

Research Article

Speed Distribution Prediction of Freight Vehicles on Mountainous Freeway Using Deep Learning Methods

Yuren Chen,¹ Yu Chen ,¹ and Bo Yu²

¹Key Laboratory of Road and Traffic Engineering of the Ministry of Education, Tongji University, Shanghai, China

²University of Michigan Transportation Research Institute, 2901 Baxter Rd, Ann Arbor, MI, USA

Correspondence should be addressed to Yu Chen; superpeasant@163.com

Received 9 August 2019; Revised 25 November 2019; Accepted 20 December 2019; Published 10 January 2020

Guest Editor: Joyoung Lee

Copyright © 2020 Yuren Chen et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Driving speed is one of the most critical indicators in safety evaluation and network monitoring in freight transportation. Speed prediction model serves as the most efficient method to obtain the data of driving speed. Current speed prediction models mostly focus on operating speed, which is hard to reveal the overall condition of driving speed on the road section. Meanwhile, the models were mostly developed based on the regression method, which is inconsistent with natural driving process. Recurrent neural network (RNN) is a distinctive type of deep learning method to capture the temporary dependency in behavioral research. The aim of this paper is to apply the deep learning method to predict the general condition of driving speed in consideration of the road geometry and the temporal evolutions. 3D mobile mapping was applied to obtain road geometry information with high precision, and driving simulation experiment was then conducted with the help of the road geometry data. Driving speed was characterized by the bimodal Gauss mixture model. RNN and its variants including long short-term memory (LSTM) and RNN and gated recurrent units (GRUs) were utilized to predict speed distribution in a spatial-temporal dimension with KL divergence being the loss function. The result proved the applicability of the model in speed distribution prediction of freight vehicles, while LSTM holds the best performance with the length of input sequence being 400 m. The result can be related to the threshold of drivers' information processing on mountainous freeway. Multiple linear regression models were constructed to be a contrast with the LSTM model, and the results showed that LSTM was superior to regression models in terms of the model accuracy and interpretability of the driving process and the formation of vehicle speed. This study may help to understand speed change behavior of freight vehicles on mountainous freeways, while providing the feasible method for safety evaluation or network efficiency analysis.

1. Introduction

In the driving process of freight vehicles, driving speed is one of the most important indicators in safety evaluation and efficiency analysis. Currently, the speed prediction model plays an important role in obtaining the driving speed of vehicles [1]. On mountainous freeways, the poor terrain conditions allowed relatively unfavorable road geometry design, and combined vertical and horizontal alignments are not uncommon in mountainous freeways, which directly produces many accident-prone road sections [2]. Meanwhile, freight vehicles are always the most critical elements in the cause of accidents. Therefore, it is essential to study the regularity of speed change behavior of freight vehicles.

On the selection of the variables in the input of the speed prediction model, parameters are mostly related to environmental conditions (e.g., road geometry, roadside vegetation, guardrail, and delineator) [3, 4]. On mountainous freeways in China, the unfavorable terrain condition allowed road alignment design with relatively lower standard. Combined vertical and horizontal alignments are not uncommon. With a tradeoff with acceptable cost, such alignments may force drivers to handle high-speed gradients, which can produce higher crash frequency [5–7]. Typical road alignment features include horizontal and vertical curvature, curve length, guardrails, roadside vegetation, and road delineator [8]. The previous studies agree that road geometry plays a dominant role affecting operating speed under most circumstances.

Considering the possible difference between two-dimensional road design method and three-dimensional road scenario, speed prediction on three-dimensional freeway alignment attracts the attention of some researchers in recent years [9, 10]. In order to fully characterize the feature of road geometry, objective characteristics of the road alignment and drivers' subjective perception in a 3D real space should both be considered [11, 12]. Driving speed serves as the most representative variable due to a clear correlation between drivers' perception of the road and behavior change [13–16]. Meanwhile, studies which adopted sight distance as the input of the speed prediction model found that the index is highly correlated to driver's operating speed, as well as driving safety [17–19].

On the selection of the output in the speed prediction models, operating speed or other representative value of speed was selected to be the predicted value of the model [20–23]. In practice, drivers choose their driving speed according to the road condition including road geometry, roadside vegetation, guardrails, and traffic condition in the environment. Design speed of the road does not necessarily mean the actual driving speed. Operating speed, on the other hand, is a statistical value of the speed dataset. The value can be representative under most circumstances, but the detailed information of the whole speed dataset remains unrevealed. Numerous studies have also supported the view that the driving speed is subject to normal distribution or even multimodal distributions, as regarding to road condition and sample drivers. Some of the researchers tried to fit the distribution of speed by multimodal Gaussian or Bayesian distribution and reach favorable fitting performance [24, 25].

On the selection of speed prediction models, regression is the most frequently used method in driving behavior prediction [26–30]. Various regression models such as linear regression, log-linear regression, nonparametric multivariate adaptive regression splines, and parametric logistic regression have all shown their adaption in different practical problems [4–6]. However, the majority of existing studies on speed prediction modeling treat the input as static variables, instead of considering speed changing process in time series. Although a few studies took the geometric features of previous and oncoming elements into consideration [31–33], the models were not able to match the actual driving process when negotiating with the complexity of the road environment. Some researchers also conducted valuable comparative studies based on the classification of road geometric features and road scenarios [12, 13], commonly with a great fitting performance, but not able to mimic actual driving process. For instance, for a cloud model based on fuzzy neural logic system, the input time sequence data were regarded as isolated variables, so the problem of correlations between input sequences remains [15]. Currently, the study on speed prediction model targeting freight vehicles is still insufficient.

Recently, powerful deep learning methods have been applied to the transportation studies. Recurrent neural networks (RNNs), such as long-short term memory (LSTM), were developed based on bionics to capture long-term

temporary dependency [34, 35], which made them effective in transportation studies [36]. Connections were created among hidden layers to transmit information. The application of RNN and its variants in traffic flow prediction and trajectory prediction also proved its advantages in processing time series data in other fields [37, 38]. In driving behavior field, LSTM had been applied to the prediction of vehicle trajectory based on naturalistic driving data and reached an accuracy of 96.83% in a 10 s prediction horizon [39]. A study on vehicle speed estimation was conducted based on in-vehicle accelerometer and gyroscope data. The study compared the robustness of LSTM model with various layers and the absolute error reached 1.61 km/h [40]. There also exist relevant studies such as identifying behavioral change among drivers using LSTM [41, 42]. However, the complex structure of LSTM usually took longer training time. As a relatively new model which was first introduced in 2014, gated recurrent units (GRUs) are derived from LSTMs, with the appearance of a simpler structure and a faster training speed, and are more convenient to solve. So far, GRUs have not been employed for operation speed prediction. The empirical study comparing GRU and LSTM showed mixed results [43–46].

On the study on freight transportation, travel time, break-taking behaviors, emission, etc., are the most popular topics [47–49]. Some studies have taken the freight transportation process as a consecutive sequence, and uncertainty factors in the transportation process were taken into consideration as well [50–52]. These current studies mainly focus on the macroscopic scale; there are few research studies on the microcosmic behavior of freight vehicles such as speeding, accelerating and decelerating, speed change, etc. Meanwhile, the microcosmic behavior is the ultimate cause of the macrofeatures of freight transportation. Driving speed is one of the most critical indicators in safety evaluation and network monitoring in freight transportation. It is essential to conduct relevant studies to explore the mechanism of the microcosmic features of speed change behavior of freight vehicles.

