Driver Activity Recognition for Intelligent Vehicles: A Deep Learning Approach

A Seminar Report

Submitted to the APJ Abdul Kalam Technological University

in partial fulfillment of requirements for the award of degree

Bachelor of Technology

in

Information Technology

by

Jinso Raj

TRV18IT032

DEPARTMENT OF INFORMATION TECHNOLOGY

GOVERNMENT ENGINEERING COLLEGE BARTON HILL

KERALA

January 2022

DEPT. OF INFORMATION TECHNOLOGY

GOVERNMENT ENGINEERING COLLEGE BARTON HILL

2021 - 22

CERTIFICATE

This is to certify that the report entitled **Driver Activity Recognition for Intelligent Vehicles: A Deep Learning Approach** submitted by **Jinso Raj** (TRV18IT032), to the APJ Abdul Kalam Technological University in partial fulfillment of the B.Tech. degree in Information Technology is a bonafide record of the seminar work carried out by him under our guidance and supervision. This report in any form has not been submitted to any other University or Institute for any purpose.

Dr. Shamna H R
(Seminar Guide)
Associate Professor
Dept.of IT
Government Engineering College Barton Hill
Thiruvananthapuram

Dr. Haripriya AP
(Seminar Coordinator)
Associate Professor
Dept.of IT
Government Engineering College Barton Hill
Thiruvananthapuram

Dr Vijayanand K S
Associate Professor and Head
Dept.of IT
Government Engineering College Barton Hill
Thiruvananthapuram

DECLARATION

I Jinso Raj hereby declare that the seminar report **Driver Activity Recognition for Intelligent Vehicles: A Deep Learning Approach** , submitted for partial fulfillment of the requirements for the award of degree of Bachelor of Technology of the APJ Abdul Kalam Technological University, Kerala is a bonafide work done by me under supervision of Dr. Shamna H R

This submission represents my ideas in my own words and where ideas or words of others have been included, I have adequately and accurately cited and referenced the original sources.

I also declare that I have adhered to ethics of academic honesty and integrity and have not misrepresented or fabricated any data or idea or fact or source in my submission. I understand that any violation of the above will be a cause for disciplinary action by the institute and/or the University and can also evoke penal action from the sources which have thus not been properly cited or from whom proper permission has not been obtained. This report has not been previously formed the basis for the award of any degree, diploma or similar title of any other University.

Jinso Raj

Thiruvananthapuram

17-01-2022

Abstract

It is important to drive safely because it can save our life. Driver decisions and behaviors are essential factors that can affect the driving safety. The accident rate can be reduced by 10% to 20% with a precise driver behavior monitoring system. To understand the driver behaviors, a driver activities recognition system is designed based on the deep convolutional neural networks (CNN). Specifically, seven common driving activities are identified, which are the normal driving, right mirror checking, rear mirror checking, left mirror checking, using in-vehicle radio device, texting, and answering the mobile phone, respectively. Among these activities, the first four are regarded as normal driving tasks, while the rest three are classified into the distraction group. The experimental images are collected using a low-cost camera, and ten drivers are involved in the naturalistic data collection. The raw images are segmented using the Gaussian mixture model (GMM) to extract the driver body from the background before training the behavior recognition CNN model. To reduce the training cost, transfer learning method is applied to fine tune the pre-trained CNN models. Three different pre-trained CNN models, namely, AlexNet, GoogLeNet, and ResNet50 are adopted and evaluated. Then, the CNN models are trained for the binary classification task and identify whether the driver is being distracted or not. Finally, the data will be further analyzed, and the model will be updated to increase the system robustness and detection accuracy. Meanwhile, the system will be tested and used for driver or passenger behavior analysis on the partially automated vehicles in the real world.

Acknowledgement

I take this opportunity to express my deepest sense of gratitude and sincere thanks to everyone who helped me to complete this work successfully. I express my sincere thanks to **Dr Vijayanand K S**, Head of Department, Information Technology, Govt. Engineering College Barton Hill Thiruvananthapuram for providing me with all the necessary facilities and support.

I would like to express my sincere gratitude to **Dr. Haripriya AP**, department of Information Technology, Govt. Engineering College Barton Hill Thiruvananthapuram for their support and co-operation.

