IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received June 14, 2019, accepted June 28, 2019, date of publication July 3, 2019, date of current version July 24, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2926494

Driving Style Classification Based on Driving Operational Pictures

GUOFA LI [1,2], FANGPING ZHU [1], XINGDA QU [1], BO CHENG [2], SHEN LI [3], AND PAUL GREEN [4,5]

[1] College of Mechatronics and Control Engineering, Institute of Human Factors and Ergonomics, Shenzhen University, Shenzhen 518060, China
[2] State Key Laboratory of Automotive Safety and Energy, School of Vehicle and Mobility, Tsinghua University, Beijing 100084, China
[3] Department of Civil and Environmental Engineering, University of Wisconsin–Madison, Madison, WI 53706, USA
[4] University of Michigan Transportation Research Institute (UMTRI), University of Michigan, Ann Arbor, MI 48109, USA
[5] Department of Industrial and Operations Engineering, University of Michigan, Ann Arbor, MI 48109, USA

Corresponding author: Xingda Qu (quxd@szu.edu.cn)

This work was supported in part by the National Natural Science Foundation of China under Grant 51805332 and Grant 31570944, in part by the Natural Science Foundation of Guangdong Province under Grant 2018A030310532, in part by the Young Elite Scientists Sponsorship Program by China Society of Automotive Engineers, and in part by the State Key Laboratory of Automotive Safety and Energy under Project KF1801.

ABSTRACT Accurately describing and classifying driving style is crucial for driving safety intervention strategies in the design of advanced driver assistance systems (ADASs). This paper presents a novel driving style classification method based on constructed driving operational pictures (DOPs) which map sequential data from naturalistic driving into 2-D pictures. By using the nested time window method, 798/1683/1153 DOPs sized 42 (features) $\times$ 60 (seconds) were generated for three different driving styles (low-risk, moderate-risk, and high-risk), respectively. The three kinds of neural network algorithms, i.e., convolutional neural network (CNN), long short-term memory (LSTM) network, and pretrain-LSTM were applied to recognize driving styles based on DOPs. The results showed that CNN performed the best with an accuracy of 98.5%, better than the traditional support vector machine (SVM) method. This study provides a new perspective to classify driving style which may help design ADASs operating characteristics to improve driving comfort and safety.

INDEX TERMS Driving style, driving comfort and safety, driving operational pictures, neural network, naturalistic driving.

I. INTRODUCTION

Driving style is defined as a set of individual driving habits formed gradually with the accumulation of driving experience [1]. It significantly influences driving safety [2] and fuel economy [3]. Knowledge about driving style can play an important role in the design of advanced driver assistance systems (ADASs) [5], [6]. However, even the same driver may exhibit different driving styles in different scenarios or the same scenario at different times [7]. Therefore, characterizing and determining drivers' driving style is particularly challenging.

Driving style can be determined subjectively or objectively. Among the subjective approaches, questionnaires/surveys and expert assessment are the most widely accepted [8], [9]. The multidimensional driving style inventory (MDSI) is a 44-item questionnaire for driving style evaluation [10], of which the questions are categorized according to the four pre-defined driving styles, i.e., drivers' reckless and careless styles (11 items), anxious styles (19 items), angry and hostile styles (5 items), and patient and careful styles (9 items). MDSI was first applied in Israel and has been widely employed in driving style classification research elsewhere [11]–[13]. To improve the objectivity and directness of the questionnaire, Hong et al. added the record of driving violations as an evaluation criterion [14].

To avoid the reliability problems from drivers' self-reported answers, expert evaluation is an alternative subjective method. Unfortunately, subjective consciousness still exists and there is no uniform standard to guide consistent expert judgments [15].

Although subjective evaluation can be effective for classifying driving style, the method is excessively labor intensive and requires experts to always be in the vehicle. This is often not possible, especially if the intended sample size is large or the focus is on higher level automation, where an

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

expert evaluator often is not normally in the vehicle. Therefore, finding alternative (objective) ways to estimate driving style is desired.

Most of the reported objective efforts on driving style classification were based on driver operation and vehicle movement signals. Han *et al.* [7] evaluated driving styles based on longitudinal speed and throttle opening. Xue *et al.* [16] extracted features from acceleration, relative speed, and relative distance from vehicle trajectory data for driving style recognition. Suzdaleva and Nagy [1] found that fuel consumption, vehicle speed, throttle position, and gear position contributed to the recognition of drivers' driving style. Eboli *et al.* [14] used vehicle speed and accelerations to determine drivers' driving style by counting the number of data out of the pre-defined safety domain. Besides these operational-level signals, maneuver-level features also can be used for driving style classification. Li *et al.* [17] found that driving maneuver transitions could be used to classify driving style with a higher accuracy than when using the event numbers of those maneuvers. Bejani and Ghatee [18] divided driving profile into four maneuvers and three time intervals and used that information to estimate the risk level of driving behavior.

To classify driving style based on the above-mentioned features, machine learning techniques show great potential. Support vector machine (SVM) is a simple but effective algorithm, which has been widely used and improved in driving style classification tasks. Woo *et al.* [19] classified driving style using SVM based on the extracted features from vehicle movement variables, and found an average of 71.0% for the evaluated F_1-score. Wang *et al.* [20] employed a semi-supervised SVM to model driving style into aggressive and normal groups with an accuracy of 86.6%. Random forest algorithm, one of the top two conventional classifiers demonstrated by Fernández-Delgado *et al.* [21], is also reported in [17] to successfully classify driving styles into three groups with an accuracy of 93.0%. Also, algorithms based on Bayes theorem have also been used in [7] to successfully classify driving styles with a reported accuracy of 93.5%. Besides these supervised learning algorithms, a typical unsupervised learning method k-means was applied in [1] to label the ground-truth of driving style into seven groups without clearly describing the meaning of each group.

Within the last three years, newly developed neural network methods such as the convolutional neural network (CNN) and recurrent neural network (RNN) have been widely used in classification tasks with satisfactory performance [22]–[24]. However, none of the previous studies have employed these advanced methods on driving style classification. If these advanced methods could be successfully applied, it might significantly improve the adaptive design of ADASs for drivers with different driving preferences [25], [26]. Many of the previous studies assume that the driving style of a driver is stable. However, as was noted earlier, driving style changes with the driving situations [5]. High-risk drivers may drive cautiously when not being irritated, and low-risk drivers could drive aggressively under high time pressure to reach a destination. Therefore, person-based driving style classification may not be applicable for real-time or quasi-real-time applications in vehicles.