In summary, the speed prediction model connects the complex road environment and driving speed. Road geometry is always the representative variable in road environment, which holds the dominant place in speed prediction. The output of the prediction model varies around driving speed, but the most appropriate way to characterize driving speed is in the way of statistical distribution. On the selection of prediction model, the continuity of driving process should be taken into consideration, which means that the model should allow the input variables to be sequence data, while holding the ability to handle consecutive information at the same time.

Therefore, the objectives of this paper are summarized as follows:

- (1) Road geometry information should be obtained with high precision to provide input variables for the speed distribution model
- (2) Deep learning network should be applied to predict speed changes of freight vehicles in consideration of

the spatial dependencies and temporal evolutions of road geometry

- (3) Driving speed should be obtained and characterized in a way of statistical distribution

This paper is organized as follows. In Section 2, 3D mobile mapping is introduced as a method to obtain road geometry information, driving simulation and the Gauss mixture model are introduced, and RNN and its variants are presented. In Section 3, an empirical study is presented including driving simulation experiment and model training and testing results. In Section 4, the results are presented and analyzed. The last part concludes remarks and scopes of future studies.

2. Methodology

2.1. 3D Mobile Mapping. In order to provide road geometry data as the input variables of speed prediction model, information including road alignment condition, terrain condition, and roadside vegetation are all needed with high precision. 3D mobile mapping then becomes an ideal solution.

The proposed 3D mobile mapping system utilized in the study was model TOPCON IP-S2, which combines 3 SICK 511 laser scanners, 1 high-precision GNSS antenna, 1 IMU module, and a Ladybug panoramic camera. The laser scanners provide LiDAR point clouds and images of the road environment with its sampling frequency being 75 Hz. The resolution of GLONASS and GPS signal received by the GNSS antenna provides the spatial coordinates of the point cloud. The IMU module guarantees the continuity of the GNSS data. The panoramic camera provides consecutive panoramic picture of the road environment with its sampling interval being 5 m. Registration of panoramic pictures and point clouds was made by the mainframe, giving RGB information of the pixels to the points. The facility and a sample of true color point cloud are presented in Figure 1.

In order to obtain adequate information about road geometry, road alignment 3D available sight distance was selected as the main index. Road alignment information includes curvature and curve length on horizontal, vertical, and cross sections, respectively. Fitting of road alignment on horizontal and vertical section information requires a series of points or a multisegment line along the road for reference. Considering that laser scanners could provide the reflection intensity information of target points, landmarkings can be easily distinguished from the point cloud by its higher reflection intensity. Therefore, landmarkings are selected to be reference in the fitting of road alignment. Sampling step on the landmarking was selected to be 5 m to guarantee sample density, as shown in Figure 2.

The fitting of the horizontal and vertical alignment was conducted in Civil 3D, as shown in Figure 3.

The fitting result shows that the average bias with the fitted alignment and the source multisegment line was no more than 0.052 m, which is acceptable in practice.

The information of cross section includes cross section width and cross slope. It is then essential to extract the points

on the road cross section in the massive point cloud. The point cloud consists of consecutive scanlines, which are parallel to road cross sections. Meanwhile, the points on the road are always the lowest points on altitude compared with the surroundings. Therefore, traverse of the scanlines for the consecutive lowest points on the road surface could provide the road alignment information on the cross section. The extraction work was completed in Civil 3D, as shown in Figure 4, where blue points are the points on road surface, and the sampling interval was set to 5 m.

The width of the cross section could be measured easily by the length of the road surface in the scanline. But the calculation of cross slope needs the spatial attitude of the road surface. Considering the high density of point cloud, the points on a scanline are not strictly a line but a narrow surface. Therefore, plane fitting was conducted by the least square method, and the tangent vector $K_T = (m, n, p)$ and normal vector $K_N = (A, B, C)$ of the plane could be calculated. The cross slope φ ($0 < \varphi < \pi/2$) could then be calculated by

$$\sin \varphi = \frac{|Am + Bn + Cp|}{\sqrt{A^2 + B^2 + C^2} \sqrt{m^2 + n^2 + p^2}} \quad (1)$$

Available sight distance (ASD) refers to the longest distance drivers can see on their normal height in their normal driving process, which reflects the visual field supply of road geometry condition. But the measurement of ASD is always challenging since the value of ASD is affected by road alignment design, roadside vegetation, terrain condition, etc. In order to measure ASD accurately, observers and targets are needed to be set in the real space of point cloud model. According to Chinese Design Standard for Road Alignment, the height of target obstacle was set to 0.1 m. Considering the height of the target four-axis freight vehicles, height of observers was set to 2.0 m. Observers and obstacle targets are placed every 5 m along the road. Spatial sight lines are then constructed between observers and targets. The works were completed in the Spatial Factory software, as shown in Figure 5.

The measurement of ASD was conducted based on the actual road alignment. In order to detect collision between sight lines and road infrastructure or roadside vegetation, "Rigid-body" command in Spatial Factory was used to transform the points in the point cloud to cubed rigid bodies with the side length of 0.1 mm so that collision can be detected. Traverse of the sight lines was then conducted till collision. The number of the sight line was then recorded in Spatial Factory. The measurement of ASD could then be conducted by the "Measurement" command in Spatial Factory as shown in Figure 6.

An example of 3D ASD measurement result is shown in Figure 7. The measuring method mentioned above was applied to a 900 m road segment in China, a horizontal curve on a mountainous freeway.

2.2. Driving Simulation. Traditional methods in obtaining driving behavior data include naturalistic driving experiment, video surveillance, and driving simulation experiments. Naturalistic driving experiment provides the opportunity to observe actual driving process. But the result is affected by the

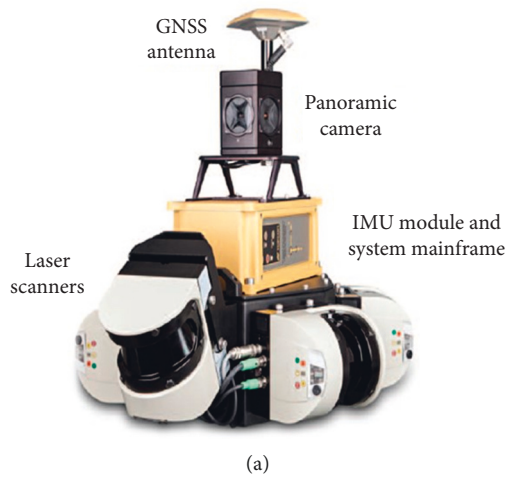


FIGURE 1: 3D mobile mapping facility and true color point cloud.

