2018 IEEE Intelligent Vehicles Symposium (IV)
Changshu, Suzhou, China, June 26-30, 2018

End-to-End Driving Activities and Secondary Tasks Recognition Using Deep Convolutional Neural Network and Transfer Learning

Yang Xing, Jianlin Tang, Hong Liu, Chen Lv, *Member, IEEE, Dongpu Cao, Member, IEEE, Efstathios Velenis, Fei-Yue Wang, Fellow, IEEE*

***Abstract*— Drivers' decision and their corresponding behaviors are important aspects that can affect the driving safety, and it is necessary to understand the driver behaviors in real-time. In this study, an end-to-end driving-related tasks recognition system is proposed. Specifically, seven common driving activities are identified, which are normal driving, right mirror checking, rear mirror checking, left mirror checking, using in-vehicle video device, texting, and answering mobile phone. Among these, the first four activities are regarded as normal driving tasks, while the rest three are divided into distraction group. The images are collected using a consumer range camera, namely, Kinect. In total, five drivers are involved in the naturalistic data collection. Before training the identification model, the raw images are first segmented using a Gaussian mixture model (GMM) to extract the driver region from the background. Then, a pre-trained deep convolutional neural network (CNN) model is trained to classify the behaviors, which directly takes the processed RGB images as the input and outputs the identified label. In this work, the AlexNet is selected as the pre-trained CNN model. Then, to reduce the training cost, the transfer learning mechanism is applied to the CNN model. An average of 79% detection accuracy is achieved for the seven driving tasks. The proposed integration model can be used as a low-cost driver distraction and dangerous tasks recognition modeL.**

I. INTRODUCTION

Driver is the most important part within the Traffic-Driver-Vehicle (TDV) loop. Drivers' decision and their behaviors are the major aspects which can influence the driving safety. It is reported that more than 90% light vehicles accidents are caused by human driver misbehaviors in the United States, and the accident rate can be reduced by 10% to 20% with a correct driver behavior understanding [1]-[5]. Therefore, the recognition of driver behaviors is one of the most important task for the intelligent vehicles. For conventional vehicles, driver is the major component within the TDV loop. The understanding of driver behaviors enables the driver assistance system (DAS) to figure out the optimal vehicle control strategy [6]-[9]. In terms of the intelligent and partially automated vehicles, such as the Level three automated vehicles (based on the definition in Society of Automotive Engineers standard J3016), driver is responsible for taking over the vehicle control authority under certain emergency condition. At that moment, the real time driver behavior and activity monitoring will determine whether the driver is able to takeover or not. This strategy will prevent the vehicle from losing control.

Driver behavior is a widely studied topic. Previous studies mainly focus on the driver attention and distraction (either physical distraction or cognitive distraction) detection [37], driver intention [8] [10], driver drowsiness and fatigue [11]-[13], etc. All of these studies require capturing the driver body information, such as head pose [14], eye gaze [9], hand motion [15], foot dynamics [16], and even physiological signals [17] [18]. In [19], the video information for driver head movement along with audio signals were collected to identify the secondary tasks during driving. In [20], driver's head pose, eye gaze direction, and hand movement were combined together to identify the driver activities. In [21], driver head pose estimation was proposed and applied to the rear-end crash avoidance system. The physiological signals, such as electroencephalogram (EEG) and electrooculography (EOG) are also been widely studied and used for real-time driver status monitoring. In [22], EEG signals were collected to predict driver braking intention. In [23], the EEG and EOG signals were used to monitor driver drowsiness and fatigue status.

According to the aforementioned works, most of existing driver behavior and activity monitoring systems require extracting some specific features such as the head pose angle, gaze direction, and the position of hand and body joints [24]. These features are not always easily to be obtained, and some even require high-cost hardware devices, which increase both the temporal and financial cost. Recently, significant progress has been gained in the computer vision area due to the development of CNN model and large-scale annotated dataset. CNN models have achieved the state-of-art results in many object detection, object classification, and segmentation tasks. Meanwhile, it has been successfully applied to the driver monitoring tasks [25] [26]. Therefore, in this work, a low-cost end-to-end CNN model is proposed to recognize the driver behavior and secondary tasks. The sensor required in this study is just a RGB camera. According to the report in [27], seven common in-vehicle tasks are selected. These tasks contain both normal driving activities and secondary tasks to reflect the distraction behavior. The CNN model will take the processed images directly without any complex feature extraction. With the help of transfer learning, the pre-trained CNN model is fine-tuned to satisfy the behavior detection task.