To advance the state-of-the-art and overcome concerns with previous methods, this paper innovatively proposes a nested time window method to construct drivers' quasi-real-time driving operational pictures (DOPs), based on which advanced machine learning techniques are used to classify driving style. The developed DOPs describe driving style from multiple aspects in the operational level (steering wheel angle, vehicle speed, acceleration, etc.) by combining various statistical functions. Advanced methods including CNN, LSTM (long short-term memory), and pretrain-LSTM were adopted to classify naturalistic driving data collected on highways. The contributions of this presented approach include:

(1) The idea of using DOPs to classify driving style is new. The DOPs make it intuitive to see the driving style patterns, and the collection of features in a DOP will be easier for understanding and evaluation than when the features are presented as tables of numerical data.

(2) We utilized CNN and LSTM to effectively classify driving style based on the constructed DOPs. Although these methodologies are not novel in the field of computer science, this is the first application of them to driving style classification.

(3) The proposed approach is applicable in the classification of dynamically changing driving style. A timely recognition of drivers' driving style would provide supports on adaptive strategies for driver assistance or intelligent driving applications.

The remainder of this paper is organized as follows: Section II clearly described the details about the naturalistic driving experiment, the ground-truth labeling of driving style, and the collected data. How the DOPs are constructed based on the collected data is introduced in section III. The employed advanced algorithms are briefly introduced in section IV. Section V lists the evaluation criterion of the examined algorithms. Section VI presents and discusses the results.

II. NATURALISTIC DRIVING EXPERIMENT AND DATA COLLECTION

To collect naturalistic driving data for driving style classification, 28 participants (18 males and 10 females) were recruited to drive on the highway from Beijing to Xianghe, a town in Hebei province in China. The age of the drivers ranged from 27 to 59 years old with a mean of 42. They had a mean driving experience of 13.0 years, ranging from 2 to 33. The round-trip distance was about 146 km with a posted speed limit of 120 km/h. Six cameras were mounted on the test vehicle to record drivers' face images, operation of the pedals, following time headway, front road scenes, and left and right side scenes. See Figure 1. Six basic features (BFs) from the CAN-bus were collected at 10 Hz, including throttle pedal position (Thro), vehicle speed (Speed), brake pedal position (Brake), steering wheel angle (SWA), lateral acceleration (LAcc), and yaw-rate (Yaw). Figure 2 presents a brief overview of the collected BFs.

FIGURE 1. Collected camera images.

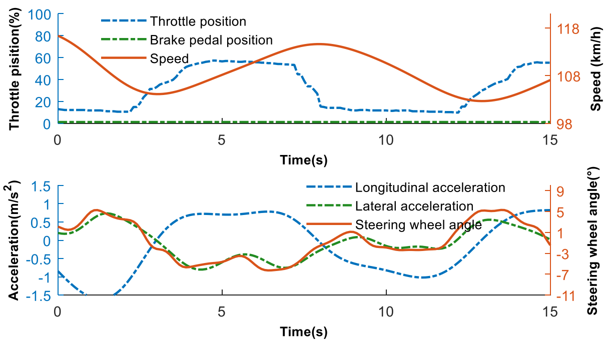


FIGURE 2. A brief overview of the collected BFs.

To provide a standard for comparison, three licensed driver, experienced in rating driving behavior, rated the driving style label of each subject using a three-point scale (1: low-risk, 2: moderate-risk, 3: high-risk). They all had been involved in driving behavior analysis and related projects for more than three years. Ratings were discussed and re-rated as necessary to obtain a consistent estimate of the probability the driver was going to be involved in a crash. When there was a conflict between their subjective evaluations, the majority rating score would be adopted as the final style label. If all the three ratings were different, the corresponding data sample needed to be re-checked and re-rated.

III. DRIVING OPERATIONAL PICTURE (DOP) CONSTRUCTION

A nested time window method was developed to map the collected operational level signals in a picture for driving style classification. The DOP construction method was described in detail in Figure 3. The nested time window included a big time window (T_B) and a small time window (T_S) sliding in T_B along with time. The big time window describes drivers' operational behavior in a relatively long period, while the small time window reveals drivers' transient operational behavior. The selection of different time window length has no influence on the following networks for driving style classification, but will correspond to different classification performance because the DOP quality will be affected by the time window selection. In this study, the length of T_B and T_S was set as 60 seconds and 2 seconds respectively according to our experience. The time step of T_S was set as half of T_S, which means that the overlap between each two adjacent T_S was 1 second. Similarly, the time step of T_B was set as half of T_B, i.e., 30 seconds.

The T_S nested in T_B was proposed to describe the driving information in each T_B. The information was described from multiple aspects by seven statistical functions (SFs) including mean, minimum, maximum, median, 25% percentile, 75% percentile, and standard deviation. Therefore, a column vector sized 42×1 (6 BFs $\times$ 7 SFs) was generated for each T_S. Therefore, in each T_B, there would be 60 ($2 \times T_B/T_S$) column vectors working together to map a complete DOP, which means that the size of the DOP was 42×60. The pixel of a DOP represented a specific SF of a certain BF during a T_S. All the features were normalized to eliminate the effect of different feature dimensions in DOPs.

IV. NEURAL NETWORKS FOR DRIVING STYLE CLASSIFICATION

In this paper, three advanced neural network algorithms including convolutional neural network (CNN), long short-term memory (LSTM) network, and pretrain-LSTM were applied for driving style classification.

A. CONVOLUTIONAL NEURAL NETWORK (CNN)

CNN has been widely used in classification tasks based on images [27]. Given the generated DOPs as inputs, CNN can be trained for driving style classification. The developed CNN architecture used in this study is shown in Figure 4. The architecture includes two convolution-pooling layers and a fully-connected layer. The size of the convolution kernel was 42×5 while the size of the max-pooling kernel was 1×2 in the first convolution-pooling layer. The corresponding sizes of the kernels were 1×3, 1×2 in the second convolution-pooling layer. The following fully-connected layer gave three outputs, which represented the three driving styles to be categorized into. Based on the three outputs, a softmax function [28] was used to calculate the classification probabilities into different driving styles. All the activation function used in this network was rectified linear units (ReLU).

