IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received May 18, 2019, accepted May 30, 2019, date of publication June 24, 2019, date of current version July 22, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2924699

Quantitative Assessment on Truck-Related Road Risk for the Safety Control via Truck Flow Estimation of Various Types

YINLI JIN[1,2], ZHEN JIA[1,2], PING WANG[1,2], (Member, IEEE), ZHU SUN[1,3], KAIGE WEN[1,2], AND JUN WANG[4]

[1]Institute for Transportation System Engineering Research, Chang'an University, Xi'an 710064, China
[2]School of Electronics and Control Engineering, Chang'an University, Xi'an 710064, China
[3]School of Information Engineering, Chang'an University, Xi'an 710064, China
[4]Toll Collection Center for Shaanxi Freeway, Xi'an 710021, China

Corresponding author: Ping Wang (wang0372@e.ntu.edu.sg)

This work was supported in part by the National Natural Science Foundation of China under Grant 51505037, in part by the Key Science and China Postdoctoral Science Foundation under Grant 2016M600814, in part by the Xi'an Intelligent Freeway Information Fusion and Control Key Laboratory under Grant 201805062ZD13CG46, and in part by the Fundamental Research Funds for the Central Universities under Grant 300102329202.

ABSTRACT Traffic conditions of truck flow is one of the critical factors influencing transportation safety and efficiency, which is directly related to traffic accidents, maintenance scheduling, traffic flow interruption, risk control, and management. The estimation of the truck flow of various types could be better to identify the irregular flow variation introduced by various trucks and quantitatively assessed the corresponding road risks. In this paper, the dynamics of truck flow are estimated first. The stochastic and uncertain trucks flow data are obtained in terms of small, medium, heavy, and the oversize truck type and regulated corresponding flow in the time series within five minutes. In order to dig the spatial-temporal correlations behind those data, the deep learning-based method is improved on the basis of the gated recurrent unit (GRU) to estimate the truck flow for various types. To quantitatively assess the truck-related effect for road risk, a multiple logistic regression method is further proposed to classify into safe, risky, and dangerous road risks levels. Different risk level could guide the traffic control and management and traffic information that broadcast drivers to help them to choose travel route. The proposed prediction of the road risk is tested in the randomly selected road segment and shows superior compared to other methods. This could promote road safety in the development of intelligent transport system (ITS).

INDEX TERMS Truck flow prediction, road risk assessment, various types of truck, remote traffic microwave sensor (RTMS), gated recurrent unit (GRU).

I. INTRODUCTION

Accurately predicting traffic flow plays an important role in the development of intelligent transportation system (ITS), which could provide support for numerous transportation services. For example, the estimated traffic flow could provide useful guideline for planning travel path for travelers [1]. Besides, road administrator could foresee congestion conditions in advance based on predicted traffic flows and then allocate road resources [2]. Due to the special characteristics of trucks, traffic flow of trucks often has a significant impact on traffic conditions, road risks, the travel experience of passenger vehicles and so on [3]. Up to 50,400 road traffic accidents involving trucks happened in 2016, China, causing 25,000 deaths and 46,800 injuries, accounting for 30.5%, 48.23%, and 27.81% of the total number of automobile liability accidents, respectively. Moreover, it has been found that the proportion of accidents involving trucks is large under normal nature conditions, and the accident rate involving trucks is increased by 19% in adverse weather conditions, and the probability of serious collisions increased to 2.3 times and 4.5 times [4]. It can be concluded that rucks are prone to traffic accidents, and this problem urgently requires a solution.

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

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see http://creativecommons.org/licenses/by/4.0/

Traffic flow prediction is based on the potential information and features extracted from the historical traffic flow. Hence, the related technologies, including data collection devices, data mining, transmission, etc., have a significant impact on the accuracy of prediction result [5]. In recent years, enormous traffic detectors have been deployed on freeway networks, for example, remote traffic microwave sensors (RTMS), video detector, radio radar, ultrasonic detectors, infrared detectors, etc., which have collected ample and a tremendous number of traffic data. Also, with the technology of development, the accuracy and diversity of data have been greatly improved. The next question is how to efficiently and reasonably use those data in modern ITS.

In the process of predicting traffic flow, the most crucial question is to think about how to extract the characteristics of traffic flow completely. Moreover, the features of nonlinearity and time sequence in the traffic flow data make the prediction difficulty significantly increased. A valuable tool to overcome this problem is machine learning. The motivation of machine learning lies in the establishment and simulation of a neural network for human brain analysis and learning, which mimics the mechanisms of the human brain to interpret data such as images, sounds, and text. The development of machine learning has played an essential role in promoting the prediction of traffic flow and has made significant achievements.

Assessing road risk is a crucial issue in the field of transportation. It mainly evaluates road safety performance based on traffic data to predict the type of collision [6] and the severity of accidents, to implement safety control measures, which is of great significance in preventing traffic accidents. At the same time, fine-grained truck flow data can be applied to support passenger and truck separation strategies [7], and predict the speed of traffic and the number of traffic accidents [8], [9]. It can be concluded that the application of truck flow data to freeway risk assessment will yield considerable improvements.

In this paper, the two-month truck flow data collected by RTMS are thoroughly pre-processed using in practice. Stochastic and uncertainty of the trucks are regulated in terms of small, medium, heavy and oversize truck types, and the fine-grained truck flow data are pre-processed to traffic flow at 15-min time interval. Then, the GRU method is improved to accurately estimate the traffic flow for each type of truck. Finally, to quantitatively assess the truck-related effect for road risk, a multi-nominal logistic regression method is proposed to classify into safe, risky and dangerous road risks. The rest of the paper is managed as follows. Section II introduces some of the previous work on traffic data, traffic flow prediction, and road risk assessment. Section III describes the principle of the improved GRU algorithm and its network structure, multi-nominal logistic regression, and the overall architecture of this research. The results of the experiment are presented in Section IV, including data sources, prediction on fine-grained truck flow, and assessment on truck-related future road risk. The last part is the conclusion and discussion.

II. RELATED WORK

This section provides some previous research achievements about traffic data, traffic flow prediction, and road risk assessment.

A. DATA SOURCE

Researchers have been studying traffic flow prediction since the 1980s [10], in which accurate and diverse traffic flow data is critical to traffic flow prediction based on data-driven approaches. There are a wide variety of traffic flow data in lots of earlier studied, which can be divided into measured data and existing data sets based on data sources. For example, Polson *et al*. used data from twenty-one loop-detectors installed on a northbound section of Interstate I-55 [11]. Yasdi *et al*. measured traffic data by inductive loops in the local roads. They considered six data classes: for ordinary Monday; Tuesday to Thursday; Friday; Saturday; Sunday and Holidays; and one class of special day for expected events [12]. Tan *et al*. collected data through a detector on National, Guangzhou, Guangdong, China from January 1, 2005 to December 30, 2005. The traffic flow data were aggregated and averaged into 1-h periods [13]. Lippi *et al*. and Lv *et al*. extracted traffic flow data from the California freeway performance measurement system (PeMS) [14], [15]. It is concluded that these data have two points in common, (1) they are all collected by loop detectors; (2) traffic data is the total flow that contains all the vehicle sizes. Conclusions indicate that there is little research on traffic flow data prediction for different vehicle sizes.