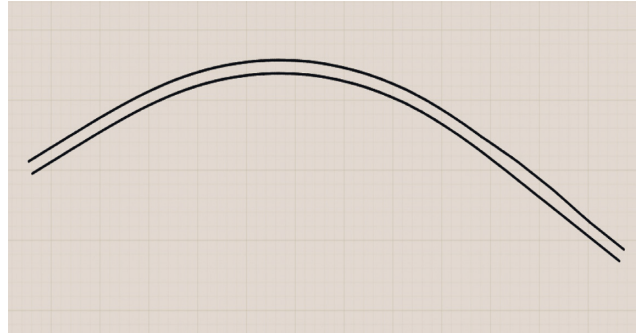
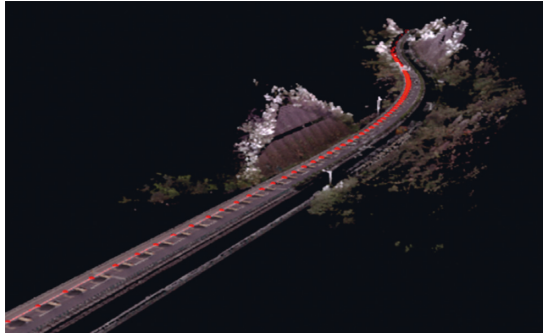


FIGURE 2: Extraction of landmarking from point cloud.

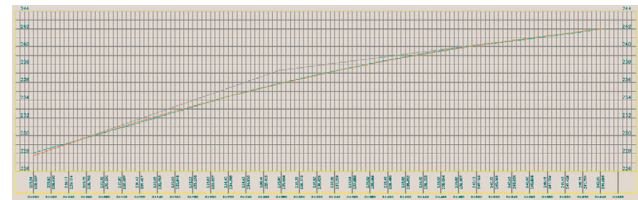
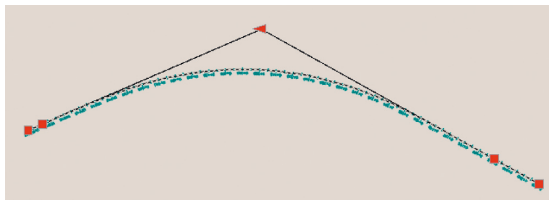


FIGURE 3: Fitting of horizontal and vertical alignment.

unexpected traffic condition, while the cost of the experiment is higher than the other ones. Video surveillance provides large quantity of data, but the continuity of the data sequence is difficult to guarantee. Driving simulation experiment enables participants to drive in a simulated environment with the help of a driving simulator. Data of driving behavior are easy to collect, and the variables in the road environment are easy to control; the cost of the experiment is also relatively low. At the same time, the simulated environment will be with high fidelity with the help of 3D mobile mapping. Therefore, driving simulation experiment was utilized to obtain driving behavior data.

Firstly, the logical layer of the road environment model is constructed in HintCAD with the road alignment information extracted from 3D mobile mapping, as shown in Figure 8.

Then, the model layer was built by 3ds Max and Google Sketchup, and road infrastructures were imported together into the driving simulation software OKTAL SCANNER, as shown in Figure 9.

Finally, scripting of data acquisition and model control was completed in SCANNER software. The steering performance of the vehicle was calibrated by the participants of the experiments so as to ensure the driving experience to be consistent with the



FIGURE 9: Model layer of road environment.

where K is the number of Gaussian distribution in GMM and $\lambda_k, \mu_k, \sigma_k$ are the weight, mean value, and standard deviation of the K th distribution, respectively. Theoretically, when K approaches infinite, the Gaussian mixture model can perfectly describe any kind of statistical distribution. In order to describe the distribution of freight vehicle speed with precision, avoiding the possibility of underfitting by any isolate distribution form, the Gaussian mixture model (GMM) is utilized to characterize the speed distribution of freight vehicles.

In practice, GMM with three or more normal distribution may include one or more distributions with very small weights, which may cause massive unnecessary calculation in the fitting process. Generally, a bimodal GMM can be qualified in the fitting of speed distribution with considerable accuracy and limited variables to be determined. The function of bimodal GMM can be written as

$$f(x) = \lambda_1 * \frac{1}{\sqrt{2\pi}\sigma_1} e^{-((x-\mu_1)^2/2\sigma_1^2)} + \lambda_2 \frac{1}{\sqrt{2\pi}\sigma_2} * e^{-((x-\mu_2)^2/2\sigma_2^2)}. \quad (3)$$

Considering that there are two distributions in bimodal GMM, it is difficult to determine the ascription of the sample points in the training dataset. Maximum likelihood estimation will be no longer suitable to estimate the parameters in the model. Therefore, EM (expectation-maximization) algorithm is applied to fit the bimodal GMM. The algorithm firstly estimates the ascription of sample points (E-step), and then iteration is conducted based on the maximum of likelihood function (M-step). The conduction of E-step is shown in formula (4), where k is set to 0 in the first step, $\mu_1, \mu_2, \sigma_1, \sigma_2$, and λ_1 are initiated to be random values, λ_2 is a dependent variable in which $\lambda_2 = 1 - \lambda_1$, N is the number of samples in the dataset, and m is the number of distributions in GMM. In the fitting of bimodal GMM, the value of m was set to 2.

$$P(m|x_n, \theta^k) = \frac{(\lambda_m/\sigma_m^k) e^{(-(1/2)((x_n - \mu_m^k)/\sigma_m^k)^2)}}{\sum_{m=1}^M (\lambda_m/\sigma_m^k) e^{(-(1/2)((x_n - \mu_m^k)/\sigma_m^k)^2)}}, \quad (4)$$

$$\begin{cases} n = 1, 2, \dots, N, \\ m = 1, 2. \end{cases}$$

Then, the conduction of M-step is shown in the following formulas:

$$\lambda_m^{k+1} = \frac{1}{N} \sum_{n=1}^N P(m|x_n, \theta^k), \quad (5)$$

$$\mu_m^{k+1} = \frac{\sum_{n=1}^N x_n P(m|x_n, \theta^k)}{N\lambda_m^{k+1}}, \quad (6)$$

$$\sigma_m^{k+1} = \left(\frac{\sum_{n=1}^N P(m|x_n, \theta^k) |x_n - \mu_m^{k+1}|^2}{N\lambda_m^{k+1}} \right)^{1/2}. \quad (7)$$

The iteration ends when the inequality in Formula (8) holds. In order to avoid local optimum solutions, the solution should also hold the inequality in formulas (8) and (9). $\varepsilon_\lambda, \varepsilon_\mu, \varepsilon_\sigma$, and ε_p are allowable error ranges of weight, mean value, standard deviation, and maximum likelihood function, and k_{\max} is the maximum iteration number. Generally, the error benchmark could be set to 1%, and reasonable value of k_{\max} could be 100.

$$\begin{aligned} \|\lambda_m^{k+1} - \lambda_m^k\| &\leq \varepsilon_\lambda \\ \|\mu_m^{k+1} - \mu_m^k\| &\leq \varepsilon_\mu \\ \|\sigma_m^{k+1} - \sigma_m^k\| &\leq \varepsilon_\sigma \\ \text{or } k &> k_{\max}, \end{aligned} \quad (8)$$

$$\log P(x|\theta) = \sum_{n=1}^N \log \left[\sum_{m=1}^M \frac{\lambda_m}{\sqrt{2\pi}\sigma_m} e^{(-(1/2)((x_n - \mu_m/\sigma_m))^2)} \right] \leq \varepsilon_p. \quad (9)$$