B. LONG SHORT-TERM MEMORY (LSTM) NETWORK

Different from CNN, Recurrent Neural Network (RNN) was designed to deal with time series problems. Long Short-Term Memory (LSTM) [29] is one of the most advanced algorithms developed based on RNN. As driving data is a typical time sequence, LSTM was employed for driving style classification in this study. An important improvement of LSTM is displacing the repeating module of standard RNN with four interacting layers, which are called "gates". The four layers

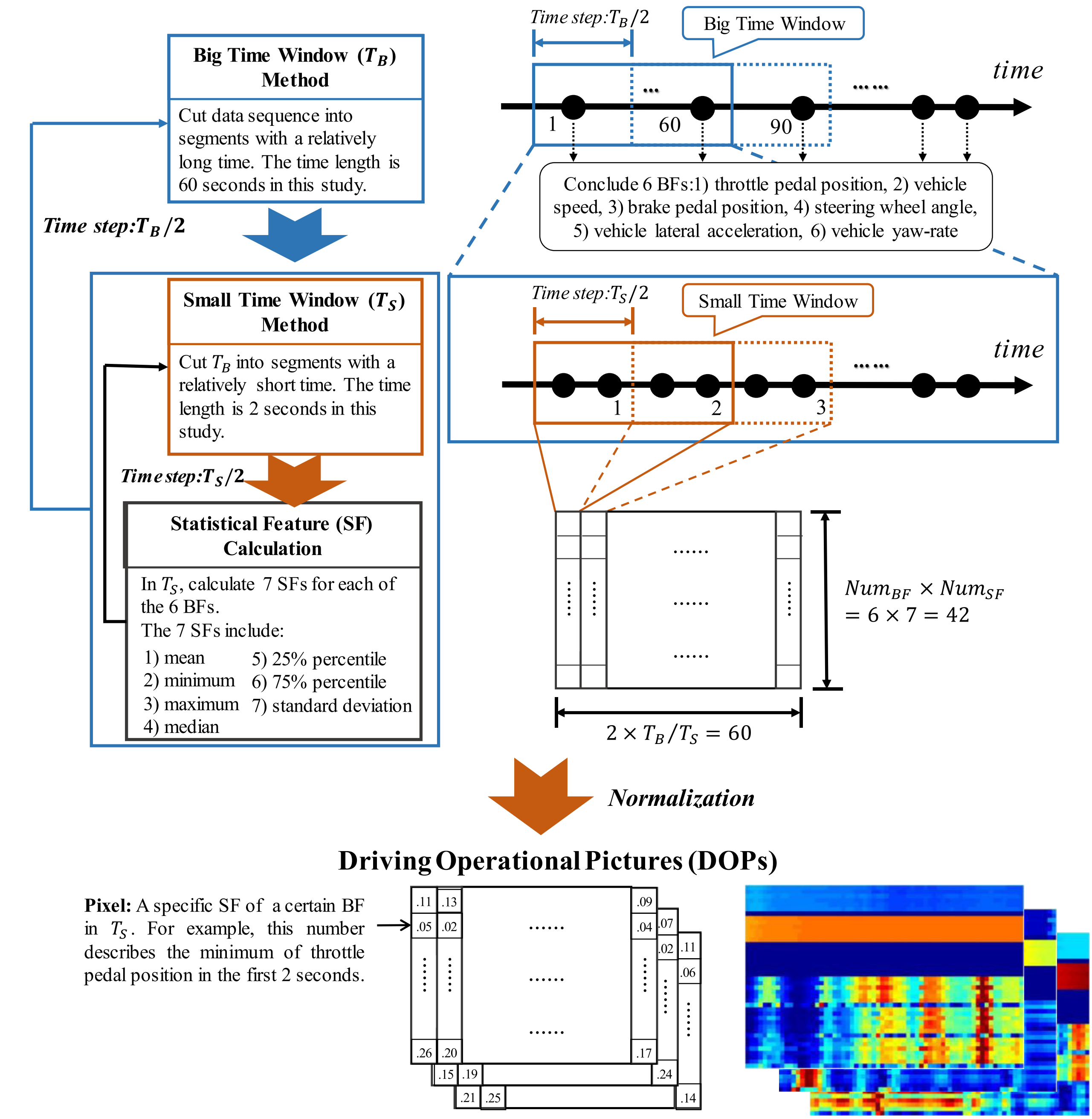


FIGURE 3. The nested time window method for driving operational picture construction.

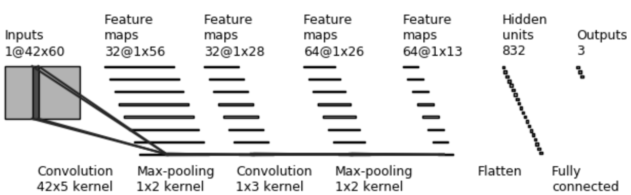


FIGURE 4. CNN architecture used for driving style classification.

are illustrated in Figure 5. Relying on the self-parameterized controlling gates, the memory cell C_t is accessed, written and cleared. It essentially acts as an accumulator of the state information. Benefiting from the property of sigmod neural net layer (σ), the input information would be checked to decide how much of them should be let through. See the working functions in Equation (1). When the input gate layer i_t is activated, the information of the new input will be stored in the cell state. Meanwhile, the past cell state C_{t-1} could be "forgotten" in proportion by the opened forget gate layer. Therefore, the cell state C_t is updated by "forgetting" useless information and "remembering" new information. Finally, the output is made up by the cell state and the output of o_t. Using the memory and gates to control information flow prevents the gradient from vanishing too quickly.

$$\begin{aligned} f_t &= \sigma(W_f \cdot [h_{t-1}, x_t] + b_f) \\ i_t &= \sigma\,(W_i \cdot [h_{t-1}, x_t] + b_i) \\ \tilde{C}_t &= \tanh(W_C \cdot [h_{t-1}, x_t] + b_C) \\ C_t &= f_t \times C_{t-1} + i_t \times \tilde{C}_t \\ o_t &= \sigma\,(W_o \cdot [h_{t-1}, x_t] + b_o) \\ h_t &= o_t \times \tanh(C_t) \end{aligned} \tag{1}$$

where b_f, b_i, b_C, b_o are the biases, W_f, W_i, W_C, W_o are the weight matrixes, x_t, h_t are the inputted DOP and outputted driving style.