Y. Xing is with Advanced Vehicle Engineering Centre, Cranfield University, Bedford, MK43 0AL, UK, and also with Vehicle Intelligence Pioneers Ltd, Qingdao, China (Email: y.xing@cranfield.ac.uk) (Y.Xing and JL.Tang contributed equally to this paper)

JL.Tang and H. Liu are with JIANGSU XCMG CONSTRUCTION MACHINERY RESEARCH INSTITUTE LTD. Add: No. 26, Tuolanshan Road, Economic Development Zone, Xuzhou, Jiangsu, P.R. China 221004. (Email:{Tangjianlin, liuhong}@xcmg.com)

C. Lv and E. Velenis are with Advanced Vehicle Engineering Centre, Cranfield University, Bedford, MK43 0AL, UK (Email:{c.lyu, e.velenis}@cranfield.ac.uk)

D. Cao is with the School of Mechanical and Mechatronics Engineering, University of Waterloo, 200 University Avenue West Waterloo, ON, N2L 3G1, Canada (Email: dongp.ca@yahoo.com)

F.-Y. Wang is with the State Key Laboratory of Management and Control for Complex Systems, Institute of Automation, Chinese Academy of Sciences, Beijing 100190, P.R. China (Email: feiyue@ieee.org).

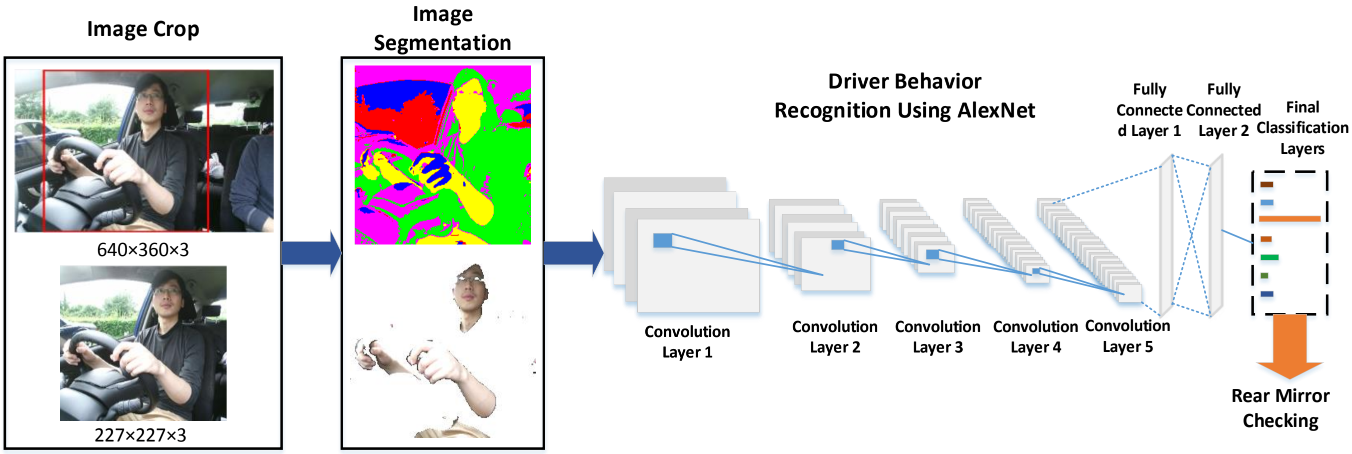

Fig. 1. Overall System Architecture.

II. Methodology

This section describe the experiment and the methodology used in this work. Fig. 1 illustrates the general system architecture. The raw images are firstly collected using Kinect. Then, the driver images are segmented using GMM. Finally, the CNN model is applied to the activities recognition task. Section II.A briefly describes the experimental setup and data collection procedure. Section II.B describes the concept of GMM and its application in image segmentation. Section II.C introduces the structure of CNN model and transfer learning mechanism.

A. Experiment and Data Collection

In this work, driver behavior data are collected using a low-cost range camera, namely, Kinect. Kinect supports the collection of multi-modal signals, such as the color image, depth image, and audio signals. It was originally designed for indoor computer-human interaction, and has been successfully applied to some driver monitoring studies [28] [29]. As described in [24], the driver head pose and upper body joints can be detected using Kinect. However, in this work, only the RGB images are captured in this work for driver activity recognition.