2.4. Deep Learning Methods. With the discussion of the weakness of traditional regression model, RNN (recurrent neural network) and its variants are ideal candidates with their great capability in handling sequence input data. The basic structure of RNN is shown in Figure 10. The output of the former time step is set to be the input of the latter one,

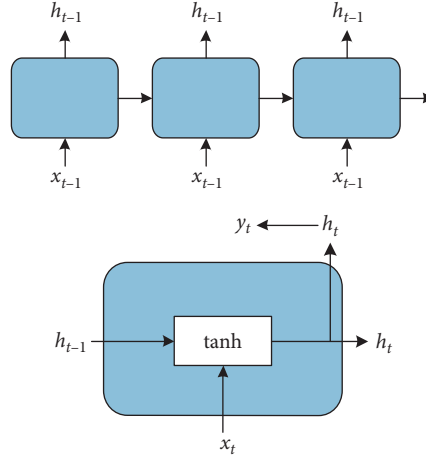


FIGURE 10: Structure of RNN and the memory unit.

and information is transmitted between hidden layers in time domain. Memory units were constructed based on the transmission functions as shown in in formulas (10) and (11).

$$h_t = \tanh(W_{xh}x_t + W_{hh}h_{t-1} + b_h), \quad (10)$$

$$\tanh(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}}. \quad (11)$$

However, with the increase of the data scale, gradient vanishing or explosion will become a common problem in RNN and information may get lost. As a special variant of RNN, long short-term memory (LSTM) improves RNN by joining “input gate,” “output gate,” and “forget gate” into the memory unit to promise the information transmission with accuracy among the hidden layers. The structure of the memory cell is shown in Figure 11. The transmission functions among the gates are shown in formulas (12)–(18).

$$f_t = \text{sigmoid}(W_{xf}x_t + W_{hf}h_{t-1} + W_{cf}c_{t-1} + b_f), \quad (12)$$

$$i_t = \text{sigmoid}(W_{xi}x_t + W_{hi}h_{t-1} + W_{ci}c_{t-1} + b_i), \quad (13)$$

$$\tilde{c}_t = \tanh(W_{xc}x_t + W_{hc}h_{t-1} + b_c), \quad (14)$$

$$c_t = f_t \otimes c_{t-1} + i_t \otimes \tilde{c}_t, \quad (15)$$

$$o_t = \text{sigmoid}(W_{xo}x_t + W_{ho}h_{t-1} + W_{co}c_t + b_o), \quad (16)$$

$$h_t = o_t \otimes \tanh(c_t), \quad (17)$$

$$\text{sigmoid}(x) = \frac{1}{1 + e^{-x}}, \quad (18)$$

where i_t is the status of input gate, which controls the information to update from the former memory unit, f_t is the status of forget gate, which controls the information to forget

from the former memory unit, and o_t is the status of output gate, which controls the information to transmit to the next status.

The delicate designed structure of LSTM could handle the problem of gradient vanishing or explosion. But the complex structure may also increase the burden of training. And it will be easier for LSTM to be overfitting.

Another variant of RNN is GRU (gated recurrent unit). This kind of variant developed the structure of memory unit by coupling the input gate and forget gate in LSTM into an update gate. The reset gate served as the output gate to control the recurrent connections, as shown in Figure 12. The transmission functions among the gates are shown in formulas (19)–(22).

$$r_t = \text{sigmoid}(W_{xr}x_t + W_{hr}h_{t-1} + b_r), \quad (19)$$

$$z_t = \text{sigmoid}(W_{xz}x_t + W_{hz}h_{t-1} + b_z), \quad (20)$$

$$\tilde{h}_t = \tanh(W_{xh}x_t + r_t \otimes W_{hh}h_{t-1} + b_h), \quad (21)$$

$$h_t = z_t \otimes \tilde{h}_t + (1 - z_t) \otimes h_{t-1}, \quad (22)$$

where r_t is the status of reset gate, which decides the information to update from the former memory unit, and z_t is the status of update gate, which controls the information to transmit to the next status.

3. Empirical Analysis

3.1. Experimental Settings. Experiment was conducted on a section of G15 freeway from Xinchang to Baihe in Zhejiang Province, China. The 30 km length test segment is 2-lane mountainous freeway with a speed limit of 100 km/h, as shown in Figure 13.

Firstly, 3D mobile mapping was applied on the road section to obtain information of road geometry and 3D ASD.

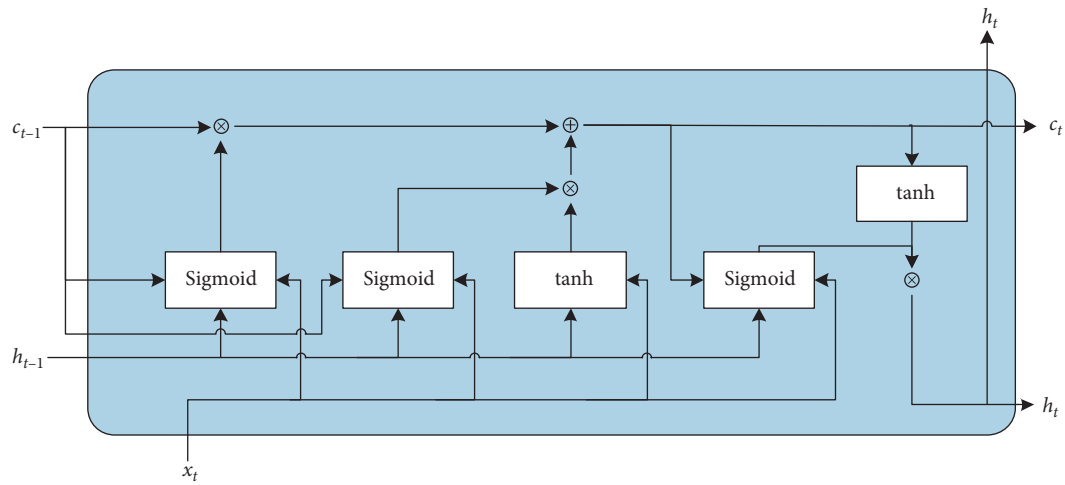


FIGURE 11: Structure of the memory unit in LSTM.

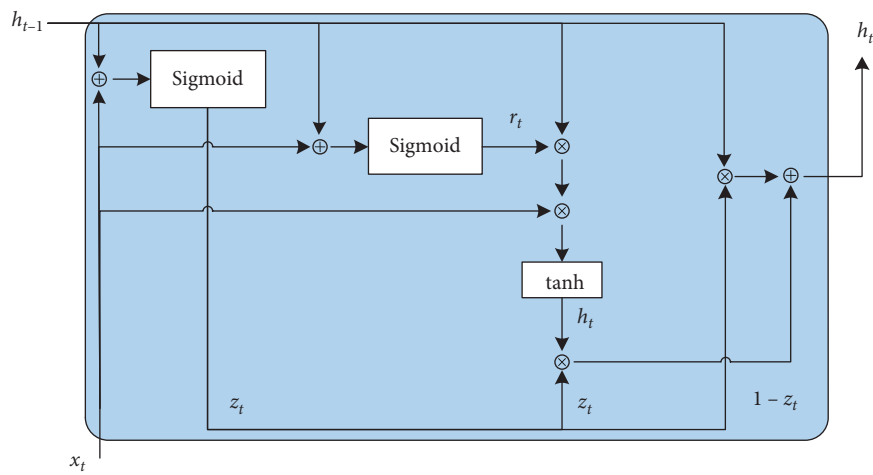


FIGURE 12: Structure of the memory unit in GRU.