I would like to place on record my sincere gratitude to my seminar guide **Dr. Shamna H R**, Associate Professor, Information Technology, Govt. Engineering College Barton Hill for the guidance and mentorship throughout the course.

Finally I thank my family, and friends who contributed to the succesful fulfilment of this seminar work.

Jinso Raj

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Driver is in the center of the Road-Vehicle-Driver loop. So driver decision and behaviors are the major aspects that can affect driving safety. The accident rate can be reduced by 10% to 20% with a precise driver behavior monitoring system. Therefore, the recognition of driver behaviors is becoming one of the most important tasks for intelligent vehicles. Regarding the intelligent and highly automated vehicles, such as the Level-3 automated vehicles, the driver is responsible for taking over the vehicle control under emergencies. At this moment, the real-time driver behavior and activity monitoring system has to decide whether the driver can take over or not. The recognition models are trained to identify seven common driving-related tasks and also to determine whether the driver is being distracted or not. With this end-to-end approach, intelligent vehicles can better interact with human drivers and properly making decisions and generating human-like driving strategies.

Driver behaviors have been widely studied over the past two decades. Previous studies mainly focus on the driver attention and distraction (either physical distraction or cognitive distraction), driver intention, driver styles, driver drowsiness and fatigue detection, etc. To understand the driver behaviors, most of the studies require capturing the driver status information, such as the head pose, eye gaze, hand motion, foot dynamics, and even the physiological signals. These features are not always easy to be obtained, and some even require specific hardware devices, which will increase either the temporal or the financial cost. Therefore, in this work, an end-to- end driver activity recognition system is proposed based on the deep CNN models.

Chapter 2

Experiment and data collection

2.1 System overview

In this work, three different CNN models will be evaluated for driver activities recognition and distraction detection tasks. The only sensor required in this study is a low-cost RGB camera. Based on the report in, seven most common in-vehicle activities for both manual driving and automated driving vehicles are selected, which contains normal driving activities as well as secondary tasks. The CNN models take the processed images directly without any manual feature extraction procedure. By applying the transfer learning scheme, the pre-trained CNN models can be efficiently fine-tuned to satisfy the behaviors detection task.

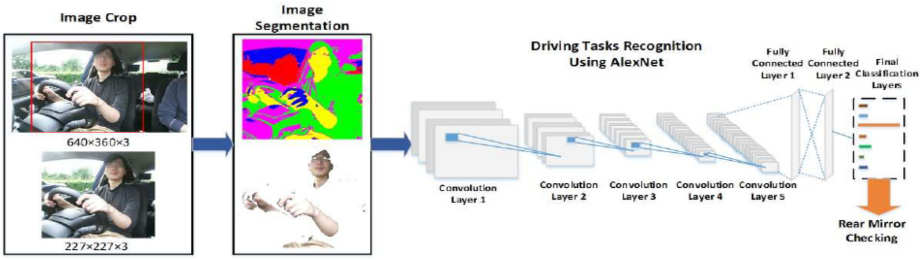

Figure 2.1: Design of the system

This system can be summarized as follows.

1. A deep learning-based approach is applied to identify driver behaviors.

2. Transfer learning is applied to fine-tune the pretrained deep CNN models.

3. An unsupervised GMM-based segmentation method is applied to process the raw images and extract the driver body region from the background.

2.2 Experiment

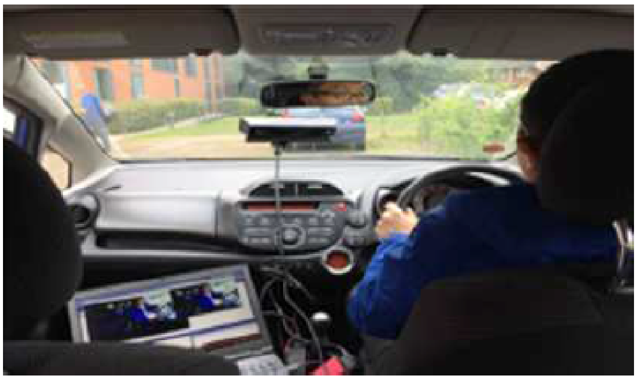

Fig. 2. Experiment setup. The Kinect is mounted on the middle of the front window and data are collected using a laptop.