B. PREDICTING TRAFFIC FLOW

In the past few decades, many prediction models have emerged to solve the problem of traffic flow estimation. These methods can be roughly divided into parametric and non-parametric models. The classical parametric model is the autoregressive integrated moving average (ARIMA), which was applied to predict traffic flow and achieved significant results [16]. Then, a series of enhanced ARIMA models such as fractional-ARIMA, SOM/ARIMA, and switching ARIMA also were applied to forecast traffic flow [17], [18], [19]. In recent years, with the rapid development of neural networks, data-driven approaches have a series breakthrough for travel-time and short-term traffic flow prediction with complex data [14], [20], for example k-nearest neighbor (KNN) [21], support vector machine (SVM) and its hybrid models [22], [23], radial basis function (RBF) [24], deep neural network (DNN) [25], staked autoencoder (SAE) [15], [26], [27]. Further than that, many of hybrid models were proposed for traffic flow prediction, for example, long short-term memory (LSTM) combined with SAE [28], convolution neural network (CNN) combined with GRU. Among them, the recurrent neural network (RNN) is more potent at training for data with time series characteristics [29]. Especially, LSTM and GRU have great achievements in forecasting traffic flow as kinds of extraordinary RNN. Currently, GRU becomes increasingly popular because most of its properties are the same as LSTM and it needs fewer parameters than LSTM so that it is easier to train [30], and it doesn't have separate memory cells, which makes GRU more efficient when training the data [31].

C. QUANTITATIVE ASSESSING ROAD RISK

In recent years, traffic accidents, interruption, and other traffic risks influencing transportation safety and efficiency are becoming increasingly severe with the development of transportation. Therefore, assessment on road risk had become a significant research hotspot in the transportation field.

Traffic states could be represented by at least two of the three foundational traffic parameters: traffic volume, speed, and density [44]. Researchers have found that high variations in speed and high-density traffic conditions, as critical precursors to traffic accidents, represent the riskiest traffic conditions [32]. Besides, a series of speed-related indicators, for example, traffic speed, standard deviation of speed, average speed, and coefficient of speed variation (CSV), are key factors influencing road risk. CSV as a dimensionless variable is often used as an indicator to assess road risk. In addition, CSV is associated with the risk of accidents in any traffic state, and the greater the coefficient of variation, the higher the risk of accidents [33]. At the same time, an appropriate time window can fully capture the unstable state of traffic. Oh *et al.* determined the standard deviation of the speed at 5-min interval as the best indicator of accident prediction [34]. Hyun *et al.* tested the time window length ranging from 5 min to 60 min in 5-min increments and resulted in the recommended period of 15 min, which could show traffic state better [35].

The assessment on road risk can be divided into qualitative and quantitative analysis methods according to research methods. Quantitative analysis requires a direct calculation of risk level values, for examples, Zhang *et al.* analyzed the impact of truck ratio on traffic safety according to the CSV. It was found that the CSV would increase with the increase in the proportion of trucks and concluded that when the truck ratio is between 0.25 and 0.5, the traffic would be more dangerous [36]. Based on traffic flow, weather, geometry, and collision data, Xu *et al.* developed a four-stage stochastic parametric logistic regression model to predict the collision type of accidents quantitatively. The qualitative analysis does not need to calculate risk level values. Golob *et al.* designed a safety performance assessment tool based on cluster analysis and nonlinear canonical correlation analysis to assess the real-time safety level of any traffic flow pattern on the freeway [37]. In further research, they implemented a tool that can be used either in real-time monitoring of the safety level of any freeway traffic flow, or for forecasting the safety aspects of changes in traffic flows [38].

Previous studies showed that speed-related indicators are one of the critical factors in collisions, and traffic flow is one of the essential data to reflect traffic conditions. Moreover, in the research of assessing road risk, the quantitative analysis compared with qualitative analysis will more intuitively display the risk level, and it is more convenient to apply in practice.

III. METHODOLOGY

This section surveys the GRU networks and designs their parameters for predicting the fine-grained truck flow. And the multi-nominal regression method is investigated at the same time. Finally, the overall structure of this research is introduced.

A. IMPROVED GRU

GRU was proposed by Cho *et al.* in 2014 to achieve that each recurrent unit can adaptively capture the correlation of different time scales [39]. The typical structure of the GRU cell and its real output are shown in Fig.1. A typical GRU cell consists of an update gate r and a reset gate z. The update gate is applied to control status information brought to the current moment from the previous moment, while the reset gate is applied to control the degree of ignoring the status information of the previous moment. Also, the candidate hidden state stores the past related information as current memory content by reset gate. The output of hidden layer delivers the information of the current time step to the next time step.

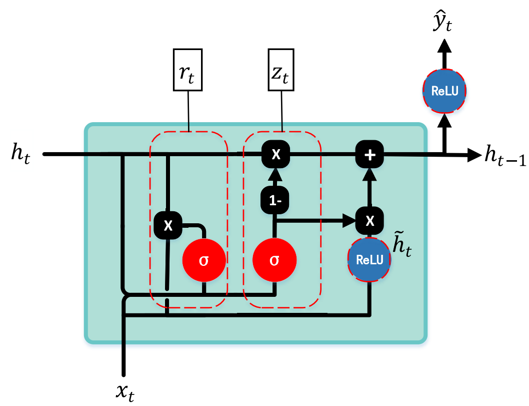


FIGURE 1. An improved GRU cell, replacing Tanh with rectified linear unit (ReLU).

The output of the hidden layer at time t is determined by the hidden layer output at time t-1 and the input series at time t. $X_t = (x_1^t, x_2^t, \ldots, x_n^t)$ represents the input time series at time t, $\hat{Y}_t$ represents the real output at time t, $h_t = (h_1^t, h_2^t, \ldots, h_n^t)$ represents the hidden state output at time t and $h_{t-1} = (h_1^{t-1}, h_2^{t-1}, \ldots, h_n^{t-1})$ represents the output of the hidden state at time t-1. As shown in Fig. 1, r_t is the value of reset gate, z_t is the value of update gate, and $\tilde{h}_t$ is a candidate hidden state of h_t at time t. Their calculations are defined as follows

$$z_t = \sigma\left(W_z h_{t-1} + U_z X_t\right) \tag{1}$$

$$r_t = \sigma\left(W_r h_{t-1} + U_r X_t\right) \tag{2}$$

$$h_t = (1 - z_t)\, h_{t-1} + z_t \tilde{h}_t \tag{3}$$

$$\sigma(x) = \frac{1}{1 + e^{-x}} \tag{4}$$

It is worth mentioning that the activation function is Tanh in the original structure. In this research, Tanh is replaced by a rectified linear unit (ReLU). ReLU is a kind of nonlinear activation function, which is a typically existent to inject nonlinearity elements for complex data. Compared to Tanh activation functions, ReLU offers some advantages: (1) it is simple and fast to execute, (2) it can effectively mitigate the gradient disappearance problem, (3) it induces sparseness [31], [40]. In fact, the prediction accuracy is significantly improved when using ReLU as the activation function, which is defined as

$$ReLu(x) = max(0,x) \tag{5}$$

So, the candidate hidden state and the real output are calculated by the following formulas.

$$\tilde{h}_t = ReLu(r_t \odot (W_{\tilde{h}} h_{t-1}) + UX_t) \tag{6}$$

$$\widehat{Y}_t = ReLU(W_y h_t) \tag{7}$$

In the above-mentioned Eqs. (1-7), W indicates the weight matrices, so the corresponding weights matrices of r_t, z_t, $\tilde{h}_t$, $\hat{Y}_t$ are W_r, W_z, $W_{\tilde{h}}$, and W_y. U is the weight of x at each gate.

Specifically, the GRU architecture is improved through two strategies. (1) Replacing the Tanh with ReLU in computing the candidate hidden state, (2) adding three fully connected layers after the GRU cell. As shown in Fig.2, the architecture of improved GRU is designed for forecasting fine-grained truck flow, which is consist of a GRU cell, three fully connected layers, and an output layer.