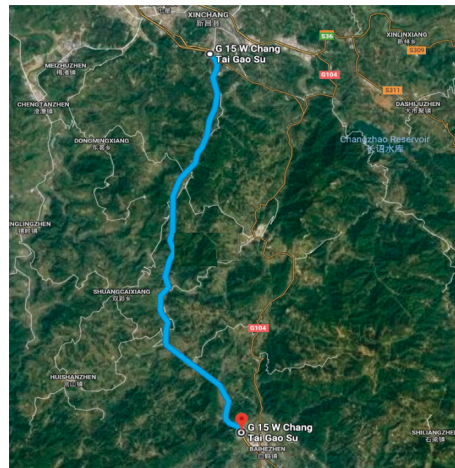


FIGURE 13: The test road section.

Secondly, independent repeated driving simulation experiments were designed to reflect the characteristic of the driver group. In order to obtain enough sufficient training data for the fitting of speed distribution on any road sections, totally 40 freight vehicle drivers were recruited to take the experiment. The information of the participants is shown in Table 1. Considering that there exist a certain number of female drivers of freight vehicles in China, the sample group contained 7 female participants to represent the real condition of the driver group of freight vehicles. The road environment model was built in SCANNER based on the point cloud model. The vehicle type was set to be four-axis freight vehicles.

Each participant was asked to drive through the road section once (total of 40 independent repeated experiments). The participants were required to drive at a safe, reasonable, and comfortable speed according to the road environment on the SILAB driving simulator of Tongji University. Totally, 3000 groups of data were collected with 40 sample points of driving speed in a group.

Thirdly, bimodal GMM fitting was conducted in each group of data. An example of bimodal GMM fitting of speed distribution is shown in Figure 14.

As shown in Figure 14, the histogram represents the statistical quantity of the observed driving speed. The probability density curve in solid line represents the fitted bimodal GMM, while the two-dotted line represents the two component distributions in the GMM. The parameters of the fitted bimodal GMM are shown in Table 2.

3.2. Construction and Training of Deep Learning Models. Considering the stochastic characteristic of neural networks, optimized structure does not necessarily mean better performance. Tests are needed to verify the applicability of the three mentioned models in the prediction of driving speed. The three models were constructed with the same structure in Python, and the performance was then compared with the same set of training dataset.

For the basic structure of the model, full-connected model was selected to improve potential performance of the model while avoiding limitations of artificial intervention.

For the input variables of the models, totally 7 variables are selected to form the input vector, including road alignment features and 3D ASD, as shown in Table 3.

For the output variables of the model, considering that driving speed was described in term of bimodal GMM, the output variables are set to be the parameters in the GMM, as shown in Table 4.

Since $\lambda_2 = 1 - \lambda_1$, there is no need for λ_2 to take an additional spot in the output vector.

For the selection of numbers of nodes in the hidden layer, there has not been universal method in determining the certain number of nodes in NN. Generally, the number of samples in the training dataset should be 2–10 times more than the number of connection weights in the hidden layers. Considering the number of nodes in the input layer and output layers, it is appropriate to select the number of nodes in the hidden layer to be 25.

For the selection of numbers of hidden layers, it has been proved that more hidden layers could improve the accuracy of the model, but the generalization ability would decrease at the same time. Generally, the profit from increasing number of nodes is larger than the profit from increasing number of hidden layers. Therefore, the number of hidden layers is set to 1 in this study.

For the selection of loss function, mean relative error (MRE), mean square error (MSE), and mean absolute error (MAE) are the representative ones to measure the accuracy of NN models. However, the aim of the models applied in this study is to predict an accurate distribution of speed. Prediction accuracy of any isolated parameters in the GMM will be less significant to reveal the performance of the model. It will be more appropriate to measure the precision of the predicted distribution by measuring its divergence with the actual distribution. Therefore, Kullback–Leibler divergence (KL divergence) was selected to be the loss function considering its physical significance in distribution comparison. The loss function is shown in formulas (23) and (24). The value of the function represents the difference between distribution $P(x)$ and $Q(x)$.

$$KL(P \parallel Q) = \int P(x) \log \frac{P(x)}{Q(x)} = E_{P(x)} \log \frac{P(x)}{Q(x)}, \quad (23)$$

$$\begin{cases} KL(P \parallel Q) = 0, & P(x) = Q(x), \\ KL(P \parallel Q) > 0, & P(x) \neq Q(x). \end{cases} \quad (24)$$

Since $KL(P \parallel Q) \neq KL(Q \parallel P)$, $P(x)$ was set to be the actual speed distribution and $Q(x)$ was the predicted speed distribution.

For the selection of length of input sequence, freight vehicle drivers' perception and memorization should be taken into consideration. A short length of input sequence may not be able to cover all the necessary information, while longer length of input sequence may cause redundancy. It has been proved that short time memory lasts for no longer than 20 s. Assuming the driving speed to be 90 km/h on mountainous freeway, drivers will remember the information of former 500 m road condition. In order to find out an appropriate length of the input sequence, five different lengths of input sequence including 100 m, 200 m, 300 m, 400 m, and 500 m will be tested, respectively.

The bias vectors and weight matrixes are initiated randomly between $(-1, 1)$, and the optimization of the weight matrixes is completed by backpropagation through time (BPTT).

According to the convention of the training of neural networks, the proportions of the training set could be 60–90% considering the scale of the dataset [53]. Considering that there are totally 3000 groups of data in the dataset, 2700 groups of data were used for the training and establishment of the model, and the remaining 300 groups were used to test the performance and validate the model.

4. Results

The comparison of the three models with different lengths of input sequence is shown in Table 5; the values in the table refer to the average KL divergence between the predicted

TABLE 1: Information of participants.

Gender	Age	Driving experience
Male: 33	Mean value: 34.4	Mean value: 7.3
Female: 7	SD: 2.9	SD: 2.2
	Range: 31–47	Range: 4–11

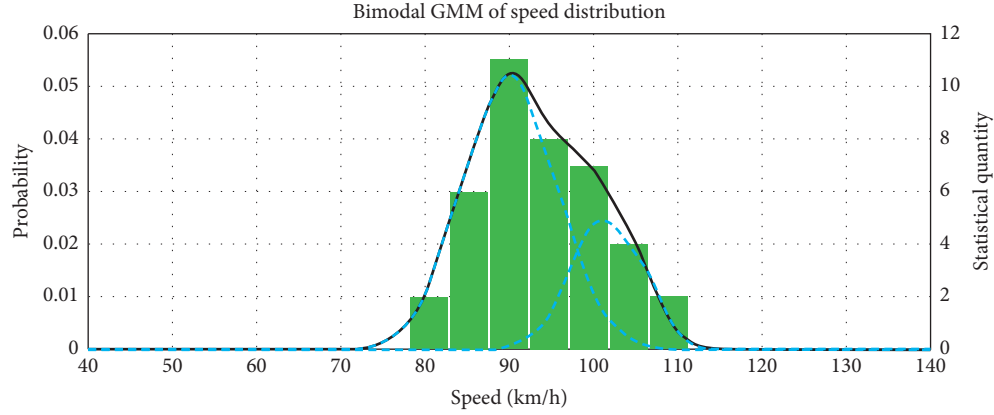


FIGURE 14: Example of bimodal GMM fitting result.