First, raw RGB images are collected using the Kinect camera. Then, the cropped images are segmented using the GMM algorithm. Finally, the CNN model is adopted for the activities recognition task. Specifically, driver behavior images are collected with a Kinect camera. The Kinect enables the collection of multi-modal signals, such as the color image, depth image, and audio signals. It was initially designed for indoor human-computer-interaction and has been successfully used for driver monitoring systems. The drivers' head poses, and upper body joints also can be detected using the Kinect. While in this study, only the RGB images are used. According to the Kinect application requirements, it was mounted in the middle of the front window,facing the upper body of the driver. The data are recorded with an Intel Core i7 2.5GHz CPU, and the codes are written in C++ based on the Windows Kinect SDK and OpenCV.
Ten drivers are involved in the experiment. They were asked to perform seven activities, which consist of four normal driving tasks (normal driving, left mirror checking, right mirror checking, and rear mirror checking) and three secondary tasks (using in-vehicle radio/video device, answering mobile phone,and texting). It took about 20 to 30 minutes for each driver to finish all these tasks, and about 34 thousands

images were captured in total. In this experiment, five drivers were asked to perform these tasks during driving in a testing field, while the rest five drivers were asked to mimic the driving tasks and not drive the vehicle.

Chapter 3

Methodologies

Methodologies includes the algorithm framework that is used in this study. Specifically, it introduces the image pre-processing and segmentation based on the GMM algorithm, also the three deep CNN frameworks as well as the transfer learning scheme.

3.1 Image Pre-processing and Segmentation

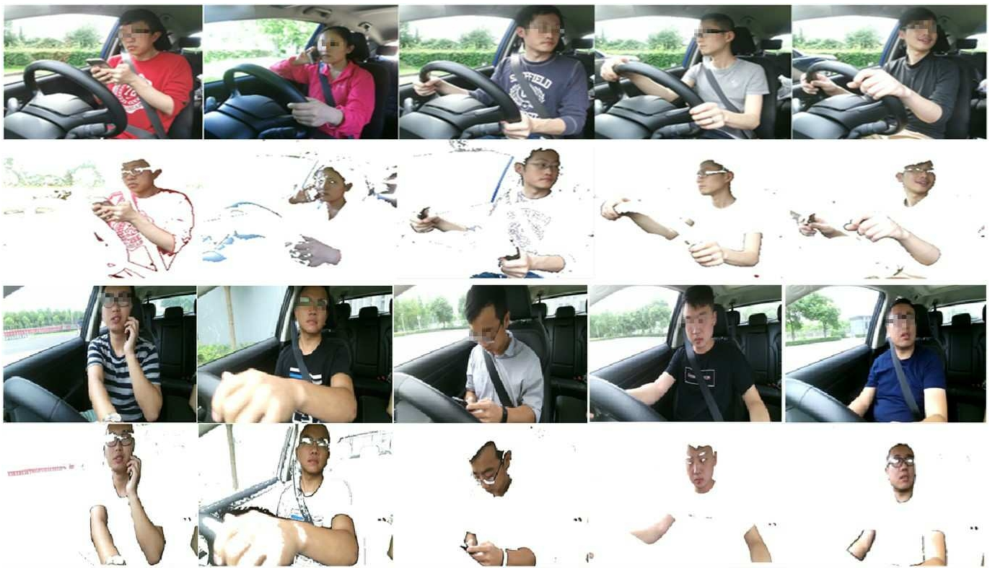

Figure 3.1: Illustration of the raw images and segmented images.

The original images are stored in the format of 640ˆ360ˆ3 . The raw images are cropped to speed up the CNN training process and increase the classification accuracy. After the raw images are cropped, these images are transformed into the size of 227ˆ227ˆ3 to satisfy the input requirement of the AlexNet and 224ˆ224ˆ3 for the GoogLeNet and ResNet, respectively. Then, the GMM algorithm is applied to segment the images and extract the driver body region from the background. GMM is an unsupervised machine learning method, which can be used for data clustering and data mining. It is a probability density function that is represented by a weighted sum of sub- Gaussian components. One of the advantages of using GMM to unsupervised segment the images is it requires no manual labeling and can be flexible to modify the model by adjusting the cluster centers. To train a GMM-based segmentation model, 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.