According to the Kinect mounting requirements [30], Kinect is mounted in the middle of the front window, facing the upper body of the driver so that it does not interfere driver's field of view. The sampling rate for the image collection is eight frames per seconds. According to the study in [19], short-term driver behaviors, such as mirror checking can last from 0.5 to 1 second. Therefore, the sampling rate is fast enough to capture these behaviors. The data are sampled with an Intel Core i7 2.5GHz computer and the codes are written in C++ based on the Windows Kinect SDK and OpenCV. The raw images are compressed to $640 \times 360 \times 3$ format to increase the computation efficiency. There are five drivers involved in the experiment. They were required to perform seven common activities, which contains four normal driving tasks (normal driving, right mirror checking, rear mirror checking, and left mirror checking) and three secondary tasks (using in-vehicle video device, answering mobile phone, and texting). It takes around 20 minutes for each driver to perform all the tasks, and there are 2.4k images are captured in total. The general device setup is shown in Fig. 2.

Fig. 2. Experiment setup. The Kinect is mounted in the middle of the front window and data are collected using a laptop.

A. Image Pre-processing and Segmentation

As aforementioned, the raw images are stored in the format of$640 \times 360 \times 3$. To speed up the CNN training process and increase the classification accuracy, the raw images are firstly cropped to find the rough driving area for the driver. An interest of region (ROI) which covers the upper body of the driver is selected. Fig. 1 illustrates the raw image and selected area.

After the raw images are cropped, these images are firstly transformed into the size of $227 \times 227 \times 3$ to satisfy the input requirement of the AlexNet. Then, the GMM algorithm is applied to segment the images and extract the driver from the background. GMM is an unsupervised machine learning method, which can be used for data clustering and data mining. GMM is a probability density function that represented by a weighted sum of sub Gaussian components [31]. One of the advantage of using GMM to segment the images is as an unsupervised learning method, GMM requires no training labels and can be very flexible by choosing different number of clusters. To train a GMM for image segmentation, each image is represented by a feature vector according to the pixel intensity. The feature vector for the GMM is a three-dimensional vector that contains the RGB intensity of each pixel.

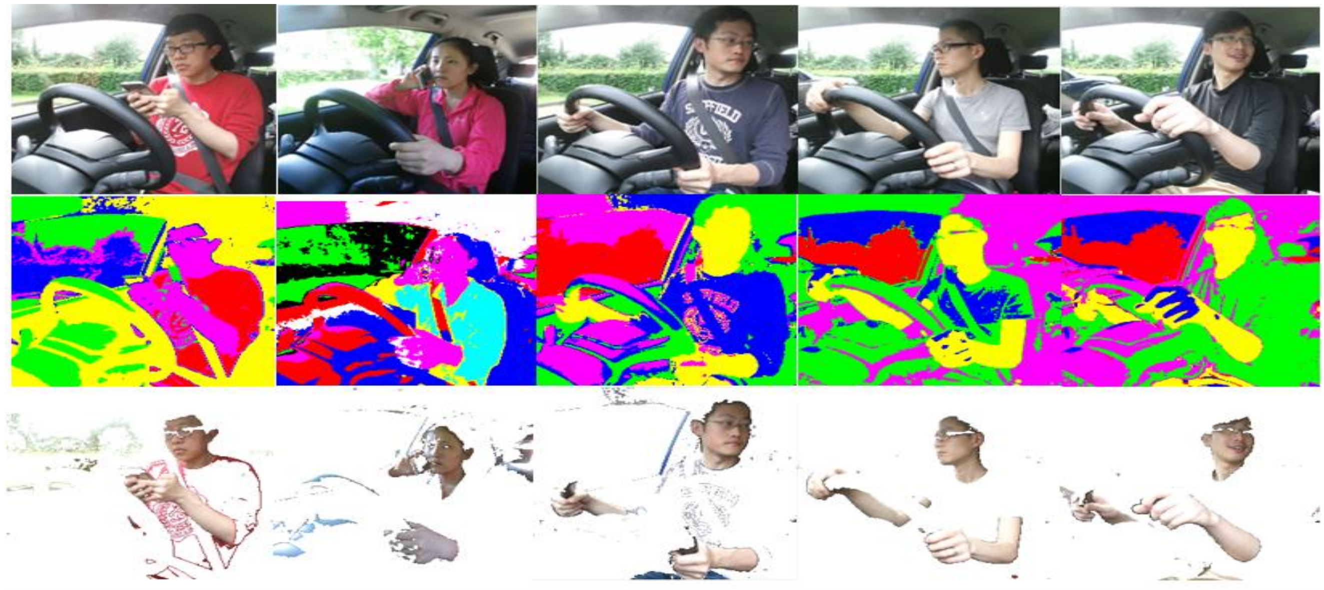

Fig. 3. Image Segmentation Results using GMM. The top row images are the original images. The middle row images are segmented images. The final row images are the corresponding highlighted images.

$$x_k = [R(u,v),\ G(u,v),\ B(u,v)] \quad (1)$$

where x is the feature vector of a cropped image, k is the index of x which has a maximum value of 227×227. (u,v) is the image coordinates.