TABLE 2: Parameters in the bimodal GMM fitting example.

Distribution	λ	μ	σ
1	0.73	89.97	5.60
2	0.27	101.62	4.20

TABLE 3: Classification and explanation of input variables.

Variables	Explanation
K_h	Horizontal curvature
S_h	Horizontal curve length
K_v	Vertical curvature
S_v	Vertical curve length
K_c	Cross slope
S_c	Cross-section width
ASD	3D available sight distance

TABLE 4: Classification and explanation of output variables.

Variables	Explanation
λ_1	Weight of the 1 st Gauss distribution
μ_1	Mean value of the 1 st Gauss distribution
μ_2	Mean value of the 2 nd Gauss distribution
σ_1	Standard deviation of the 1 st Gauss distribution
σ_2	Standard deviation of the 2 nd Gauss distribution

TABLE 5: KL divergence of the deep learning models with different lengths of input sequence.

Type of model	Length of the input sequence				
	100 m	200 m	300 m	400 m	500 m
RNN	0.44	0.35	0.32	0.32	0.32
LSTM	0.25	0.18	0.12	0.11	0.11
GRU	0.34	0.27	0.25	0.25	0.25

distribution and the actual distribution. An example of the predicted distribution of LSTM with the length of input sequence being 400 m and the actual distribution is shown in Figure 15. The parameters of the actual distribution and the predicted distribution are shown in Table 6.

It can be found that either type of network shows its applicability in fitting the speed prediction model with the minimum value of the loss function being no more than 0.32. The best performance belongs to LSTM with the input sequence of 400 m with its KL divergence reaching 0.11. It is also worth noticing that LSTM and GRU always outperformed RNN in accuracy with the same length of input sequence. It can be inferred that the additional gates in RNN could partly solve the problem of time dependency. Among those, LSTM holds the best performance all the time. The simplified structure of GRU resulted in its poorer performance.

Besides, suppose that the driving speed to be 90 km/h, a length of 400 m means a memorization span of 16 s, which means that drivers are able to remember the information of road geometry for about 16 s on mountainous freeway. The result is highly correlated with the terrain condition on mountainous freeways. The character of road geometry is likely to change in 400 m in mountainous freeway, which is a relatively large burden for the freight vehicle drivers to take. So, it can be explained that drivers are likely to get used to the oncoming road condition and forget the previous information if there is a significant difference between the adjacent sections. The length of input sequence and the time span may change in different road conditions.

In order to validate the proposed LSTM model, the study applied multiple linear regression to make a comparative study. Since there are five parameters in the bimodal GMM,

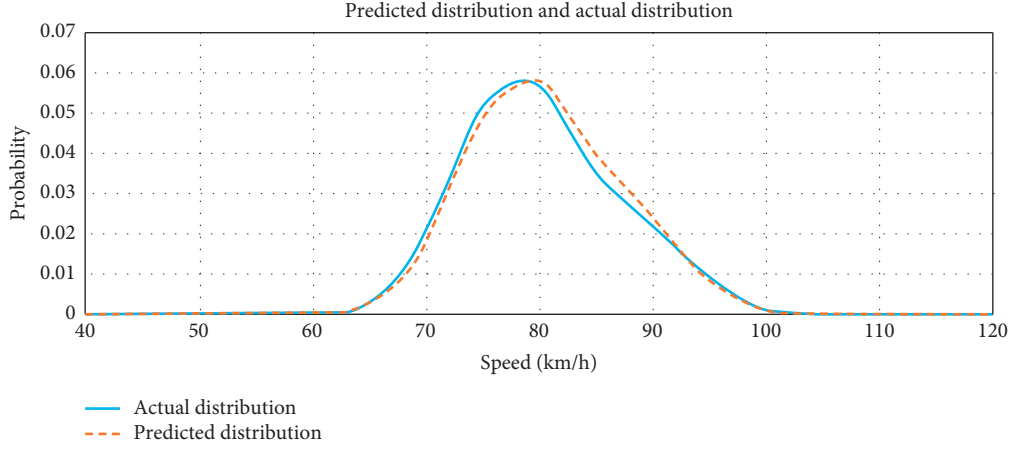


FIGURE 15: An example of the predicted distribution and actual distribution.

TABLE 6: Parameters in the predicted distribution and the actual distribution.

Variable	Parameters in the bimodal GMM					
	λ_1	μ_1	σ_1	λ_2	μ_2	σ_2
Actual	0.77	77.74	5.22	0.23	89.11	4.80
Predicted	0.75	78.03	5.19	0.25	88.54	4.69
KL divergence	0.12					

TABLE 7: Adjusted R^2 of the multiple linear regression models.

Predicted parameter	Adjusted R^2
λ_1	0.182
μ_1	0.272
μ_2	0.153
σ_1	0.208
σ_2	0.135

five multiple linear regression models were constructed to predict the parameters, respectively. The adjusted R^2 values of the models are shown in Table 7. It can be easily concluded that multiple linear regression is weaker in model accuracy and interpretability than the LSTM model since the adjusted R^2 values of the models are all smaller than 0.3, which means that it is difficult for the model to predict the parameters with precision. Meanwhile, the structure of the model is inferior than LSTM in interpreting the driving process since the isolated input variables fail to reflect the sequence characteristics of the driving process.

However, the multiple linear regression could help to understand the mechanism of the road environment affecting vehicle speed. Taking the multiple linear regression model predicting μ_1 as an example, as shown in Table 8, under the confidence level of 0.05, horizontal curvature, horizontal curve length, and 3D available sight distance have significant influence on one of the mean values of the GMM. The coefficient of the three variables can be interpreted as follows: the increase of horizontal curvature, horizontal curve length, and 3D available sight distance will lead to higher speed, while horizontal curvature has a most significant impact among the three variables.

TABLE 8: Coefficient table of the multiple linear regression model predicting μ_1 .

Variables	Standardized coefficients	Sig.
K_h	0.375	0.011
S_h	0.158	0.009
K_v	-0.242	0.088
S_v	0.272	0.059
K_c	-0.249	0.093
S_c	0.110	0.686
ASD	0.144	0.020

5. Conclusion

Targeting the weakness of traditional speed prediction models, this study applies several types of deep learning methods based on RNN to predict speed distribution on mountainous freeways. RNN, LSTM, and GRU were trained and tested, respectively, based on a driving simulation dataset. The results suggest that

- (1) It is reasonable to apply bimodal GMM to characterize the distribution of speed of freight vehicle, which improved the traditional method by using single normal distribution or lognormal distribution.
- (2) RNN, LSTM, and GRU all show their applicability in speed prediction of freight vehicles in consideration of road geometry. LSTM outperforms RNN and GRU on most tasks, and GRU shows its slight advantage over RNN on this specific issue.
- (3) The length of input sequence has a significant impact on training results of the models, which demonstrates the necessity of the application of RNN and its variants. Drivers of freight vehicles are proved to get affected by the consecutive road condition in their driving process. On mountainous freeway, drivers are likely to update their memorization of road condition with the threshold of 16 s, which is 400 in driving distance. The change of road geometry may be the main reason for the time span. The value of the

time span may change in different road conditions. This could also explain that mountainous freeway always has accident-prone road sections with heavy information load for freight vehicle drivers to take.