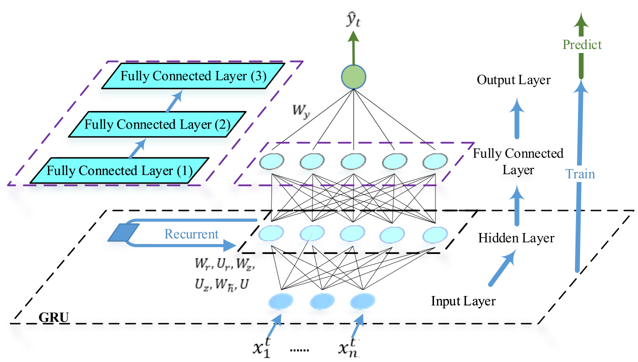


FIGURE 2. The architecture of an improved GRU. The GRU architecture is improved through two strategies. (1) Replacing the Tanh with ReLU in computing the candidate hidden state, (2) adding three fully connected layers after the GRU cell.

A detailed description of each layer is explained below:

Input layer: this model requires one-dimension data (of any digital form) as the input.

Hidden layer: Hidden layer includes six weight matrices to be trained, and the ReLU and sigmoid functions are configured for each recurrent.

Fully connected layers: they are consisting of fully connected (1-3) in the Fig.2, whose activation functions are ReLU. The six weight matrices can be better optimized through these three layers.

Output layer: it is a neural that outputs a predicted value through a weight matrix.

For the sake of minimizing training error, Adaptive gradient (AdaGrad), an amelioration of stochastic gradient descent (SGD), is applied for backpropagation through time. During training models, AdaGrad allocates appropriate learning rate for each variable according to observe their prediction errors, optimizing the problem of overall uniformity of learning rate in traditional SGD. AdaGrad can be calculated by

$$\theta_{i,t+1} = \theta_{i,t} - \frac{\varphi}{\sqrt{G_{i,t} + \varepsilon}} \nabla_{\theta_{i,t}} J(\theta) \tag{8}$$

$$G_{i,t} = G_{i,t-1} + (\nabla_{\theta_{i,t}} J(\theta))^2 \tag{9}$$

where $\theta_{i,t}$ denotes the value of ith parameter at the tth iteration. $\nabla_{\theta_{i,t}} J(\theta)$ denotes the gradient value of ith parameter at the tth iteration. φ is a stable learning rate. ϵ is a smoothing term used to avoid the denominator being 0, which is generally $1e-8$.

To intuitively display the performance of GRU and other related forecast models, three criteria are chosen, mean absolute error (MAE), mean relative error (MRE) and root mean square error (RMSE), to respectively evaluate the performance of neural networks model. Their definitions are (10-12),

$$RMSE = \sqrt{\frac{1}{N} \sum_{t=1}^{N} (\hat{y}_t - y_t)^2} \tag{10}$$

$$MAE = \frac{1}{N} \sum_{t=1}^{N} |\hat{y}_t - y_t| \tag{11}$$

$$MRE = \frac{1}{N} \sum_{t=1}^{N} \frac{|\hat{y}_t - y_t|}{y_t} \tag{12}$$

where $\widehat{y_t}$ is the estimation results, and y_t is the measured data. RMSE measures the deviation between the prediction value and true value. It is more sensitive to the marginal error than other criteria. MAE can better reflect the actual situation of the prediction error. MRE represents the actual size deviating from the true value, as well as reflects the credibility of the prediction better.

B. MULTI-NOMINAL LOGISTIC REGRESSION MODEL

Logistic regression including binomial and multi-nominal logistic regression model is a classical classification method in statistics [41]. If values of the discrete random variable Y is $1, 2, \ldots, K$, then the classical multi-nominal logistic regression model is

$$P(Y = k \,|\, x) = \frac{\exp(w_k * x)}{1 + \sum_{k=1}^{K-1} \exp(w_k * x)}, \quad k = 1, 2, \ldots, K-1 \tag{13}$$

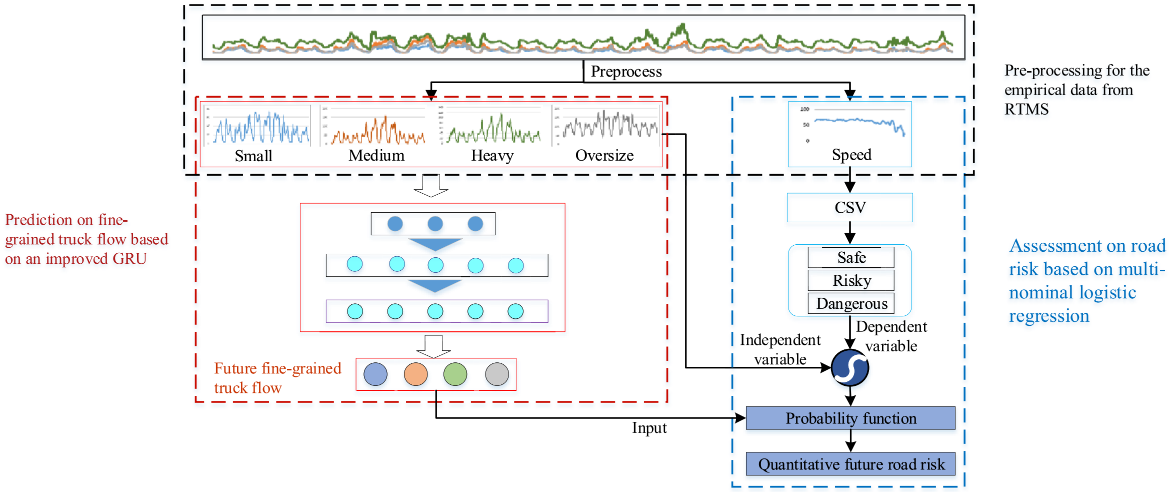


FIGURE 3. Quantitative assessment method on truck-related road risk with fine-grained truck flow includes three aspects. (a) Preprocessing for the empirical data from RTMS; (b) Prediction on fine-grained truck flow based on an improved GRU; (c) Quantitative assessment on road risk based on a multi-nominal logistic regression.

$$P\left(Y=K\mid x\right)=\frac{1}{1+\sum_{k=1}^{K-1}exp(w_k*x)} \tag{14}$$

where $\boldsymbol{x}\in\boldsymbol{R}^{n+1}$,$\boldsymbol{w_k}\in\boldsymbol{R}^{n+1}$. $\mathbf{x}$ represents independent variables affecting classification results, and $\boldsymbol{w_k}$ represents the coefficients of $\boldsymbol{x}$.

This research first utilizes the CSV to classify road risks into three categories, including Safe, Risky, and Dangerous. Then the multi-nominal logistic regression model is applied to construct a relationship between road risk and fine-grained truck flow.

CSV is used as the categorical variable that reflects the degree of dispersion of velocity. The larger the CSV, the greater the magnitude of the velocity change. Conversely, the smaller the magnitude of the velocity change. CSV is computed by

$$CSV=\frac{\sigma_s}{\bar{V}} \tag{15}$$

where σ_s represents the standard deviation of speed, and $\bar{V}$ represents the average of speed.

Therefore, the multi-nominal logistic regression model applied to calculate road risk can be defined as,

$$P\left(Y=1|x\right)=\frac{EXP\left(G_1\right)}{1+EXP\left(G_1\right)+EXP\left(G_2\right)} \tag{16}$$

$$P\left(Y=2|x\right)=\frac{EXP\left(G_2\right)}{1+EXP\left(G_1\right)+EXP\left(G_2\right)} \tag{17}$$

$$P\left(Y=3|x\right)=1-P\left(Y=1|x\right)-P\left(Y=2|x\right) \tag{18}$$

$$P=\max\left\{P\left(Y=1|x\right),P\left(Y=2|x\right),\; P\left(Y=3|x\right)\right\} \tag{19}$$

$$\begin{bmatrix}G_1\\G_2\end{bmatrix}=w*x \tag{20}$$

$$w=\begin{bmatrix}w_1\\w_2\end{bmatrix}=\begin{bmatrix}w_{11}&w_{12}&w_{13}&w_{14}&w_{15}\\w_{21}&w_{22}&w_{23}&w_{24}&w_{25}\end{bmatrix},$$

$$x=\begin{bmatrix}1\\x_1\\x_2\\x_3\\x_4\end{bmatrix} \tag{21}$$

where $P\left(Y=1|x\right)$, $P\left(Y=2|x\right)$, $P\left(Y=3|x\right)$ respectively represent the computed probability of Safe, Risky and Dangerous road status. G_1 and G_2 are two intermediate variables. x indicates matrix of truck flow including truck flow of small, medium, heavy and oversize, and w is the coefficients matrix of x. The state with the highest probability P will be recognized as road risk at that moment.