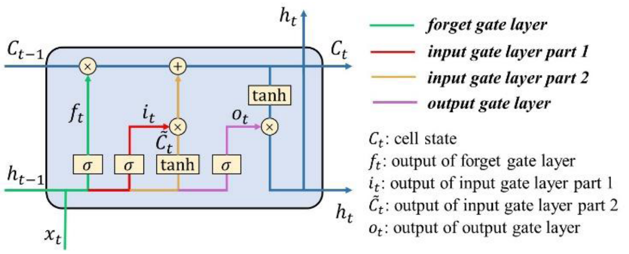


FIGURE 5. The repeating module of LSTM.

C. PRETRAIN-LSTM

CNN creates new features through convolution operation. These new features may work well on classification tasks [30]. Thus, taking the outputs from the convolution-pooling layer as the inputs of LSTM is feasible in practical. In this paper, the convolution-pooling layers of pretrain-LSTM were the same as the CNN structure built above. The output of the second convolution-pooling layer was re-shaped before feeding into LSTM. Figure 6 illustrates how the pretrain-LSTM works.

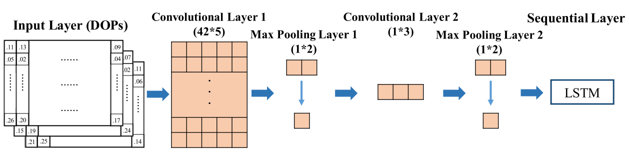


FIGURE 6. Working principle of pretrain-LSTM.

V. EVALUATION CRITERION

Table 1 presents the descriptions of the elements in the confusion matrix on the classification of low-density samples. The following terms were used to evaluate the performance of the used methods:

1) True Positives (TP): The number of low-risk samples that were classified into the correct driving style group (i.e., the low-risk group).

2) True Negatives (TN): The number of moderate- and high-risk samples that were classified into the moderate- or the high-risk group. TN = TN1 + TN2 + TN3 + TN4.

3) False Positives (FP): The number of moderate- and high-risk samples that were classified into the low-risk group. FP = FP 1 + FP2.

4) False Negatives (FN): The number of low-risk samples that were classified into the moderate- or high-risk group. FN = FN1 + FN2.

TABLE 1. Description on the elements in the confusion matrix on the classification of low-risk samples.

		Predicted density class		
		Low	Medium	High
True	Low	TP	FN1	FN2
density	Medium	FP1	TN1	TN2
class	High	FP2	TN3	TN4

To evaluate the performance of the employed advanced methods based on DOPs, prevision, recall, and F_1-score were used. These indexes are commonly adopted and well accepted in classification tasks. See Equation (2) for their calculations.

$$\begin{aligned} \text{Accuracy} &= \frac{TP+TN}{TP+TN+FP+FN} \\ \text{Precision} &= \frac{TP}{TP+FP} \\ \text{Recall} &= \frac{TP}{TP+FN} \\ F_\beta - score &= (1+\beta^2)\cdot\frac{Precision \cdot Recall}{\left(\beta^2 \cdot Precision\right) + Recall} \end{aligned} \tag{2}$$

where $\beta = 1$ is used in this study.

Based on the true positive rate ($TPR = TP\big/(TP+FN)$) and false positive rate ($FPR = FP\big/(FP+TN)$) to describe the probability of correct classification and the probability of false alarm respectively, receiver operating characteristic (ROC) curve was used to illustrate the relationship between TPR and FPR. The ROC curve is also a commonly used method to evaluate classifier performance. The more a curve in the ROC space bends to the up-left corner, the better the classification performance of the classifier is. To quantitatively describe the ROC curve, the machine learning community usually employs the area under the curve (AUC) statistic for model comparison. A higher AUC value indicates a closer bending to the up-left corner of the ROC curve, which proves better performance of the designed classifier.

VI. RESULTS AND DISCUSSION

In total, 3634 DOPs were finally obtained from the 28 drivers' naturalistic driving data on highways. Based on the subjective evaluation on driving style, 798/1683/1153 DOPs were categorized into the risk groups (low, moderate, high). To verify the effectiveness of the three neural network methods, 70% of all the DOPs were randomly selected as the training set and the remaining 30% were used as the test set. All these methods were trained and tested on the same training and test set to make fair comparisons.

TABLE 2. Results on the training dataset.

Methods	Accuracy (%)	Precision (%) by group			Recall (%) by group			F_1-score (%) by group		
		Low	Moderate	High	Low	Moderate	High	Low	Moderate	High
CNN	98.9	96.7	100	100	100	98.4	100	98.3	99.2	100
LSTM	98.3	94.9	100	100	100	97.4	100	97.4	98.7	100
pretrain-LSTM	98.9	96.7	100	100	100	100	97.7	98.3	100	98.8
SVM	93.4	94.6	91.0	96.5	83.5	95.5	97.1	88.7	93.2	96.8

TABLE 3. Results on the test dataset.

Methods	Accuracy (%)	Precision (%) by group			Recall (%) by group			F_1-score (%) by group		
		Low	Moderate	High	Low	Moderate	High	Low	Moderate	High
CNN	98.5	100	96.9	100	93.6	100	100	96.7	98.4	100
LSTM	95.7	100	91.4	100	81.7	100	100	89.9	95.5	100
pretrain-LSTM	47.5	4.0	48.7	100	18.7	100	9.5	6.5	65.5	17.4
SVM	92.2	94.0	88.6	93.5	81.7	94.9	96.2	87.4	91.7	94.9

To compare the performance of employed neural networks with other popular classification methods, SVM (support vector machine) was adopted. Since a DOP contained 2520 (42×60) features but the number of training samples was limited, it would cause overfitting problems when training with SVM. Therefore, we used a feature selection method based on conditional likelihood maximization to select the top 10 features as inputs of the SVM classifier. Details of the feature selection method can be found in [31].