The GMM can be represented as the following equation:

$$p(x_i|\theta) = \sum_{k=1}^{K} \pi_k N(x_i|\mu_k, \Sigma_k) \quad (2)$$

where x_i is the 3-Dimensional feature vector, θ is the parameters of GMM, K is the total number of components in the model (five in this case), π_k is the weight of each component Gaussian distribution function and the sum of π_k equals to one. μ and $\sum$ are the mean and covariance parameter of multivariate Gaussian function.

$N(x_i|\mu_k, \Sigma_k)$ is the univariate Gaussian distribution function in this case with the form as follows.

$$N(x|\mu, \Sigma) = \frac{1}{(2\pi)^{D/2}|\Sigma|^{1/2}} \exp[-\frac{1}{2}(x-\mu)^T \Sigma^{-1}(x-\mu)] \quad (3)$$

where D is the dimension of the data vector. A complicate GMM contains three parameters and can be represented as:

$$\theta = \{\pi_k, \mu_k, \Sigma_k\} \quad (4)$$

The most common method for GMM training is the Expectation-Maximization (EM) maximum likelihood estimation algorithm. It computes the maximization of the cost function in an iterative manner. The detail description of EM algorithm can be found in [32].

Fig. 3 illustrates the image segmentation results based on the GMM. The driver body and skin can be identified using GMM segmentation. After that, only the driver head and hand pixels are remained for CNN model training. As shown in the results section, the GMM-segmentation-based method leads to a more accurate detection compared with the model that trained with the raw images.

C. Transfer Learning and Model Training

Currently, deep convolutional neural networks have gain a tremendous improvement in the domain of computer vision. One of the major reason is the distribution of the ImageNet dataset [33]. ImageNet is a large-scale dataset which contains more than 15 million high-resolution annotated natural images of over 22,000 categories. The large amount of annotated images benefit the training of deeper and more accurate CNN models. In this work, a pre-trained deep convolutional neural network model, namely, AlexNet was chosen as the basic model structure for the recognition of driver behavior. AlexNet was first proposed by Alex Krizhevsky in 2012 [34]. The model won the ImageNet Large Scale Visual Recognition Challenge (ILSVRC12) with a much more accurate detection accuracy compared with other models at that time. The model was trained for the classification of 1000 categories. There are five convolutional layers and three fully connected neural network layers in the model with non-linearity and pooling layers between the convolutional layers. In total, AlexNet contains 60 million parameters and 650,000 neurons.

To train such a deep convolutional neural networks from scratch, a large-scale annotated dataset like ImageNet is need. However, In general, large-scale annotated datasets are not always available. Therefore, common ways of using the pre-trained deep CNN model are to treat the model as a fixed feature extractor without tuning the model parameters, or fine-tune the pre-trained model parameters with limited dataset. In this part, the AlexNet model is used in the second manner, which is to fine-tune the model with the collected driver behavior dataset. Since original AlexNet is trained to classify 1000 categories, the last few layers have to be modified so that the model satisfies the seven objects classification problem. In this work, the structure of the last three layers of AlexNet are modified. The original last fully connected layer and output layer, which generate the probabilities for 1000 categories, are replaced by a fully connected layer and a classification layer that output the probabilities for the seven categories. The basic structure and properties of the convolutional layers are remained so that these layers can keep their advantages in the feature extraction, and the knowledge that learned from the ImageNet dataset can be transferred to the driver behavior domain.

For transfer learning, a small initial learning rate is selected to slow down the updating rate for the convolutional layers. On contrary, a larger learning rate factors for the last fully connected layer is chosen to speed up the learning rate in the final layers. With this kind of combination, the new model is trained to satisfy the new classification tasks. The convolutional neural network layers of the new model maintains the key functionality of the AlexNet, which is designed for hierarchical feature extraction. The feature extraction knowledge that learned from the large-scale dataset will be transferred to the driver behavior domain. The last three layers are fine-tuned to satisfy the driver behavior classification tasks. A detailed discussion on transfer learning can be found in [38].