- (4) The accurate road geometry information provided by 3D mobile mapping may help to improve the reliability of driving simulation study and the results of the speed prediction model.
- (5) LSTM is obviously superior to regression models in terms of the model accuracy and interpretability of the driving process and the formation of vehicle speed. Further research on regression models could be conducted targeting the mechanism of road environment affecting vehicle speed.

The findings contribute to a better understand on freight vehicle drivers' driving behavior on mountainous freeways. The study also provided important insights into the road geometry design and development of transportation safety strategies.

In the future work, more elements in road environment and driving behavior could be included into the model. Effort could be put into freight vehicle drivers' memorizing mechanism of road environment, and analysis of the information load for drivers could also be made.

Data Availability

The 3D mobile mapping data, deep learning code in Python, and driving simulation data are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This research was sponsored by the Zhejiang Provincial Communication Department (2015J07).

References

- [1] Y. Yan, G. Li, J. Tang, and Z. Guo, "A novel approach for operating speed continuous predication based on alignment space comprehensive index," *Journal of Advanced Transportation*, vol. 2017, Article ID 9862949, 14 pages, 2017.
- [2] B. Yu, Y. Chen, S. Bao, and D. Xu, "Quantifying drivers' visual perception to analyze accident-prone locations on two-lane mountain highways," *Accident Analysis & Prevention*, vol. 119, pp. 122–130, 2018.
- [3] M. Giunta and F. G. Praticò, "Safety evaluation: a new operating speed model for two-lane, undivided rural roads," *Advances in Transportation Studies*, vol. 34, pp. 67–80, 2014.
- [4] C. D. Fitzpatrick, C. P. Harrington, and M. A. Knodler, "The effects of clear zone size and roadside vegetation on driver behavior," in *Proceedings of the Transportation Research Board 93rd Annual Meeting*, Washington, DC, USA, January 2014.
- [5] N. Skinner and J. D. Bullough, "Effects of roadway delineator spacing, size, and height on drivers' perception and behavior," in *Proceedings of the Transportation Research Board 88th Annual Meeting*, Washington, DC, USA, January 2009.
- [6] L. C. Trentin, F. I. Kabbach Jr., and C. Y. Suzuki, "Analysis of speed prediction for design consistency of two-lane rural highways," in *Proceedings of the 4th International Symposium on Highway Geometric Design*, Valencia, Spain, June 2010.
- [7] A. Montella and L. L. Imbriani, "Safety performance functions incorporating design consistency variables," *Accident Analysis & Prevention*, vol. 74, pp. 133–144, 2015.
- [8] R. Lamm, A. K. Guenther, and E. M. Choueiri, "Safety module for highway geometric design," in *Proceedings of the Transportation Research Board 99th Annual Meeting*, pp. 7–15, Washington, DC, USA, January 1995.
- [9] P. Colonna, R. A. Aquilino, N. Berloco et al., "Relationships among road geometry, drivers' risk perception, and speed choice: experimental study," in *Proceedings of the Transportation Research Board 92nd Annual Meeting*, Washington, DC, USA, January 2013.
- [10] S. Shi, X. Liu, K. Cao, and S. Ge, "Study on cloud model of operating speed based on road alignment and sight distance," in *Proceedings of the 2011 International Conference on Transportation, Mechanical, and Electrical Engineering (TMEE)*, pp. 648–651, Changchun, China, December 2012.
- [11] S. Cafiso and G. L. Cava, "Driving performance, alignment consistency, and road safety," *Transportation Research Record Journal of the Transportation Research Board*, vol. 2102, no. 1, pp. 1–8, 2009.
- [12] J. C. Mao, "Safety evaluation of mountain scenic highway based on operating speed of three-dimensional alignment," *Transport Research*, vol. 1, no. 3, 2015.
- [13] G. M. Gibreel, S. M. Easa, and I. A. El-Dimeery, "Prediction of operating speed on three-dimensional highway alignments," *Journal of Transportation Engineering*, vol. 127, no. 1, pp. 21–30, 2001.
- [14] A. Tizghadam and A. Leon-Garcia, "On robust traffic engineering in transport networks," in *Proceedings of the IEEE GLOBECOM 2008—2008 IEEE Global Telecommunications Conference*, pp. 1–6, New Orleans, LO, USA, December 2008.
- [15] T. Ben-Bassat and D. Shinar, "Effect of shoulder width, guardrail and roadway geometry on driver perception and behavior," *Accident Analysis & Prevention*, vol. 43, no. 6, pp. 2142–2152, 2011.
- [16] S. G. Charlton and N. J. Starkey, "Risk in our midst: centrelines, perceived risk, and speed choice," *Accident Analysis & Prevention*, vol. 95, pp. 192–201, 2016.
- [17] L. Zakowska, "The role of geometric road design parameters in driving speed negotiation and safety perception: objective and subjective measures in simulation study," *Advances in Transportation Studies*, vol. 18, no. 18, pp. 17–38, 2009.
- [18] B. Zirkel, S. Park, J. Mcfadden, M. Angelastro, and L. McCarthy, "Analysis of sight distance, crash rate, and operating speed relationships for low-volume single-lane roundabouts in the United States," *Journal of Transportation Engineering*, vol. 139, no. 6, pp. 565–573, 2013.
- [19] D. W. Harwood and K. M. Bauer, "Effect of stopping sight distance on crashes at crest vertical curves on rural two-lane highways," *Transportation Research Record Journal of the Transportation Research Board*, vol. 2486, no. 1, pp. 45–53, 2015.
- [20] A. M. P. Zuriaga, A. G. García, F. J. C. Torregrosa, and P. D'Attoma, "Modeling operating speed and deceleration on two-lane rural roads with global positioning system data," *Transportation Research Record: Journal of the Transportation Research Board*, vol. 2171, no. 1, pp. 11–20, 2010.