C. OVERALL STRUCTURE OF QUANTITATIVE ASSESSING FURURE TRUCK-RELATED ROAD RISK

In this section, the overall framework of this research will be presented, which is divided into three parts and shown in Fig.3.

The first part is the preparation and pre-processing of fine-grained traffic data. The traffic data used in this study is the truck flow and traffic speed of different sizes captured by RTMS using in practice. Data pre-processing is mainly for dealing with data loss and data noise. At the same time, the time window of traffic flow is adjusted to capture the traffic conditions accurately. The details of the data will be particularly described in Section IV.

The second part is based on the improved GRU network to estimate fine-grained truck traffic flow. The pre-processed fine-grained truck data is inputted into an improved GRU with appropriate parameters to output future fine-grained truck flow.

The third part is that a multiple logistic regression method is further proposed to quantitatively assess the truck-related effect for road risk. First, the CSV for each time window is calculated based on the pre-processed traffic speed.

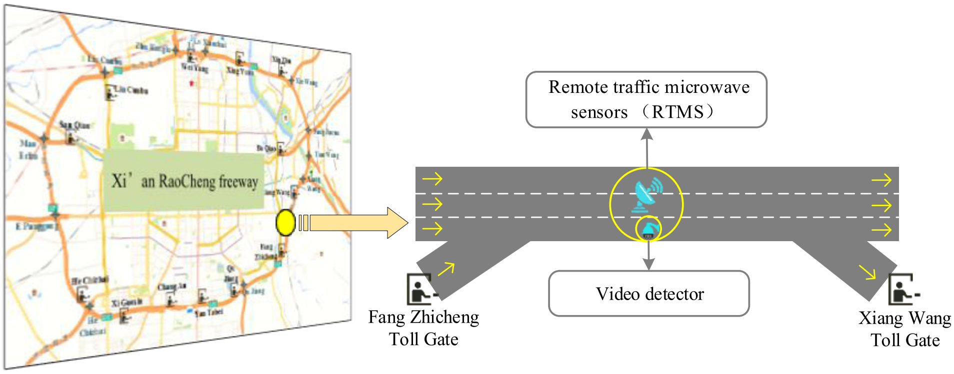


FIGURE 4. The geometric map of the Xi'an RaoCheng freeway and the detection equipment in target section. The equipment mainly relies on RTMS to capture traffic data, and its auxiliary equipment is video detectors in this section.

Second, the road risk is classified into Safe, Risky, and Dangerous based on the computed CSV value. Third, with three types of road risks as the dependent variables, the pre-processed fine-grained truck flow is used as an independent variable, the probability function of truck flow and road risk is obtained based on multi-nominal logistic regression method. Finally, the predicted future truck flow is inputted into the probability function to calculate the future road risk value and risk status.

IV. EXPERIMENTAL RESULTS

A. DATA DESCRIPTION

The data used for the paper is collected by a traffic survey equipment composed of RTMS, a video camera, a microprocessor-based computer, and video processing software, which located at Xi'an Ring Freeway from Fang Zhi Cheng to Xiang Wang Toll Gate. The equipment mainly relies on RTMS to detect traffic data, and its auxiliary method is video detection. It could identify the traffic flow of various vehicles, including small car and truck, medium car and truck, heavy truck, and record corresponding traffic flow, speed, time occupancy rate, etc. in the time series within five minutes. Moreover, its accuracy is much higher than loop detection equipment. The detailed information is shown in Fig.4.

To estimate dynamics of truck flow, the truck flow data of days from 01 April to 31 May 2018 is acquired as an experiment dataset and pre-processed with 15-min time intervals. The data from 01 April to 17 May 2018 are used as the training dataset, while the rest is testing dataset. In this research, stochastic and uncertainty of the trucks are regulated in terms of small, medium, heavy and oversize truck type, which are the same as original data. According to the standard 'General Office of the Ministry of Transport[2010] 205 [42]' in China, the truck types are categorized and represented in Table 1.

However, the empirical data obtained from the RTMS exists some problems with lost data at a low flow period and data noise. To avoid unexpected errors, the lost data is supplemented to zero. Afterward, median filtering is applied to alleviated data noise.

TABLE 1. The classification standard of trucks [42].

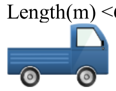
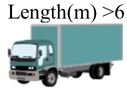
Small truck	Medium truck	Heavy truck	Oversize truck
Two axles truck and Length(m) <6	Two axles truck and Length(m) >6	Three or four axles truck	Five axles truck and above
			

TABLE 2. The parameters of improved GRU.

	GRU-HL	FCL	Epochs	LR
Number	1	3	200	0.01
Neural units	[50]	[50, 30, 10]		

GRU-HL represents hidden layers of GRU, FCL represents fully connected layer, and LR represents learning rate.

B. PREDICTION ON FINE-GRAINED TRUCK FLOW

The experiment is implemented on a desktop computer with Intel i3 3.3GHz CPU, 4 GB memory and Intel (R) HD GPU. As shown in Table 2, the optimal construction and parameters of the GRU model are determined after trial and error to avoid overfitting and under-fitting. Notably, the three fully connected layers behind the GRU can effectively improve forecasting accuracy. In addition, the learning rate will have a significant impact on the forecast results. Excessive learning rate can lead to overfitting, and instead, it will result in underfitting. For the amount of data in our research, it is sufficient that an epoch is set to 200. Also, both the input size and batch size only have little effect on prediction results.

The comparison between the predicted truck traffic flow and the measured data is shown in Fig.5. The truck flow of the small, medium, heavy, oversize is exhibited in a, b, c, and d respectively. The comparison results reveal that the truck flow for all truck size predicted by improved GRU is very similar to the measured data. To validate the efficiency of the improved GRU network, the performance is compared with other neural network approaches, which include RNN,