III. Results and Discussion

In this section, the experimental results for driving tasks recognition are proposed. The system performances are evaluated from three major aspects: the impact of GMM-based image segmentation on the recognition of driver behaviors, the classification results compared with the feature extraction methods, and driver physical distraction identification.

A. The Impact of GMM Image Segmentation on Driving Tasks Recognition

Firstly, the recognition performance for the five participants are evaluated. The seven driving related tasks are ordered as {normal driving, right mirror checking, rear mirror checking, left mirror checking, using video device, texting, and answering mobile phone}. Table1 illustrates the classification results for these tasks. Specifically, the upper part indicates the classification results with GMM image segmentation, while the lower part illustrates the classification results using the raw RGB images. T1 to T7 represents the seven tasks and D1 to D5 represents the five drivers. The models are trained using MATLAB Deep Learning toolbox. The model performance is evaluated using *leave-one-out (LOO)* cross validation method. To get the activities identification results for a certain driver, the images of this driver are used for testing purpose only, and the images of the rest four drivers are used for model training. Therefore, for these drivers, their images are completely new to the corresponding CNN model and the identification performances equals to the model generalization accuracy in this case.

As shown in Table 1, the general identification accuracy for the GMM segmentation-based AlexNet achieved an average of 79.1% accuracy, which is much better than that without using image segmentation (50.3%). The average performance is defined as the average results of the five drivers. In terms of the accuracy on each task, the right mirror checking gets the most accurate detection result among the five drivers, and the answering mobile phone behavior achieved the second best result.

TABLE 1
Classification Results for Driving Tasks Recognition

	Driving Tasks				Non-Driving Tasks			
GMM	T1	T2	T3	T4	T5	T6	T7	Ave
D1	0.921	0.990	0.929	0.398	0.904	0.897	0.999	0.831
D2	0.996	0.906	0.920	0.449	0.595	0.242	0.936	0.753
D3	0.889	0.989	0.176	1.00	0.982	0.982	0.732	0.832
D4	0.353	0.994	0.229	0.813	1.00	0.982	0.979	0.752
D5	0.958	0.998	0.934	0.236	0.666	0.816	0.998	0.786
Mean	0.823	0.975	0.637	0.579	0.829	0.783	0.929	0.791
Raw	T1	T2	T4	T4	T5	T6	T7	Ave
D1	0.188	0.999	0.998	0.00	0.00	0.942	0.893	0.534
D2	0.593	0.949	0.01	0.00	0.789	0.215	0.957	0.408
D3	0.200	0.604	0.00	0.00	1.00	1.00	0.504	0.437
D4	0.00	1.00	0.00	0.875	1.00	0.98	0.00	0.482
D5	0.292	0.996	1.00	0.00	0.787	1.00	0.993	0.657
Mean	0.254	0.906	0.402	0.175	0.715	0.827	0.669	0.503

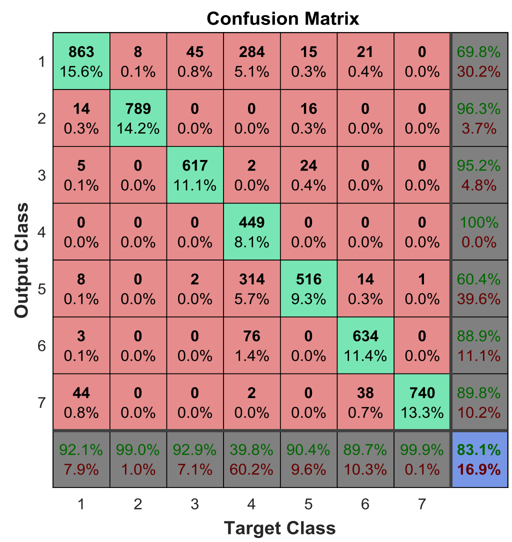

Fig. 4. Confusion matrix for driver 1.

The worst result happens on the left mirror checking case, both for using GMM segmentation and using raw images. For the normal driving activities recognition, the rear mirror checking and left mirror checking results are not as accurate as the other two mirror checking behaviors. One possible explanation is that these two behaviors show similar profile, which can confuse the CNN model sometimes. Another evidence that can be drawn from Table 1 is that the CNN model achieved better detection rates on the secondary tasks in average. This mainly due to the fact that, when performing the secondary tasks, the driver has to move their body and hands instead of only rotating their head, which makes the detection easier. Similar results were obtained in [24].