A. CLASSIFICATION PERFORMANCE USING DIFFERENT ALGORITHMS

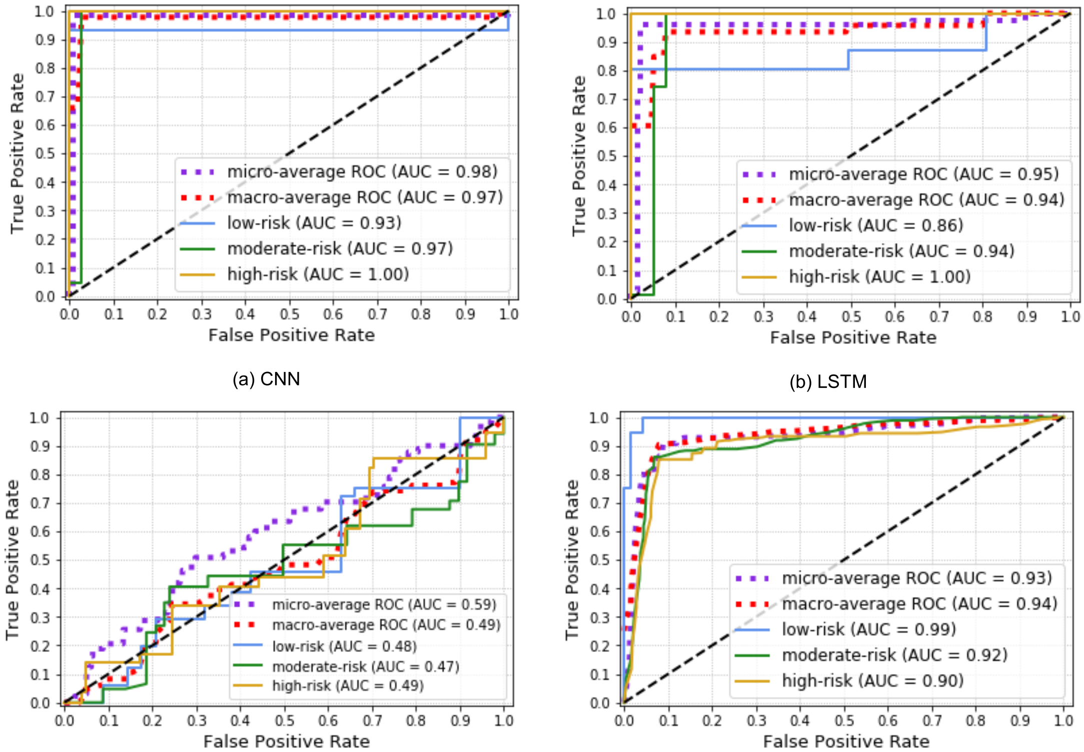
Using DOPs as inputs of the three adopted methods and SVM on the training set, the corresponding accuracies were calculated. See Table 2 for the classification results on the training set. All the classifiers achieved satisfactory performance with accuracies greater than 96% and almost all precisions, recalls, and F_1-scores were greater than 90%. However, on the test set, the performance of pretrain-LSTM was far from satisfactory. The classification accuracy was only 47.5%. See Table 3 for the classification results on the test set. Among the three networks, CNN did the best on classification with the test accuracy of 98.5%. LSTM ranked the second with an accuracy of 95.7%, following by SVM with an accuracy of 92.2%.

The confusion matrix of the three neural networks and SVM is shown in Table 4. It is clearly shown that only 16 low-risk samples were misclassified as moderate-risk by CNN. LSTM misclassified 46 low-risk samples as moderate-risk, accounting for 18.3% of all the tested low-risk samples. For the pretrain-LSTM method, the majority of real low-risk (204 out of 251) and high-risk (314 out of 347) samples were misclassified as moderate-risk. Therefore, the precision, recall, and F_1-score were 4.0%, 18.7%, and 6.5% for the low-risk group, respectively. These three numbers were 100%, 9.5%, and 17.4% for the high-risk group using pretrain-LSTM, respectively. SVM misclassified 85 DOPs among the three style groups, accounting for 7.8% of the test set.

TABLE 4. Confusion matrix on the test dataset.

Methods	Ground-truth from experts	Recognized by classifier as High	Moderate	Low	Accuracy
CNN	High	347	0	0	98.5%
	Moderate	0	492	0	
	Low	0	16	235	
LSTM	High	347	0	0	95.7%
	Moderate	0	492	0	
	Low	0	46	205	
pretrain-LSTM	High	33	314	0	47.5%
	Moderate	0	492	0	
	Low	0	204	47	
SVM	High	333	14	0	92.2%
	Moderate	12	467	13	
	Low	0	46	205	

Figure 7 illustrates the ROC curves of the four employed methods on the test set. The AUCs for the macro-average ROC curves were 0.97, 0.94, 0.49, and 0.94 for CNN, LSTM, pretrain-LSTM, and SVM, respectively. For high-risk samples, the AUC was 1.00 using either CNN or LSTM. The two numbers on low-risk samples were 0.93 and 0.86 for CNN and LSTM, respectively. However, none of the AUC values was greater than 0.5 when using pretrain-LSTM. In summary, it can be clearly distinguished from Figure 7 that CNN performed the best and pretrain-LSTM was the worst.