- [21] S. K. S. Abbas, M. A. Adnan, and I. R. Endut, "Exploration of 85th percentile operating speed model on horizontal and vertical alignments: a case study for two lane rural highways," in *Proceedings of the Road Transport Information and Control Conference and the ITS United Kingdom Members' Conference, IET*, pp. 352–363, London, UK, May 2010.
- [22] J. Lu, "Regression analysis on speed distribution characteristics of ordinary road," *Journal of Southeast University*, vol. 42, no. 2, pp. 374–377, 2012.
- [23] B.-J. Park, Y. Zhang, and D. Lord, "Bayesian mixture modeling approach to account for heterogeneity in speed data," *Transportation Research Part B: Methodological*, vol. 44, no. 5, pp. 662–673, 2010.
- [24] D. Branston, "A method of estimating the free speed distribution for a road," *Transportation Science*, vol. 13, no. 2, pp. 130–145, 1979.
- [25] C. Xu, "Distribution of vehicle free flow speeds based on Gaussian mixture model," *Journal of Highway and Transportation Research and Development*, vol. 29, no. 8, pp. 132–135, 2012.
- [26] D. B. Fambro, K. Fitzpatrick, and C. W. Russell, "Operating speed on crest vertical curves with limited stopping sight distance," *Transportation Research Record: Journal of the Transportation Research Board*, vol. 1701, no. 1, pp. 25–31, 2007.
- [27] K. Fitzpatrick, L. Elefteriadou, D. W. Harwood et al., "Speed prediction for two-lane rural highways," Report No. FHWA-RD-99-171, Federal Highway Administration, Washington, DC, USA, 2000.
- [28] W. Bo, S. Hallmark, P. Savolainen, and J. Dong, "Examining vehicle operating speeds on rural two-lane curves using naturalistic driving data," *Accident Analysis & Prevention*, vol. 118, pp. 236–243, 2018.
- [29] A. Ghasemzadeh and M. M. Ahmed, "Utilizing naturalistic driving data for in-depth analysis of driver lane-keeping behavior in rain: non-parametric MARS and parametric logistic regression modeling approaches," *Transportation Research Part C: Emerging Technologies*, vol. 90, pp. 379–392, 2018.
- [30] B. Yu, Y. Chen, and S. Bao, "Quantifying visual road environment to establish a speeding prediction model: an examination using naturalistic driving data," *Accident Analysis & Prevention*, vol. 129, pp. 289–298, 2019.
- [31] A. Polus and C. Mattar-Habib, "New consistency model for rural highways and its relationship to safety," *Journal of Transportation Engineering*, vol. 130, no. 3, pp. 286–293, 2004.
- [32] A. García, D. Llopis-Castelló, F. Camacho-Torregrosa, and A. Pérez-Zuriaga, "New consistency index based on inertial operating speed," *Transportation Research Record: Journal of the Transportation Research Board*, vol. 2391, no. 1, pp. 105–112, 2013.
- [33] D. Llopis-Castelló, F. Bella, F. J. Camacho-Torregrosa, and A. García, "New consistency model based on inertial operating speed profiles for road safety evaluation," *Journal of Transportation Engineering, Part A: Systems*, vol. 144, no. 4, Article ID 04018006, 2018.
- [34] J. S. Wijnands, J. Thompson, G. D. P. A. Aschwanden, and M. Stevenson, "Identifying behavioural change among drivers using long short-term memory recurrent neural networks," *Transportation Research Part F: Traffic Psychology and Behaviour*, vol. 53, pp. 34–49, 2018.
- [35] A. Graves, A. R. Mohamed, and G. Hinton, "Speech recognition with deep recurrent neural networks," in *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing*, pp. 6645–6649, Vancouver, Canada, May 2013.
- [36] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Computation*, vol. 9, no. 8, pp. 1735–1780, 1997.
- [37] Q. Wang, Y. Gu, J. Liu, and S. Kamijo, "DeepSpeedometer: vehicle speed estimation from accelerometer and gyroscope using LSTM model," in *Proceedings of the 2017 IEEE 20th International Conference on Intelligent Transportation Systems (ITSC)*, pp. 1–6, Yokohama, Japan, October 2018.
- [38] F. Altché and A. De La Fortelle, "An LSTM network for highway trajectory prediction," in *Proceedings of the IEEE 20th International Conference on Intelligent Transportation*, Yokohama, Japan, October 2017.
- [39] Z. Zhao, W. Chen, X. Wu, P. C. Y. Chen, and J. Liu, "LSTM network: a deep learning approach for short-term traffic forecast," *IET Intelligent Transport Systems*, vol. 11, no. 2, pp. 68–75, 2017.
- [40] D. Kang, Y. Lv, and Y. Y. Chen, "Short-term traffic flow prediction with LSTM recurrent neural network," in *Proceedings of the 2017 IEEE 20th International Conference on Intelligent Transportation Systems (ITSC)*, pp. 1–6, Yokohama, Japan, October 2017.
- [41] H. Zhu, X. Yang, and Y. Wang, "Prediction of daily entrance and exit passenger flow of rail transit stations by deep learning method," *Journal of Advanced Transportation*, vol. 2018, Article ID 6142724, 11 pages, 2018.
- [42] Z. Duan, Y. Yang, K. Zhang, Y. Ni, and S. Bajgain, "Improved deep hybrid networks for urban traffic flow prediction using trajectory data," *IEEE Access*, vol. 6, pp. 31820–31827, 2018.
- [43] K. Cho, B. Van Merriënboer, C. Gulcehre et al., "Learning phrase representations using RNN encoder-decoder for statistical machine translation," in *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, Doha, Qatar, October 2014.
- [44] S. Cascianelli, G. Costante, T. A. Ciarfuglia, P. Valigi, and M. L. Fravolini, "Full-GRU natural language video description for service robotics applications," *IEEE Robotics and Automation Letters*, vol. 3, no. 2, pp. 841–848, 2018.
- [45] A. H. Mirza, "Online additive updates with FFT-IFFT operator on the GRU neural networks," in *Proceedings of the 2018 26th Signal Processing and Communications Applications Conference (SIU)*, pp. 1–4, Izmir, Turkey, May 2018.
- [46] R. Fu, Z. Zhang, and L. Li, "Using LSTM and GRU neural network methods for traffic flow prediction," in *Proceedings of the 2016 31st Youth Academic Annual Conference of Chinese Association of Automation (YAC)*, pp. 324–328, Wuhan, China, November 2017.
- [47] E. Demir, G. T. Bektaş, and G. Laporte, "A review of recent research on green road freight transportation," *European Journal of Operational Research*, vol. 237, no. 3, pp. 775–793, 2014.
- [48] X. Li and R. Bai, "Freight vehicle travel time prediction using gradient boosting regression tree," in *Proceedings of the 2016 15th IEEE International Conference on Machine Learning and Applications (ICMLA)*, IEEE Computer Society, Anaheim, CA, USA, December 2016.
- [49] D. Tian, X. Shan, Z. Sheng, Y. Wang, W. Tang, and J. Wang, "Break-taking behaviour pattern of long-distance freight vehicles based on GPS trajectory data," *IET Intelligent Transport Systems*, vol. 11, no. 6, pp. 340–348, 2017.
- [50] X. Li and R. Bai, "Freight vehicle travel time prediction using sparse Gaussian processes regression with trajectory data," in *Intelligent Data Engineering and Automated*

- Learning—IDEAL 2016*, Springer International Publishing, Berlin, Germany, 2016.
- [51] J. Mendes-Moreira, A. M. Jorge, J. F. de Sousa, and C. Soares, "Comparing state-of-the-art regression methods for long term travel time prediction," *Intelligent Data Analysis*, vol. 16, no. 3, pp. 427–449, 2012.
 - [52] C. M. Monsere, M. Wolfe, H. Alawakiel et al., *Developing Corridor Level Truck Travel Time Estimates and Other Freight Performance Measures from Archived ITS Data*, Oregon Department of Transportation, Oregon Transportation Research and Education Consortium, Portland, OR, USA, 2009.
 - [53] M. T. Hagan, H. B. Demuth, and M. H. Beale, *Neural Network Design*, 2002.