Fig. 4 illustrates the confusion matrix for driver 1. The green diagonal shows the correct detection cases for the class. The bottom row shows the classification accuracy with respect to the target class, while the right most column shows the classification accuracy with respect to the predicted labels. As shown in Fig. 4, all the driving tasks except the fourth task (left mirror checking) achieved reasonable detection rates. In terms of the fourth task, there are 284 cases are misclassified into the normal driving case and 314 cases are misclassified into the video device setup group. The general detection time cost for the each image is about 13ms using the AlexNet.

B. Comparison between Transfer Learning and Feature Extraction

As discussed in the last section, a pre-trained CNN model normally can be used in two different manners. The first one is only fine-tune the last few layers to make the model satisfy the new task, as aforementioned.

The second common usage is using the pre-trained CNN model as a feature extractor [35]. The deep CNN models are trained on large-scale dataset, which makes the convolutional layers have a strong representation ability of the objects. The lower level of the convolutional layers are more concentrate on the local features, such as edges, and corners. While the deeper convolutional layers focus on the higher-level features. The combination of feature extraction and conventional machine learning algorithms like Support Vector Machine (SVM) is more convenient to use. Therefore, in this part, vision-based feature extraction method for driving activities recognition will be studied.

Two different features are collected based on the segmented images. The first feature set is generated using Histogram of oriented gradients (HOG) [36], with a block size of 2×2 and cell size 8×8. The second features set is generated by the fifth convolutional layer of the AlexNet. The Principal Component Analysis (PCA) algorithm is used to reduce the dimension of the feature vector. In this work, the dimension of the feature vector is reduced to 50. Meanwhile, two different classification models are evaluated, which are the SVM and Feedforward Neural Network (FFNN). The classification results are shown in Table 2. In Table 2, S+C means SVM and Conv5 feature. S+H means SVM and HOG features. Similarly, F+C means FFNN and Conv5 feature. F+H means FFNN and HOG features. As shown in Table 2, the feature extraction methods are unable to accurately identify the driving tasks. The average results for the four combination are much lower compared with transfer learning method. Hence, transfer learning is proved to be more suitable for this task.

C. Driver Distraction Detection Using Binary Classifier

Lastly, we evaluate the model performance on the binary classification case, which is to separate the normal driving and the secondary tasks. At this moment, the first four tasks are grouped together and the last three tasks constitute another group. As shown in Fig. 5. The identification accuracy for the normal driving tasks and the distracted tasks are 88.2% and 93.9%, respectively. The general accuracy is 90.3%. From the safety perspective of view, it is acceptable or less dangerous to misclassify the normal driving tasks into distracted group, although this may disrupt the drivers and reduce their confidence on the assistance system. While, if taking the normal driving as positive class, the false positive rate should be reduced as much as possible to keep the driver concentrate on the driving task.

It should be noticed that, in this case, there are no smoothing schemes are applied to the distraction warning module. In most cases, the driver assistance system will only warn the driver if distraction cases happen continuously in a period. Therefore, if applying short period smoothing or voting technique, the binary classification accuracy is expected to be increased.

TABLE 2
CLASSIFICATION RESULTS USING FEATURE EXTRACTION

	D1	D2	D3	D4	D5	Ave
S+C	0.497	0.344	0.331	0.236	0.249	0.331
S+H	0.509	0.395	0.270	0.147	0.122	0.288
F+C	0.304	0.298	0.537	0.250	0.289	0.335
F+H	0.428	0.480	0.454	0.173	0.136	0.334