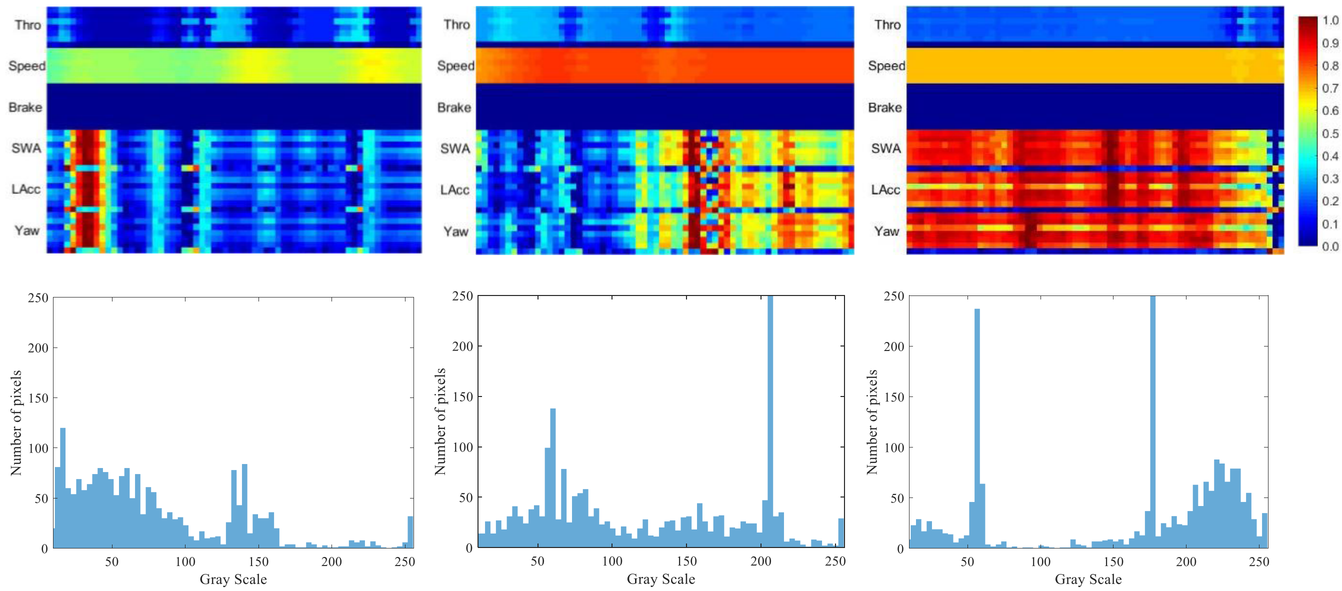
(a) CNN

(b) LSTM

(c) pretrain-LSTM

(d) SVM

FIGURE 7. ROC curves of the neural network algorithms and SVM (The macro-average ROC computes the matrix independently for each group and then take the average (hence treating all groups equally), whereas the micro-average ROC aggregates the contributions of all groups to compute the average.).



(a) Low-risk driving style

(b) Moderate-risk driving style

(c) High-risk driving style

FIGURE 8. DOP example and the corresponding histogram for low-risk, moderate-risk, and high-risk driving styles.

Figure 8 illustrates DOP examples for the low-, moderate-, and high-risk driving styles. A warmer color indicates a higher value of the corresponding 'pixel' after regularization, and a colder color indicates a lower value. The DOPs show that driving style of higher risk usually correlates with faster speed and frequent operations on vehicles' lateral movement. Statistically significant differences ($p < 0.001$) were found on all the 42 features among the style groups, which indicates

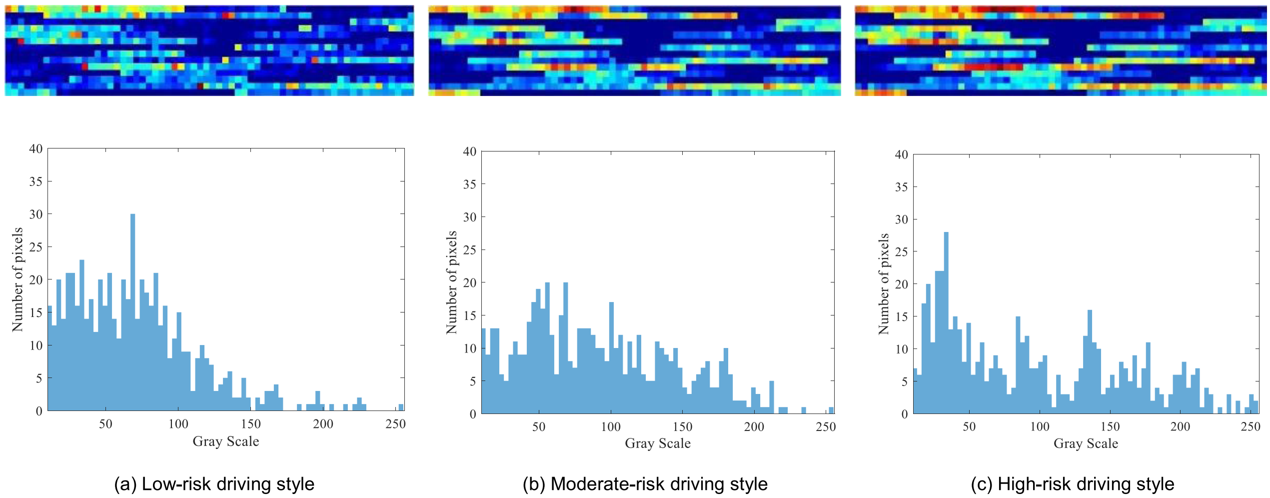


FIGURE 9. The outputs from the second convolution-pooling layer and their histograms.

that the driving operations of drivers with different driving styles greatly differed from each other. Therefore, design of ADASs and intelligent vehicles should take this difference between drivers seriously for the improvement on driving comfort and safety.

Compared with driving style classification performance in previous studies, the classification accuracies of driving style varied from 71.0% to 93.5% when using operational level variables as inputs of classifiers [1], [19], [20]. When using maneuver transition probabilities as classifier inputs, the reported classification accuracy was 93.0% in [17]. In this study, the classification accuracy was 98.5% on the test set when using the developed DOPs as inputs, better than when using the traditional operational variables or maneuver transition features.

B. MISCLASSIFICATION BETWEEN LOW-RISK AND MODERATE-RISK DOPS

The presented results clearly show that misclassifying low-risk driving style as moderate-risk is the main classification error of either CNN or LSTM. The factors leading to this error may be attributed to: (1) The subjective evaluation of driving style (the ground truth) is based on replayed videos as illustrated in Figure 1, and the experts lack the feeling of the real immersed longitudinal and lateral movements experienced by participants. Therefore, the ground-truth style label for a real low- or moderate-risk sample may be mislabeled. (2) Low- and moderate-risk driving perform similarly, which is difficult to be distinguished, even for experienced experts. However, high-risk driving usually has obvious aggressive operations like fast approaching and sharp lane changing which is easy to be distinguished even from videos.

C. OVERFITTING OF PRETRAIN-LSTM

As for the overfitting problems of the pretrain-LSTM method, the inputs of LSTM are the outputs from the second convolution-pooling layer. As shown in Figure 9, there is no obvious distinction between the pictures from different driving style groups, which is different from the qualitative presentation in Figure 8. This is because CNN disintegrates the time sequence relationship in the original DOP by convolution and pooling, which brings too much noise in the inputs of the following LSTM. As LSTM is good at dealing with time sequence problems, the collapsed time sequence relationship and the included noise in the outputs from the convolution-pooling layers lead to the failure of the pretrain-LSTM network.