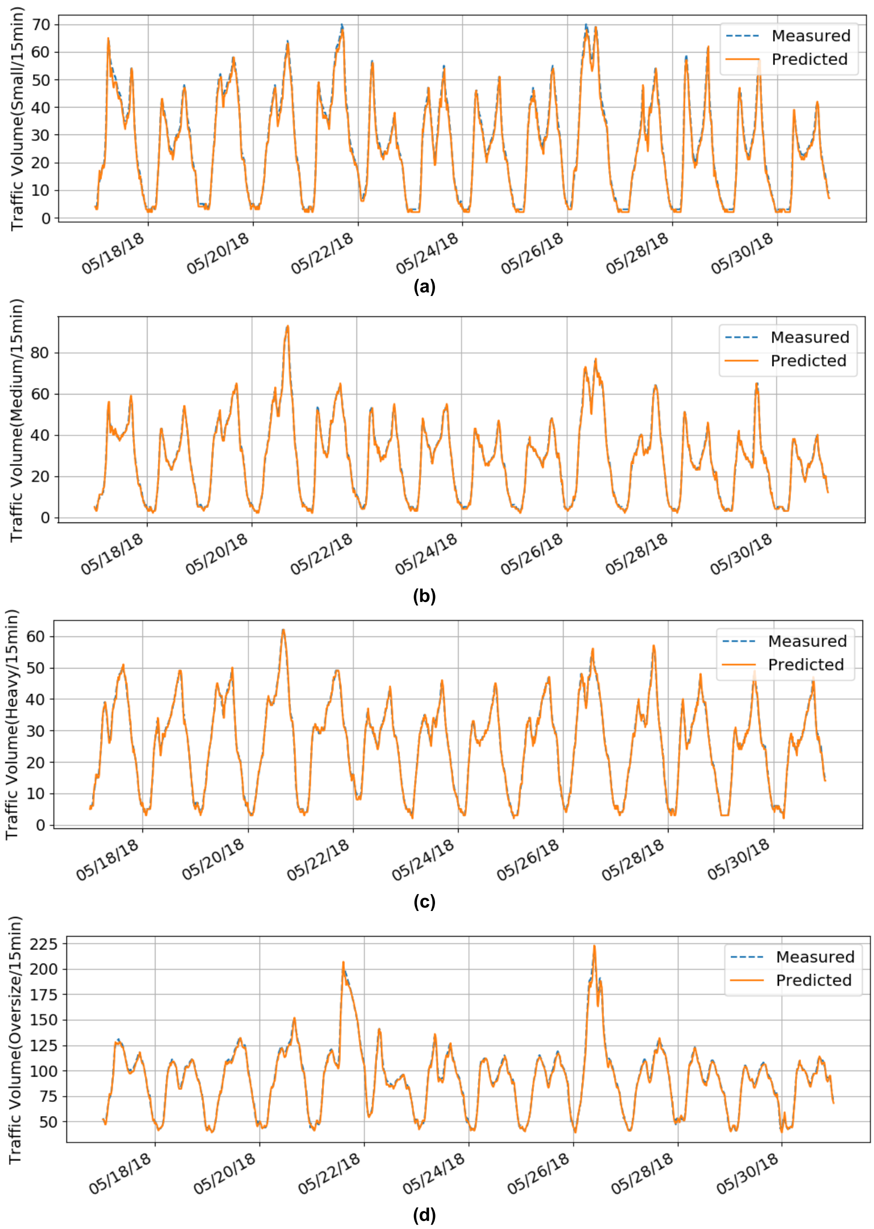


FIGURE 5. **Comparison of measured data and predicted data at different truck types.(a) Small truck, (b) Medium truck, (c) Heavy truck, (d) Oversize truck.**

LSTM, DNN, and SAE. Based on the forecast results of four kinds of truck flow sampled in 15-min increments, the MAEs, RMSEs, and MREs of different prediction models are compared in Table 3. The MREs of improved GRU are 0.05, 0.048, 0.041, and 0.017 respectively. It can be concluded that the improved GRU usually has the minimum error in three criteria compared with other models. Therefore, it can be known from prediction results that the improved GRU is very efficient and credible in the field of predicting traffic flow.

MRE, as one of the assessment criteria, can reflect the dependability of each approach very well by calculating the magnitude of the predicted value to the measured value. Hence, to demonstrate the distinction among five neural networks clearly, the calculated MREs are displayed in Fig.6 by boxplots, including small, medium, heavy, and oversize trucks flow. Based on the three sets of MRE values for each truck size, the boxplots for each model are drawn, showing the maximum, upper quartile, median, lower quartile, and minimum values of the MRE values. As can be seen from four boxplots, the MRE results of the improved GRU model are optimal and most stable. DNN's performance is second only to the improved GRU, and slightly better than RNN's. Besides, the SAE has the worst effect and unstable.

C. ASSESSMENT ON ROAD RISK

In this section, the prediction results are applied to assess the road risk of freeways. Taking the traffic volume and speed of

TABLE 3. Prediction performance of different models for different truck types.

Truck types	Error	GRU	RNN	LSTM	SAE	DNN
Small truck	MAE	**0.854**	0.950	1.322	0.897	0.951
	RMSE	**1.308**	1.359	1.799	1.327	1.349
	MRE	**0.050**	0.060	0.077	0.056	0.062
Medium truck	MAE	**0.843**	0.922	0.947	1.708	0.915
	RMSE	**1.205**	1.278	1.309	2.101	1.229
	MRE	**0.048**	0.055	0.054	0.092	0.050
Heavy truck	MAE	**0.709**	0.737	0.899	0.925	0.758
	RMSE	**1.011**	1.027	1.218	1.166	1.019
	MRE	**0.041**	0.044	0.047	0.058	0.042
Oversize truck	MAE	**1.407**	1.719	1.934	2.522	1.497
	RMSE	**2.006**	2.315	2.621	3.040	2.117
	MRE	**0.017**	0.020	0.023	0.030	0.018

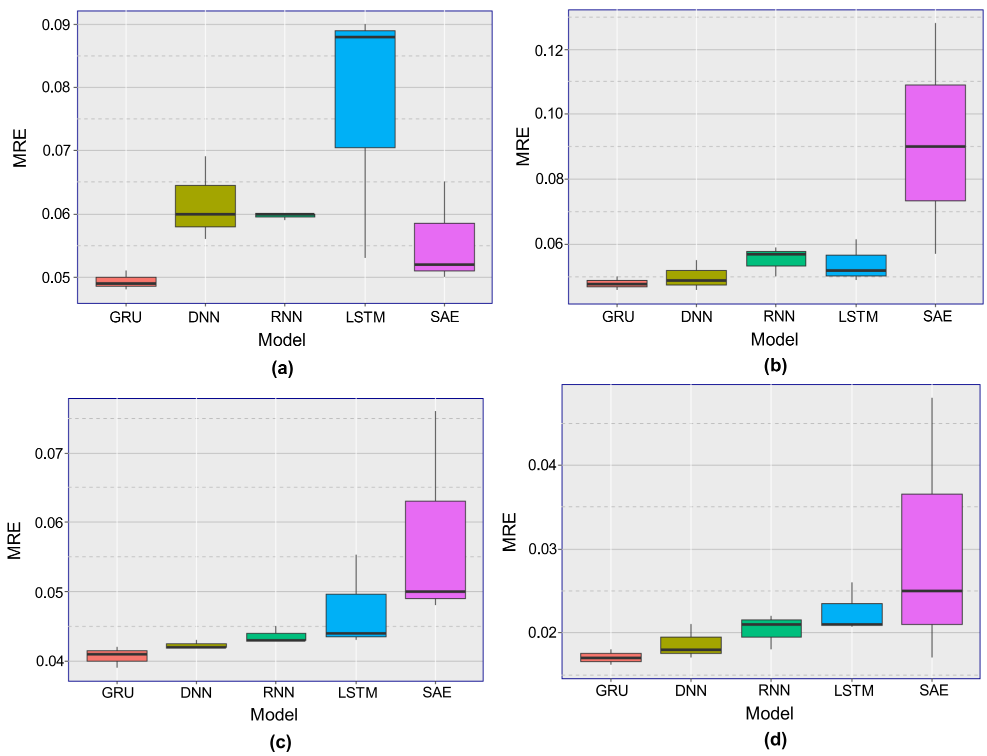


FIGURE 6. The boxplot of MRE error with different truck types. (a) Small, (b) Medium, (c) Heavy, (d) Oversize.

vehicles on April 1, 2018 as an example, the impact of trucks on road risk is explored in this road section.

First, as shown in Fig.7 (a), as the proportion of the total traffic volume of trucks increases, the overall situation of the trucks traffic flow on the road section can be obtained by observing the trend of four kinds of trucks flow. It can be concluded that oversize trucks have the most traffic volume in the truck traffic. Meanwhile, the truck traffic volume of heavy, medium, and small trucks is roughly similar, and far less than the traffic flow of oversize trucks.