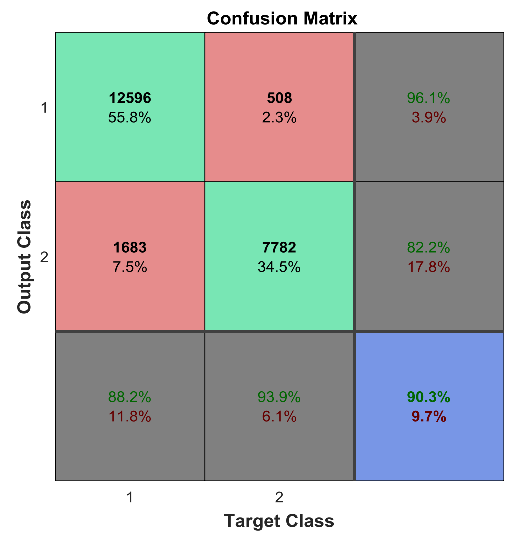

Fig. 5. Confusion matrix for the binary classification result

IV. CONCLUSION

In this work, an end-to-end driving related activities recognition system based on CNN model is proposed. The deep CNN model is trained with transfer learning method. Specifically, the pre-trained AlexNet is fine-tuned to satisfy the seven driving tasks recognition problem. To increase the identification accuracy, the raw RGB images are first processed using a GMM-based segmentation algorithm. The GMM method can efficiently identify the drivers from the background environment and remove the irrelevant objects. The classification results indicate that the GMM segmentation based CNN model gives a much better result than that the model which use raw images. Another comparison is made between the transfer learning and feature extraction methods. Results show that no matter using HOG or CNN features, feature extraction methods are not able to accurately identify the driver behaviors. Finally, if using the CNN model as a binary classifier, the distraction detection rate can achieve nearly 94% accuracy. Although the classification accuracy in this work is no better than the result in [24], the end-to-end transfer learning scheme only require RGB images as input. Meanwhile, it does not need complicate feature extraction and selection process, which makes it much easier to be implemented.

As shown in this work, some behaviors cannot be detected accurately. Future works may concentrate on evaluating different CNN models such as VGG and GoogLeNet. Meanwhile, the depth information can be adopted to increase the robustness of the behavior identification.

REFERENCES

[1] Wang, Fei-Yue, et al. "Parallel driving in CPSS: a unified approach for transport automation and vehicle intelligence." *IEEE/CAA Journal of Automatica Sinica* 4.4 (2017): 577-587.

[2] Foroutan, N.,Hamzeh, A. "Discovering the Hidden Structure of a Social Network: A Semi Supervised Approach." IEEE Transactions on Computational Social Systems, 3(4): 151-163, 2016.

[3] Tong, G., et al. "Effector Detection in Social Networks." IEEE Transactions on Computational Social Systems, 4(1): 14-25, 2017.

[4] Lv, C., Wang, H., and Cao, D. "High-Precision Hydraulic Pressure Control Based on Linear Pressure-Drop Modulation in Valve Critical Equilibrium State." IEEE Transactions on Industrial Electronics, 2017.

[5] Liang, Yulan, and John D. Lee. "A hybrid Bayesian Network approach to detect driver cognitive distraction." Transportation research part C: emerging technologies 38 (2014): 146-155.

[6] Xing, Yang, et al. "Driver workload estimation using a novel hybrid method of error reduction ratio causality and support vector machine." *Measurement* 114 (2018): 390-397.

[7] Lv, Chen, et al. "Analysis of autopilot disengagements occurring during autonomous vehicle testing." *IEEE/CAA Journal of Automatica Sinica* 5.1 (2018): 58-68.

[8] Jain, Ashesh, et al. "Car that knows before you do: Anticipating maneuvers via learning temporal driving models." Proceedings of the IEEE International Conference on Computer Vision. 2015.

[9] Vicente, Francisco, et al. "Driver gaze tracking and eyes off the road detection system." IEEE Transactions on Intelligent Transportation Systems 16.4 (2015): 2014-2027.

[10] McCall, Joel C., et al. "Lane change intent analysis using robust operators and sparse bayesian learning." *IEEE Transactions on Intelligent Transportation Systems* 8.3 (2007): 431-440.

[11] Wang, Xuesong, and Chuan Xu. "Driver drowsiness detection based on non-intrusive metrics considering individual specifics." *Accident Analysis & Prevention* 95 (2016): 350-357.

[12] Chai, Rifai, et al. "Driver fatigue classification with independent component by entropy rate bound minimization analysis in an EEG-based system." *IEEE journal of biomedical and health informatics* 21.3 (2017): 715-724.

[13] Mandal, Bappaditya, et al. "Towards detection of bus driver fatigue based on robust visual analysis of eye state." *IEEE Transactions on Intelligent Transportation Systems* 18.3 (2017): 545-557.

[14] Murphy-Chutorian, Erik, and Mohan Manubhai Trivedi. "Head pose estimation and augmented reality tracking: An integrated system and evaluation for monitoring driver awareness." IEEE Transactions on intelligent transportation systems 11.2 (2010): 300-311.

[15] Das, Nikhil, Eshed Ohn-Bar, and Mohan M. Trivedi. "On performance evaluation of driver hand detection algorithms: Challenges, dataset, and metrics." Intelligent Transportation Systems (ITSC), 2015 IEEE 18th International Conference on. IEEE, 2015.

[16] Tran, Cuong, Anup Doshi, and Mohan Manubhai Trivedi. "Modeling and prediction of driver behavior by foot gesture analysis." Computer Vision and Image Understanding 116.3 (2012): 435-445.