Driver head and body region can be identified with the GMM segmentation method. The driver body region can be determined based on a set of pre-defined points which are located around the drivers head position. The points around the head position and the corresponding label will be used to indicate the driver regions.

3.2 Model preparation and Transfer Learning

3.2.1 AlexNet Model

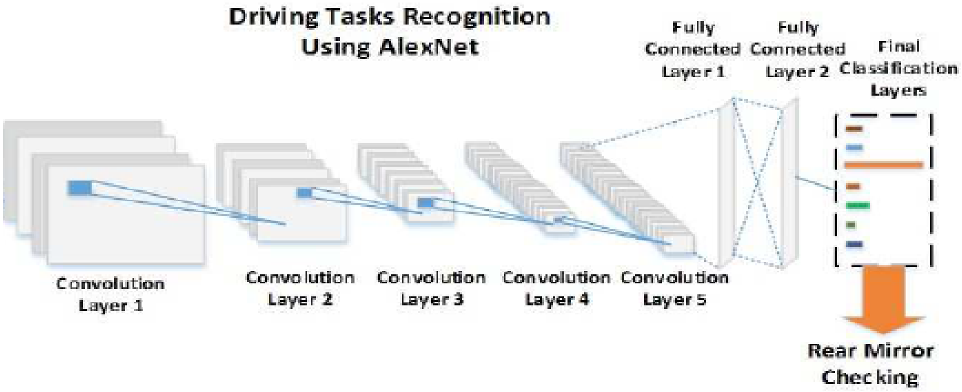

Figure 3.2: AlexNet.

The model was trained for the classification of 1000 categories in the ImageNet dataset. There are five convolutional layers and three fully connected neural network layers with non-linearity and pooling layers between the convolutional layers. In total, AlexNet contains 60 million parameters and 650,000 neurons.

3.2.2 GoogLeNet Model

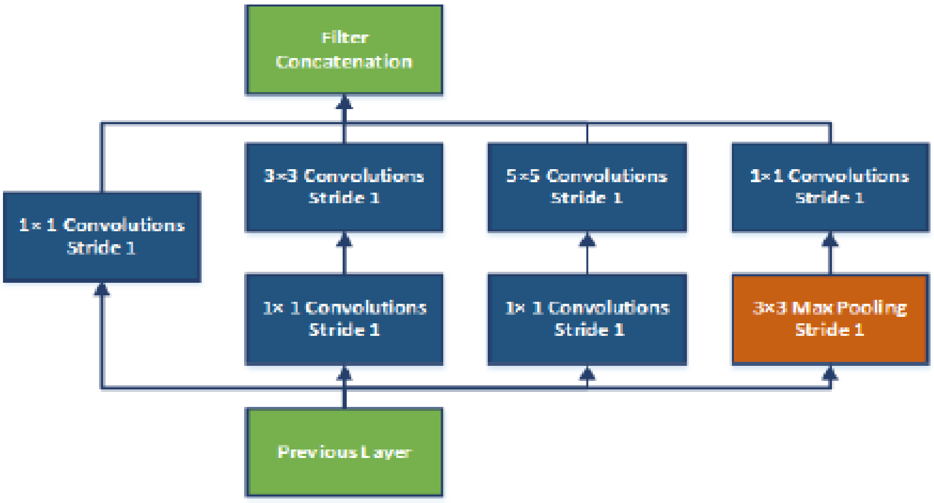
Filter Concatenation
1× 1 Convolutions Stride 1
3×3 Convolutions Stride 1
5×5 Convolutions Stride 1
1×1 Convolutions Stride 1
1× 1 Convolutions Stride 1
1× 1 Convolutions Stride 1
3x3 Max Pooling Stride 1
Previous Layer

Figure 3.3: Inception layer of GoogLeNet.

GoogLeNet is another deep CNN model, which is significantly deeper than the AlexNet, and it achieved more accurate classification results on the ImageNet dataset. Despite the model depth, the main contribution of GoogLeNet is the utilization of Inception architecture. The most common ways of increasing CNN model performance are to improve the network size (either the depth or the width of the model). However, it gives rise to the requirement for larger scale dataset and more computational burden. Based on this, the Inception layers was introduced into the CNN model to increase the sparsity among the layers, and reduce the number of parameters. Each Inception layer consists of six basic convolution filters and one max pooling filter. With different scales, the parallel-arranged convolutional filters will have more accurate detailing and a broader representation for the information from previous layers.