As shown in Fig.7 (b), with the proportion of trucks increasing, the speed of oversize trucks doesn't vary significantly, the speed of heavy and medium trucks decreases slightly, and the speed of minivans and passenger cars demonstrates significant reductions and the amplitude of variation of speed increases. Especially after the proportion of trucks piles up to 70%, the speed of passenger cars changed more drastically. Therefore, it can be concluded that the road risk of passenger cars is exceedingly relevant to the truck flow in this road section.

Second, as shown in Fig.8, it is a CSV cartogram of the passenger car in April 1st, 2018. The turning point (CSV = 0.34) represents a change in road risk [43]. Therefore, road risk is divided into three categories based on CSV trends,

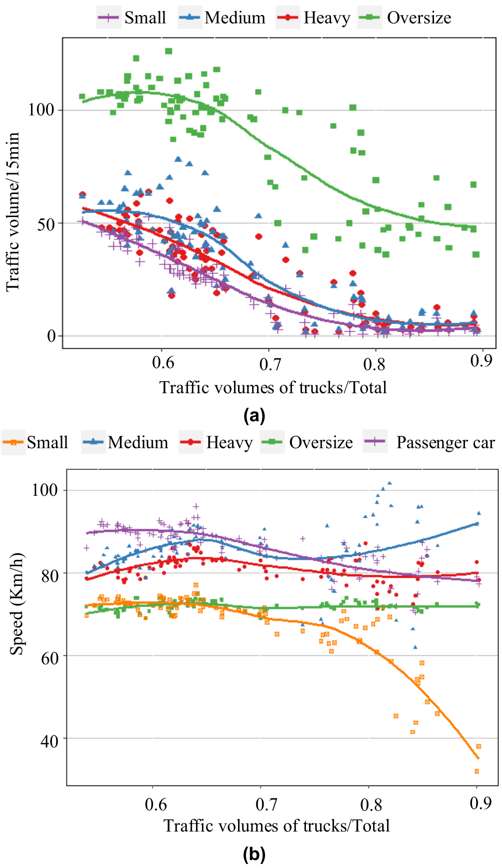


FIGURE 7. Trend of Traffic volume and speed under the proportion of overall traffic volume of trucks in total traffic volume. (a) The traffic volume trend of different truck types, (b) The traffic speed trend of different vehicle types.

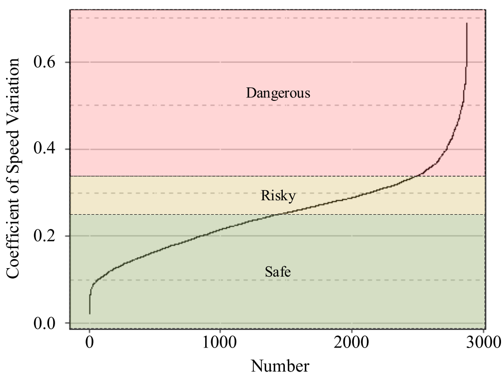


FIGURE 8. Coefficient of Speed Variation (CSV) of passenger cars.

including Safe, Risky, and Dangerous. Behind the inflection point is the Dangerous zone, while before the turning point (CSV from 0.25 to 0.34) is the Risky zone, and the rest is the Safe zone.

Third, the multi-nominal logistic regression is applied to analysis the impact of truck flow on road risk. In mathematical statistics software SPSS, the dependent variable is road risk, while the independent variables are the truck flow of the small, medium, heavy, and oversize. In the regression

TABLE 4. The result of likelihood ratio tests.

Effect	Model Fitting Criteria	Likelihood Ratio Tests		
	-2 Log Likelihood of Reduced Model	Chi-Square	df	Sig.
Intercept	4493.646	41.554	2	**0.000**
Small	4553.987	101.895	2	**0.000**
Medium	4463.995	11.903	2	**0.003**
Heavy	4639.708	187.616	2	**0.000**
Oversize	4463.341	11.248	2	**0.004**

Chi-Square = the difference in -2 log-likelihoods between the final model and a reduced model, df = degree of freedom. Sig. = significance.

TABLE 5. The correct of classification.

Observed	Predicted			
	Safe	Risky	Dangerous	Percent Correct
Safe	1304	181	109	81.8%
Risky	521	246	150	26.8%
Dangerous	80	65	222	60.8%
Overall Percentage	66.2%	17.1%	16.7%	61.6%

analysis of this study, the maximum likelihood estimation is applied and the validity of the model is tested by a likelihood ratio test. The result of likelihood radio tests is shown in Table 4. The likelihood ratio test is a test method that utilizes a likelihood function to assess whether a hypothesis is valid. It can be seen that this model is very effective because all the values of Sig. are less than 0.05. Besides, the correct rate of classification is shown in Table 5, which displays that the correct percentage of Safe, Risky, and Dangerous respectively is 81.8%, 26.8%, and 60.8%. The correct rate of Risk status is so low because most of the Risk states are assessed as Safe status. The reason is considered to be that the boundary between Risk status and Safe status is not obvious enough. The overall correct percentage is 61.6%, which isn't the best result, but it can be relied upon to forecast the road risk. The parameters of the multi-nominal logistic regression function are shown in Table 6, which shows the parameters used to calculate the road risk probabilities of Safe and Risky, then the road risk probability of Dangerous is calculated based on the result of Safe and Riaky. Wald reflects the degree of influence of independent variables on the dependent variable. The larger Wald, the more significant impact. Moreover, when Sig. is less than 0.05, the independent variable has a significant influence on the dependent variable, indicating that the analysis of the problem is valuable. So, the conclusion is that truck traffic flow is significantly related to road risk of passenger cars. Especially in the Risky parameters, the Sig. of the oversize truck flow is the smallest, being 0.002. It can be seen that the traffic flow of oversize trucks has the greatest impact on road risks. Therefore, according to Table 6, The coefficient w can be obtained as

$$w = \begin{bmatrix} -1.845 & -0.098 & 0.048 & 0.142 & 0.013 \\ -1.486 & -0.032 & 0.039 & 0.044 & 0.018 \end{bmatrix} \tag{22}$$

TABLE 6. Value of parameters.

	Parameter	B	Std. Error	Wald	df	Sig.	Exp(B)
Safe	Intercept	**-1.845**	0.294	39.436	1	0.000	
	Small	**-0.098**	0.016	36.758	1	0.000	0.907
	Medium	**0.048**	0.015	10.721	1	0.001	1.049
	Heavy	**0.142**	0.019	58.579	1	0.000	1.153
	Oversize	**0.013**	0.006	4.809	1	0.028	1.013
Risky	Intercept	**-1.486**	0.0289	26.385	1	0.000	
	Small	**-0.032**	0.016	3.983	1	0.046	0.969
	Medium	**0.039**	0.015	7.080	1	0.008	1.040
	Heavy	**0.044**	0.019	5.614	1	0.018	1.045
	Oversize	**0.018**	0.006	9.279	1	0.002	0.019

B = coefficient, Std. Error = standard error, Wald = chi-square, df = degree of freedom, Sig. = significance, Exp(B) = B's index.

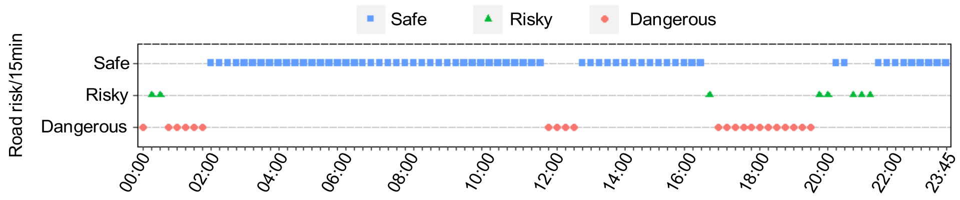


FIGURE 9. Road risk assessed by predicted truck flow in April 1 2018.

