China. He has wide research interests, mainly including machine learning, data mining, pattern recognition and artificial intelligence.

Bin Chen received his Ph.D degree in Tokyo University of Agriculture and Technology, Japan, in 2013. He is a lecture at the School of Computer Science, Hangzhou Dianzi University, China. His research interests include software engineering, machine Learning, and handwriting character recognition.

Yanli Shao received her Ph.D degree in Zhejiang University, China, in 2016. She used to work at the Institute of Automation, Chinese Academy of Sciences from 2017 to 2018. Currently, she is an associate researcher at the School of Computer Science, Hangzhou Dianzi University, China. Her research interests include computer aided design, software engineering, and artificial intelligence.