3.2.3 ResNet Model

Kaming, et al. introduced a novel deep CNN model, namely, Residual Networks (ResNet) to enable the construction of deeper convolutional neural networks. By introducing the residual learning scheme, the ResNet achieved the first place on the ILSVRC 2015 classification competition and won the ImageNet detection, ImageNet localization, COCO detection 2015, and COCO segmentation.

As in Fig 3.4 the underlying mapping function for the basic residual block can be assumed as H(x).The x represents the inputs to the first layer. The residual network supposes an explicit residual mapping function F(x). The core idea behind the residual network block is that although both H(x) and the F(x)+x mapping is able to approximate the desired functions asymptotically, it is much easier to learn the mapping of F(x)+x. The added layers through the shortcut connection are the identity mapping.

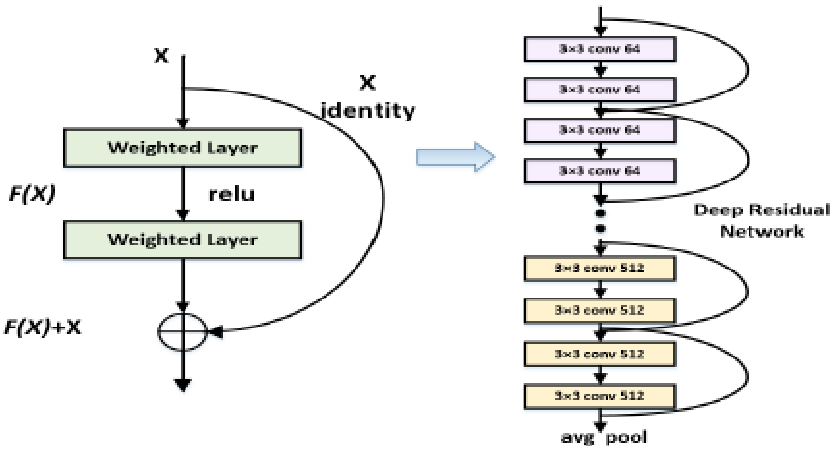

Figure 3.4: Residual learning block and deep residual network.

3.2.4 Transfer Learning

In general, large-scale annotated datasets are not always available for specific tasks. Therefore, the common ways to use the pre-trained deep CNN model are either treating the model as a fixed feature extractor without tuning the model parameters or fine-tune the pre-trained model parameters with a small-scale dataset. In this study, the CNN models will be used in the second manner, which is to fine tune the last few layers of the models with the driver behavior dataset. Since the original models are trained to classify the 1000 categories, the last few layers have to be modified so that the models can satisfy the seven objects or the binary classification task. The basic structure and properties of the convolutional layers is remained so that these layers can keep their advantages in the feature extraction and representation. Meanwhile, the knowledge that learned from the large-scale ImageNet dataset can be transferred to the driver behavior domain. A small initial learning rate is selected to slow down the updating rate of the convolutional layers.

Chapter 4

Results

Table 4.1: Classification results for driving tasks recognition using AlexNet

No.	T1	T2	T3	T4	T5	T6	T7	Ave
D1	0.825	0.929	0.011	0.225	0.840	1.0	0.972	0.771
D2	0.875	0.234	0.571	0.229	0.516	0.928	0.836	0.813
D3	0.564	0.684	0.0	0.711	0.747	0.983	0.983	0.908
D4	0.825	0.469	0.927	0.399	0.0	0.958	0.994	0.786
D5	0.797	0.20	0.10	0.843	0.60	0.959	0.996	0.843
D6	0.957	0.928	0.852	0.977	0.783	0.926	0.999	0.928
D7	0.993	0.921	0.915	0.951	0.913	0.290	0.981	0.878
D8	0.990	0.989	0.417	1.0	0.991	0.996	0.736	0.880
D9	0.353	0.994	0.229	0.813	1.0	0.982	0.979	0.752
D10	0.528	0.724	0.447	0.798	0.274	1.0	0.995	0.684
Mean	0.786	0.869	0.545	0.802	0.771	0.932	0.945	0.816