So, the mathematical forms of the multinomial logistic regression used in this analysis is as follows.

$$G_1 = -1.845 - 0.098x_1 + 0.048x_2 + 0.142x_3 + 0.013x_4 \quad (23)$$

$$G_2 = -1.486 - 0.032x_1 + 0.039x_2 + 0.044x_3 + 0.018x_4 \quad (24)$$

where G_1 and G_2 are intermediate variables mentioned in equations (16) to (20). Then the three road risks represented by equations (16) to (18) can be calculated through the G_1 and G_2.x_1, x_2, x_3, and x_4 represent the traffic flow of small, medium, heavy and oversize trucks, respectively, at 15-min intervals.

Finally, the road risk calculated by predicted truck flow is shown in Fig.9, from which the state of road risk can be obtained at every moment. The figure shows that the road is in a safe state for most of the time on April 1st, 2018. The Dangerous time is mainly concentrated in the 00:00-02:00, 11:45-12:30 and 16:45 -19:45. Risky status take up less time and generally exist at a transitional point between Safe and Dangerous.

V. CONCLUSIONS AND DISCUSSIONS

This study aims to estimate the traffic flow of various trucks and identify the future changes of road risk. A method for quantitative assessing truck-related road risk in freeway was proposed based on estimating fine-grained truck flow by an improved GRU.

Truck flow data from RTMS are regulated in terms of small, medium, heavy, and oversize truck types. The empirical data is slightly processed by median filtering to solve lost data and reduce data noise. It is worth noting the quality of data has a considerable impact on the prediction results. With the pre-processed fine-grained truck flow dataset, the proposed GRU deep-learning network was improved through replacing the activation function and adding three fully connected layers behind it. The prediction accuracy is significantly enhanced after improving the GRU model. Then, the performance of forecast on GRU, LSTM, RNN, SAE, and DNN are compared through the MAE, MRE, and RMSE. The results showed that the estimation results from GRU are superior compared to other methods.

A multi-nominal logistic regression method was further proposed to identify the road risk that related with truck flow. The model assessment results showed that unsafe traffic time can be identified by truck flow in target section with 15-min intervals. In practical application, Different risk levels could guide the traffic control and management and traffic information that broadcast to drivers to help them to choose travel route and assist managers set out passenger and truck separation strategy in advance. The proposed prediction of the road risk is used in the randomly selected road segment to show the advances to promote road safety in part of the development of intelligent transport system.

It can be seen from the experiment that the performance of GRU predicting truck traffic flow is slightly better than RNN, LSTM, DNN, and SAE. In the later stage, we will try to increase the amount of data or change the prediction duration according to different research demands, observe the running results, and further optimize the model. About the application of traffic flow, we apply the traffic data of different trucks to evaluate the road risk. Although a threshold of CSV is obtained, the reliability of the threshold remains to be further studied. In the later work, the traffic flow, speed, time occupancy rate and other information of passenger cars and trucks with multiple sections will be integrated, and the comfort and driving psychology of the small vehicle drivers will be collected in the form of questionnaires. Finally, the above information is comprehensively considered for conducting an assessment to identify periods and sections that impact the comfort and safety of drivers.

ACKNOWLEDGEMENTS

The authors appreciate the data provided by Toll Collection Center for Shaanxi Freeway, P. R. China. They would like to express appreciation to Miss Yiwen Gao, Mr. Erlong Tan, Mr. Wanrong Xu, Mr. Wenbang Hao, and Mr. Yiming Zhai for their help to collect data and present work in various work.

REFERENCES

[1] J. Yuan, Y. Zheng, X. Xie, and G. Sun, "Driving with knowledge from the physical world," in *Proc. ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, Aug. 2011, pp. 316–324.

[2] 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.

[3] S. Peeta, P. Zhang, and W. Zhou, "Behavior-based analysis of freeway car–truck interactions and related mitigation strategies," *Transp. Res. B, Methodol.*, vol. 39, no. 5, pp. 417–451, Jun. 2005.

[4] M. A. Mohamed, R. Franke, K. Ksaibati, and D. S. Shinstine, "Effects of truck traffic on crash injury severity on rural highways in Wyoming using Bayesian binary logit models," *Accid. Anal. Prev.*, vol. 117, pp. 106–113, Aug. 2018.

[5] B. Ran, P. J. Jin, D. Boyce, T. Z. Qiu, and Y. Cheng, "Perspectives on future transportation research: Impact of intelligent transportation system technologies on next-generation transportation modeling," *J. Intell. Transp. Syst.*, vol. 16, no. 4, pp. 226–242, Aug. 2013.

[6] C. Xu, Y. Wang, P. Liu, W. Wang, and J. Bao, "Quantitative risk assessment of freeway crash casualty using high-resolution traffic data," *Rel. Eng. Syst. Saf.*, vol. 169, pp. 299–311, Jan. 2018.

[7] Y. B. Ji, "Research on dynamic control of separating trucks from passenger cars of bidirectional multi-lane freeway," M.S thesis, Highway Collage Chang'an Univ., Xi'an, China, 2017.

[8] B. Coifman and S. Neelisetty, "Improved speed estimation from single-loop detectors with high truck flow," *J. Intell. Transp. Syst.*, vol. 18, no. 2, pp. 138–148, Apr. 2014.

[9] B. Dadashova, "Simulation study of the effect of decreasing truck traffic flow on safety on almeria-barcelona corridor," Congreso de Ingeniería del Transporte València, Univ. Politècnica de València, Valencia, Spain, Tech. Rep. CIT2016—XII, 2016.

[10] I. Okutani and Y. J. Stephanedes, "Dynamic prediction of traffic volume through Kalman filtering theory," *Transp. Res. B, Methodol.*, vol. 18, no. 1, pp. 1–11, Feb. 1984.

[11] N. G. Polson and V. O. Sokolov, "Deep learning for short-term traffic flow prediction," *Transport. Res. C, Emerg. Technol.*, vol. 79, pp. 1–17, 2017.

[12] R. Yasdi, "Prediction of road traffic using a neural network approach," *Neural Comput. Appl.*, vol. 8, no. 2, pp. 135–142, 1999.

[13] M.-C. Tan, S. C. Wong, J.-M. Xu, Z.-R. Guan, and P. Zhang, "An aggregation approach to short-term traffic flow prediction," *IEEE Trans. Intell. Transp. Syst.*, vol. 10, no. 1, pp. 60–69, Mar. 2009.

[14] M. Lippi, M. Bertini, and P. Frasconi, "Short-term traffic flow forecasting: An experimental comparison of time-series analysis and supervised learning," *IEEE Trans. Intell. Transp. Syst.*, vol. 14, no. 2, pp. 871–882, 2013.

[15] Y. Lv, Y. Duan, W. Kang, Z. Li, and F.-Y. Wang, "Traffic flow prediction with big data: A deep learning approach," *IEEE Trans. Intell. Transp. Syst.*, vol. 16, no. 2, pp. 865–873, 2015.

[16] M. Levin and Y. D. Tsao, "On forecasting freeway occupancies and volumes (abridgment)," *Transp. Res. Rec.*, vol. 773, pp. 47–49, Aug. 1980.

[17] C. Williamson, "Synthetic traffic generation techniques for ATM network simulations," *Simulation*, vol. 72, no. 5, pp. 305–312, May 1999.

[18] H. Chen, S. Grant-Müller, L. Mussone, and F. Montgomery, "A study of hybrid neural network approaches and the effects of missing data on traffic forecasting," *Neural Comput. Appl.*, vol. 10, no. 3, pp. 277–286, 2001.

[19] Y. M. Chen and D. Y. Xiao, "Traffic network flow forecasting based on switching model," *Control Decis.*, vol. 24, no. 8, pp. 1177–1180, 2009.