D. CLASSIFICATION PERFORMANCE OF CNN ON DIFFERENT DRIVERS

We also examined the performance of CNN on different drivers using the leave-one-subject-out method. The DOPs from one driver were selected as the test set, and the DOPs from all the other drivers were combined as the training set. Among the classification results on the 28 drivers, the DOPs from 17 drivers were classified with an accuracy higher than 98%, but the DOPs from another 6 drivers were classified with an accuracy lower than 70%. This probably may be caused by the subjective evaluation error. Therefore, subjective evaluation from more experts should be used to establish a more reliable ground truth in future efforts. In addition, to alleviate the lack of real immersed feeling of vehicle movements from videos, at least some of the expert ratings should be collected onsite in the test vehicle in future efforts.

VII. CONCLUSION

This paper proposes an innovative and effective approach to classify driving style based on constructed driving operational pictures (DOPs) using advanced neural networks. This work extends previous efforts on driving style estimation from post-driving person-based or trip-based classification to quasi-real-time dynamic classification. The application of convolutional neural network (CNN) based on constructed DOPs performs to be the best among the examined networks. The innovatively developed DOP method could be further expanded (e.g., picture dimension expansion by including more features) and applied in driving related studies by combining the state-of-the-art deep learning approaches. Future efforts should focus on the following aspects: (1) Subjective evaluations from experts in the test vehicle are needed to establish a more reliable ground truth. (2) The classification performance when using nested time windows with different lengths needs to be examined to find the optimal combination.

REFERENCES

[1] E. Suzdaleva and I. Nagy, "An online estimation of driving style using data-dependent pointer model," *Transp. Res. C, Emerg. Technol.*, vol. 86, pp. 23–36, Jan. 2018.

[2] F. Sagberg, Selpi, G. F. Piccinini, and J. Engström, "A review of research on driving styles and road safety," *Hum. Factors*, vol. 57, no. 7, pp. 1248–1275, Nov. 2015.

[3] T. Wilhelem, H. Okuda, B. Levedahl, and T. Suzuki, "Energy consumption evaluation based on a personalized driver–vehicle model," *IEEE Trans. Intell. Transp. Syst.*, vol. 18, no. 6, pp. 1468–1477, Jun. 2016.

[4] C. Lv, X. Hu, A. Sangiovanni-Vincentelli, Y. Li, C. M. Martinez, and D. Cao, "Driving-style-based codesign optimization of an automated electric vehicle: A cyber-physical system approach," *IEEE Trans. Ind. Electron.*, vol. 66, no. 4, pp. 2965–2975, Apr. 2019.

[5] H. Bellem, B. Thiel, M. Schrauf, and J. F. Krems, "Comfort in automated driving: An analysis of preferences for different automated driving styles and their dependence on personality traits," *Transp. Res. F, Traffic Psychol. Behav.*, vol. 55, pp. 90–100, May 2018.

[6] Y. Xing, C. Lv, H. Wang, H. Wang, Y. Ai, D. Cao, E. Velenis, and F. Y. Wang, "Driver lane change intention inference for intelligent vehicles: Framework, survey, and challenges," *IEEE Trans. Veh. Technol.*, vol. 68, no. 5, pp. 4377–4390, May 2019.

[7] W. Han, W. Wang, X. Li, and J. Xi, "Statistical-based approach for driving style recognition using Bayesian probability with kernel density estimation," *IET Intell. Transp. Syst.*, vol. 13, no. 1, pp. 22–30, Jan. 2019.

[8] O. Taubman-Ben-Ari, A. Eherenfreund-Hager, and C. G. Prato, "The value of self-report measures as indicators of driving behaviors among young drivers," *Transp. Res. F, Traffic Psychol. Behav.*, vol. 39, pp. 33–42, May 2016.

[9] S. Amado, E. Arıkan, G. Kaça, M. Koyuncu, and B. N. Turkan, "How accurately do drivers evaluate their own driving behavior? An on-road observational study," *Accident Anal. Prevention*, vol. 63, pp. 65–73, Feb. 2014.

[10] O. Taubman-Ben-Ari and D. Yehiel, "Driving styles and their associations with personality and motivation," *Accident Anal. Prevention*, vol. 45, pp. 416–422, Mar. 2012.

[11] D. Tao, R. Zhang, and X. Qu, "The role of personality traits and driving experience in self-reported risky driving behaviors and accident risk among Chinese drivers," *Accident Anal. Prevention*, vol. 99, pp. 228–235, Feb. 2017.

[12] T. G. Brown, M. C. Ouimet, M. Eldeb, J. Tremblay, E. Vingilis, L. Nadeau, J. Pruessner, and A. Bechara, "The effect of age on the personality and cognitive characteristics of three distinct risky driving offender groups," *Personality Individual Differences*, vol. 113, pp. 48–56, Jul. 2017.

[13] B. Ryder, B. Gahr, P. Egolf, A. Dahlinger, and F. Wortmann, "Preventing traffic accidents with in-vehicle decision support systems—The impact of accident hotspot warnings on driver behaviour," *Decis. Support Syst.*, vol. 99, pp. 64–74, Jul. 2017.

[14] L. Eboli, G. Mazzulla, and G. Pungillo, "How drivers' characteristics can affect driving style," *Transp. Res. Procedia*, vol. 27, pp. 945–952, Jan. 2017.

[15] T. Broberg and T. D. Willstrand, "Safe mobility for elderly drivers–considerations based on expert and self-assessment," *Accident Anal. Prevention*, vol. 66, pp. 104–113, May 2014.

[16] Q. Xue, K. Wang, J. J. Lu, and Y. Liu, "Rapid driving style recognition in car-following using machine learning and vehicle trajectory data," *J. Adv. Transp.*, vol. 2019, Jan. 2019, Art. no. 9085238.

[17] G. Li, S. E. Li, B. Cheng, and P. Green, "Estimation of driving style in naturalistic highway traffic using maneuver transition probabilities," *Transp. Res. C, Emerg. Technol.*, vol. 74, pp. 113–125, Jan. 2017.

[18] M. M. Bejani and M. Ghatee, "A context aware system for driving style evaluation by an ensemble learning on smartphone sensors data," *Transp. Res. C, Emerg. Technol.*, vol. 89, pp. 303–320, Apr. 2018.

[19] H. Woo, Y. Ji, Y. Tamura, Y. Kuroda, T. Sugano, Y. Yamamoto, A. Yamashita, and H. Asama, "Dynamic state estimation of driving style based on driving risk feature," *Int. J. Automot. Eng.*, vol. 9, no. 1, pp. 31–38, 2018.