[17] Bi, Luzheng, et al. "Queuing Network Modeling of Driver EEG Signals-Based Steering Control." *IEEE Transactions on Neural Systems and Rehabilitation Engineering* 25.8 (2017): 1117-1124.

[18] Teng, Teng, Luzheng Bi, and Yili Liu. "EEG-Based Detection of Driver Emergency Braking Intention for Brain-Controlled Vehicles." *IEEE Transactions on Intelligent Transportation Systems* (2017).

[19] Li, Nanxiang, and Carlos Busso. "Detecting drivers' mirror-checking actions and its application to maneuver and secondary task recognition." IEEE Transactions on Intelligent Transportation Systems 17.4 (2016): 980-992.

[20] Ohn-Bar, Eshed, et al. "Head, eye, and hand patterns for driver activity recognition." Pattern Recognition (ICPR), 2014 22nd International Conference on. IEEE, 2014.

[21] Rezaei, Mahdi, and Reinhard Klette. "Look at the driver, look at the road: No distraction! No accident!" Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2014.

[22] Kim, Il-Hwa, et al. "Detection of braking intention in diverse situations during simulated driving based on EEG feature combination." *Journal of neural engineering* 12.1 (2014): 016001.

[23] Zhang, Chi, Hong Wang, and Rongrong Fu. "Automated detection of driver fatigue based on entropy and complexity measures." *IEEE Transactions on Intelligent Transportation Systems* 15.1 (2014): 168-177.

[24] Yang, X, Lv, C., Cap, D., et al. Identification and Analysis of Driver Postures for In-Vehicle Driving Activities and Secondary Tasks Recognition. *IEEE Transactions on Computational Social Systems* 99 (2018):1-14.

[25] Le, T. Hoang Ngan, et al. "DeepSafeDrive: A grammar-aware driver parsing approach to Driver Behavioral Situational Awareness (DB-SAW)." *Pattern Recognition* 66 (2017): 229-238.

[26] Yuen, Kevan, Sujitha Martin, and Mohan M. Trivedi. "Looking at faces in a vehicle: A deep CNN based approach and evaluation." *Intelligent Transportation Systems (ITSC), 2016 IEEE 19th International Conference on*. IEEE, 2016.

[27] Sivak M, Schoettle B. Motion sickness in self-driving vehicles. UMTRI-2015-12. (2015). Ann Arbor: University of Michigan Transportation Research Institute.

[28] Gaglio, Salvatore, Giuseppe Lo Re, and Marco Morana. "Human activity recognition process using 3-D posture data." IEEE Transactions on Human-Machine Systems 45.5 (2015): 586-597.

[29] Neto, Laurindo Britto, et al. "A Kinect-based wearable face recognition system to aid visually impaired users." IEEE Transactions on Human-Machine Systems 47.1 (2017): 52-64.

[30] Mehran Azimi. Skelton Joint Smoothing White Paper [Online]. Available: https://msdn.microsoft.com/en-us/library/jj131429.aspx.

[31] Rasmussen, Carl Edward. "The infinite Gaussian mixture model." *Advances in neural information processing systems*. 2000.

[32] Xu, Lei, and Michael I. Jordan. "On convergence properties of the EM algorithm for Gaussian mixtures." *Neural computation* 8.1 (1996): 129-151.

[33] Deng, Jia, et al. "Imagenet: A large-scale hierarchical image database." *Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on*. IEEE, 2009.

[34] Krizhevsky, Alex, Ilya Sutskever, and Geoffrey E. Hinton. "Imagenet classification with deep convolutional neural networks." *Advances in neural information processing systems*. 2012.

[35] Deo, Nachiket, Akshay Rangesh, and Mohan Trivedi. "In-vehicle Hand Gesture Recognition using Hidden Markov models." *Intelligent Transportation Systems (ITSC), 2016 IEEE 19th International Conference on*. IEEE, 2016.

[36] Dalal, Navneet, and Bill Triggs. "Histograms of oriented gradients for human detection." *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on*. Vol. 1. IEEE, 2005.

[37] Tawari, Ashish, et al. "Looking-in and looking-out vision for urban intelligent assistance: Estimation of driver attentive state and dynamic surround for safe merging and braking." *Intelligent Vehicles Symposium Proceedings, 2014 IEEE*. IEEE, 2014.

[38] Pan, Sinno Jialin, and Qiang Yang. "A survey on transfer learning." *IEEE Transactions on knowledge and data engineering* 22.10 (2010): 1345-1359.