[20] U. Ali and T. Mahmood, "Using deep learning to predict short term traffic flow: A systematic literature review," in *Proc. 1st Int. Conf. Intell. Transp. Syst. Springe*, vol. 222, 2018, pp. 90–101.

[21] Y. U. Bin, S. H. Wu, M. H. Wang, and Z. H. Zhao, "K-nearest neighbor model of short-term traffic flow forecast," *J. Traffic Transp. Eng.*, vol. 12, no. 2, pp. 105–111, Mar. 2012.

[22] M. Guo, "Search on short time traffic flow forecasting method," *Appl. Res. Comput.*, vol. 25, no. 9, pp. 2676–2678, 2008.

[23] X. Shao, "Application of ARIMA-SVM model in network traffic prediction," *Appl. Res. Comput.*, vol. 29, no. 5, pp. 1901–1903, 2012.

[24] M. A. Jun, L. I. Xiao-Dong, and Y. Meng, "Research of urban traffic flow forecasting based on neural network," *Acta Electronica Sinica.*, vol. 37, no. 5, pp. 1092–1094, 2009.

[25] H. Yi, H. J. Jung, and S. Bae, "Deep neural networks for traffic flow prediction," in *Proc. IEEE Int. Conf. Big Data Smart Comput. (BigComp)*, Feb. 2017, pp. 328–331.

[26] P. Wang, W. Xu, Y. Jin, J. Wang, L. Li, Q. Lu, and G. Wang, "Forecasting traffic volume at a designated cross-section location on a freeway from large-regional toll collection data," *IEEE Access*, vol. 7, pp. 9057–9070, 2018.

[27] 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.

[28] Y. L. Jin, E. L. Tan, L. Li, G. P. Wang, J. Wang, and P. Wang, "Hybrid traffic forecasting model with fusion of multiple spatial toll collection data and remote microwave sensor data," *IEEE Access*, vol. 6, pp. 79211–79221, 2018.

[29] A. Graves, A.-R. Mohamed, and G. Hinton, "Speech recognition with deep recurrent neural networks," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process.*, vol. 38, no. 2003, May 2013, pp. 6645–6649.

[30] X. Wang, R. Jiang, L. Li, Y. Lin, X. Zheng, and F.-Y. Wang, "Capturing car-following behaviors by deep learning," *IEEE Trans. Intell. Transp. Syst.*, vol. 19, no. 3, pp. 910–920, Mar. 2018.

[31] F. Godin, J. Degrave, J. Dambre, and W. D. Neve, "Dual rectified linear units (DReLUs): A replacement for tanh activation functions in quasi-recurrent neural networks," *Pattern Recogn. Lett.*, vol. 116, pp. 8–14, Dec. 2018.

[32] C. Lee, F. Saccomanno, and B. Hellinga, "Analysis of crash precursors on instrumented freeways," *Transp. Res. Rec.*, vol. 1784, pp. 1–8, Feb. 2002.

[33] Z. Li, W. Wang, P. Liu, and C. Xu, "Evaluation of the impacts of speed variation on freeway traffic collisions in various traffic states," *Traffic Injury Prevention*, vol. 14, no. 8, pp. 861–866, 2013.

[34] J.-S. Oh, C. Oh, S. G. Ritchie, and M. Chang, "Real-time estimation of accident likelihood for safety enhancement," *J. Transp. Eng.*, vol. 131, no. 5, pp. 358–363, May 2005.

[35] K. K. Hyun, K. Jeong, A. Tok, and S. G. Ritchie, "Assessing crash risk considering vehicle interactions with trucks using point detector data," *Accid. Anal. Prev.*, 2018.

[36] S. W. Zhang, "Examining the effect of truck proportion on highway traffic safety in free flow state," *J. Wu Han Univ. Technol.*, vol. 39, no. 12, pp. 1–9, 2017.

[37] T. F. Golob, W. W. Recker, and V. M. Alvarez, "Freeway safety as a function of traffic flow," *Accid. Anal. Prevention*, vol. 36, no. 6, pp. 933–946, Nov. 2004.

[38] T. F. Golob, W. Recker, and Y. Pavlis, "Probabilistic models of freeway safety performance using traffic flow data as predictors," *Saf. Sci.*, vol. 46, no. 9, pp. 1306–1333, Nov. 2008.

[39] J. Chung, C. Gulcehre, K. H. Cho, and Y. Bengio, "Empirical evaluation of gated recurrent neural networks on sequence modeling," 2014, *arXiv:1412.3555*. [Online]. Available: https://arxiv.org/abs/1412.3555

[40] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," *Commun. ACM*, vol. 60, no. 2, pp. 1097–1105, 2012.

[41] H. Li, *Statistical Learning Method*. Beijing, China: Tsinghua Univ. Press, 2012, pp. 77–80.

[42] *Notice on Adjustment of Vehicle Type Classification and Conversion Coefficient for Highway Traffic Survey*, Gen. Office Ministry Transp., Beijing, China, 2010.

[43] G. H. Hu, Y. Lin, J. P. Gao, "Real-time safety probability evaluation of the expressway traffic flow," *J. Saf. Environ.*, vol. 15, no. 1, pp. 57–63, 2015.

[44] X. B. Qu, S. Jin, and J. X. Weng, "Analysis of the relationship between aggregated traffic, vol. and, traffic conflicts on expressways," *Transp. A, Transp. Sci.*, vol. 11, no. 7, pp. 1–11, 2015.

YINLI JIN received the B.S. degree in computer science, the M.S. degree in transportation information and control, and the Ph.D. degree in transportation planning and management from Chang'an University, China, in 1995, 2003, and 2010, respectively, where he is currently the Head of the Department of Automation and also the Founder and the Director of the Institute for Transportation Systems Engineering Research. He was a Visiting Professor with the Transportation Research Center, University of Wisconsin-Milwaukee, USA, from 2012 to 2013. His current research interests include traffic information systems, traffic control and management, and intelligent transportation systems.

ZHEN JIA received the B.E. degree in automation from the School of Electric and Control Engineering, Chang'an University, Xi'an, China, in 2017, where she is currently pursuing the M.E. degree in control science and engineering. Her current research interests include intelligent transportation systems, deep learning, and freeway emergency management.

PING WANG (M'11) received the B.S. degree in automation from Shandong University, China, in 2004, the M.S. degree in control theory and control engineering from Shanghai Jiao Tong University, China, in 2007, and the Ph.D. degree in intelligent robotics from Nanyang Technological University, Singapore, in 2011. She was a Postdoctoral Fellow with an international collaboration joint lab at Loughborough University, U.K., and Nanyang Technological University, for three years. She joined the Institute for Transportation Systems Engineering Research, Chang'an University, China, as an Associate Professor. Her current research interests include the applications of control algorithm, artificial intelligence, and intelligent transportation systems.

ZHU SUN received the B.S. and M.S. degrees in automation from Chang'an University, Xi'an, China, in 2010 and 2013, respectively, where he is currently pursuing the Ph.D. degree in traffic information engineering and control. His research interests include intelligent transportation systems and computer vision, and their applications.

KAIGE WEN received the M.S. degree in aerospace engineering and the Ph.D. degree in transportation engineering from Northwestern Polytechnical University, China, in 2006 and 2009, respectively. He was a Teacher with the School of Electric and Control Engineering, Chang'an University, Xi'an, China, until 2009. His current research interests include traffic control and intelligent transportation systems.

JUN WANG received the B.E. degree in computer science from Northwestern Polytechnical University, in 1997, and the M.S. degree in traffic and transportation engineering from Chang'an University, Xi'an, China, in 2008. He is currently a Senior Engineer and the Director of the Toll Collection Center for Shaanxi Freeway. His current research interests include management of freeway, traffic control, and intelligent transportation systems.