[20] W. Wang, J. Xi, A. Chong, and L. Li, "Driving style classification using a semisupervised support vector machine," *IEEE Trans. Human-Mach. Syst.*, vol. 47, no. 5, pp. 650–660, Oct. 2017.

[21] M. Fernández-Delgado, E. Cernadas, S. Barro, and D. Amorim, "Do we need hundreds of classifiers to solve real world classification problems?" *J. Mach. Learn. Res.*, vol. 15, no. 1, pp. 3133–3181, 2014.

[22] U. R. Acharya, S. L. Oh, Y. Hagiwara, J. H. Tan, M. Adam, A. Gertych, and R. S. Tan, "A deep convolutional neural network model to classify heartbeats," *Comput. Biol. Med.*, vol. 89, pp. 389–396, Oct. 2017.

[23] L. Mou, P. Ghamisi, and X. X. Zhu, "Deep recurrent neural networks for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 7, pp. 3639–3655, Jul. 2017.

[24] Y. Xing, C. Lv, H. Wang, D. Cao, E. Velenis, and F. Y. Wang, "Driver activity recognition for intelligent vehicles: A deep learning approach," *IEEE Trans. Veh. Technol.*, vol. 68, no. 6, pp. 5379–5390, Jun. 2019.

[25] G. Li, S. E. Li, and B. Cheng, "Field operational test of advanced driver assistance systems in typical Chinese road conditions: The influence of driver gender, age and aggression," *Int. J. Automot. Technol.*, vol. 16, no. 5, pp. 739–750, Oct. 2015.

[26] C. M. Martinez, M. Heucke, F.-Y. Wang, B. Gao, and D. Cao, "Driving style recognition for intelligent vehicle control and advanced driver assistance: A survey," *IEEE Trans. Intell. Transp. Syst.*, vol. 19, no. 13, pp. 666–676, Mar. 2016.

[27] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," *Commun. ACM*, vol. 60, no. 6, pp. 84–90, 2017.

[28] A. Graves and J. Schmidhuber, "Framewise phoneme classification with bidirectional LSTM and other neural network architectures," *Neural Netw.*, vol. 18, no. 5, pp. 602–610, 2005.

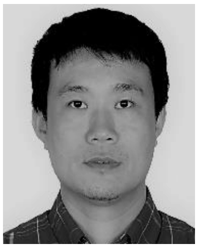
[29] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Comput.*, vol. 9, no. 8, pp. 1735–1780, 1997.

[30] Y. Guo, Y. Liu, E. M. Bakker, Y. Guo, and M. S. Lew, "CNN-RNN: A large-scale hierarchical image classification framework," *Multimed. Tools Appl.*, vol. 77, no. 8, pp. 10251–10271, Apr. 2018.

[31] G. Brown, A. Pocock, M.-J. Zhao, and M. Lujan, "Conditional likelihood maximisation: A unifying framework for information theoretic feature selection," *J. Mach. Learn. Res.*, vol. 13, pp. 27–66, Jun. 2012.

GUOFA LI received the Ph.D. degree from Tsinghua University, Beijing, China, in 2016. From 2012 to 2013, he was a Visiting Scholar with the University of Michigan Transportation Research Institute (UMTRI). He is currently an Assistant Professor with the College of Mechatronics and Control Engineering, Shenzhen University, Guangdong, China. His research interests include autonomous vehicles, driver behavior, advanced driver assistance systems, machine learning techniques, and human factors in automotive and transportation engineering. He was a recipient of the NSK Sino-Japan Outstanding Paper Prize from NSK Ltd. (2014), the Excellent Young Engineer Innovation Award from SAE-China (2017), and the Young Elite Scientists Sponsorship Program by SAE-China (2018).

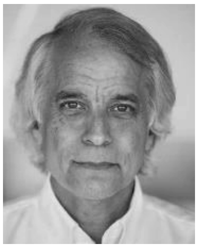
FANGPING ZHU received the B.E. degree in control engineering from Shenzhen University, Guangdong, China, in 2017, where he is currently pursuing the master's degree with the College of Mechatronics and Control Engineering. His main research interests include driving style recognition and advanced driver assistance systems development.



XINGDA QU received the Ph.D. degree in human factors and ergonomics from Virginia Tech, in 2008. He is currently a Professor with the Institute of Human Factors and Ergonomics, Shenzhen University, China. His research interests include transportation safety, occupational safety and health, and human–computer interaction.

BO CHENG received the B.E. and M.E. degrees from Tsinghua University, Beijing, China, in 1985 and 1988, respectively, and the Ph.D. degree from Tokyo University, Tokyo, Japan, in 1998. He is currently a Professor with Tsinghua University, where he is also the Dean of the Suzhou Automotive Research Institute. He has authored more than 70 journal/conference papers and holds more than 20 patents. His active research interests include autonomous vehicles, driver assistance systems, active safety, and vehicular ergonomics. He is also the Chairman of the Academic Board of SAE-Beijing, a Committee Member of the National 863 Plan, a member of the Academic Committee of China Transportation Technology Project, and a member of the Academic Committee of Ministry of Public Security Traffic Management.

SHEN LI received the B.E. degree from Jilin University, in 2012, and the Ph.D. degree from University of Wisconsin–Madison, in 2019. He has participated in many research projects funding by the National Natural Science Foundation of China, the Ministry of Science and Technology (863 projects), and the U.S. Department of Transportation. His research interests include intelligent transportation systems (ITS), architecture design of CAVH systems, cooperative control method of connected vehicles, traffic data mining based on cellular data, and traffic operations and management.



PAUL GREEN received the M.S.E. and Ph.D. degrees from the University of Michigan, Ann Arbor, MI, USA, in 1974 and 1979, respectively. He is currently a Research Professor with the University of Michigan Transportation Research Institute Driver Interface Group, Ann Arbor, MI, USA, and an Adjunct Professor with the Department of Industrial and Operations Engineering, University of Michigan. He teaches automotive human factors and human–computer interaction classes. He is the Leader of the University's Human Factors Engineering Short Course, the flagship continuing education course in the profession, now in its 60th year. His research interests include driver interfaces, driver behavior, driver workload, and the development of standards to get research into practice. He was a Past President of the Human Factors and Ergonomics Society.