Table 4.2: Classification results for driving tasks recognition using GoogleNet

No.	T1	T2	T3	T4	T5	T6	T7	Ave
D1	0.917	0.619	0.0	0.325	0.433	1.0	0.968	0.768
D2	0.892	0.362	0.0	0.042	0.230	0.784	0.815	0.767
D3	0.883	0.563	0.0	0.073	0.840	1.0	0.994	0.739
D4	0.740	0.453	0.848	0.986	0.758	0.663	1.0	0.755
D5	0.970	0.20	0.233	0.325	0.078	0.959	0.988	0.799
D6	0.951	0.966	0.807	0.936	0.967	0.075	1.0	0.829
D7	1.0	0.886	0.436	0.990	0.890	0.248	0.963	0.737
D8	0.301	0.995	0.178	1.0	1.0	0.990	0.998	0.789
D9	0.562	0.245	0.949	0.997	1.0	0.990	0.843	0.792
D10	0.990	1.0	1.0	0.685	0.882	0.012	1.0	0.810
Mean	0.835	0.766	0.648	0.796	0.819	0.678	0.948	0.786

Table 4.3: Classification results for driving tasks recognition using ResNet50

No.	T1	T2	T3	T4	T5	T6	T7	Ave
D1	0.944	0.389	0.120	0.125	0.219	1.0	0.963	0.746
D2	0.872	0.284	0.0	0.729	0.066	0.918	0.926	0.921
D3	0.919	0.938	0.195	0.040	0.814	0.998	0.993	0.753
D4	0.975	1.0	0.924	0.514	1.0	0.639	0.882	0.801
D5	0.907	0.255	0.133	0.874	0.473	0.930	0.996	0.856
D6	0.790	0.992	0.941	0.791	0.504	0.509	0.985	0.750
D7	0.996	0.857	0.629	0.922	0.950	0.301	0.973	0.786
D8	0.528	0.567	0.192	0.641	0.988	0.944	0.715	0.638
D9	0.346	0.245	0.713	0.997	0.735	0.693	0.829	0.655
D10	0.002	0.999	0.058	0.991	0.782	0.219	1.0	0.589
Mean	0.728	0.652	0.391	0.662	0.653	0.715	0.926	0.749

The seven driving-related tasks are ordered as normal driving, right mirror checking, rear mirror checking, left mirror checking, using radio/video device, texting, and answering mobile phone. Table 4.1, Table 4.2, and Table 4.3 illustrate the classification results of the seven tasks based on AlexNet, GoogLeNet, and ResNet, respectively. T1 to T7 represents the seven tasks and D1 to D10 indicates the ten different drivers. The models are trained with MATLAB Deep Learning toolbox and evaluated using the leave-one-out (LOO) cross-validation method. To get the activity identification results for each driver, the images from one driver are used as testing images, whereas the rest images of the nine drivers are used for training.

As shown in Table 4.1, the general identification accuracy for the segmentation-based AlexNet achieved an average of 81.4% accuracy. The raw-image based AlexNet was also tested, which achieved only 69.2% recognition accuracy. In Table 4.1, the average performance in the rightmost column is defined as the average detection results for each driver, while the mean accuracy in the bottom row represents the average detection rate for each task. The worst result happens in the rear mirror checking (T3) case for the three models. One explanation is that the rear mirror checking behavior require few body and head movement, which can be easily misclassified into the normal driving task.Table 4.2 indicates the activity classification results given by the GoogLeNet. The general detection results is similar to the results in Table 4.1 except that the overall detection accuracy for the ten drivers are slightly lower. The

GoogLeNet does not achieve better classification results than the AlexNet as it does on the ImageNet dataset. However, the classification results for the GoogLeNet trained with raw images are better than that in the AlexNet case. The general classification results for the GoogLeNet with the raw image is 74.7% accuracy, which is 5% higher than that for the AlexNet.

Table 4.3 illustrates the activity classification results given by the ResNet. Same to the GoogLeNet, the ResNet does not show its advantage on the activity classification task. Instead, the precision is the lowest among these three models. The general classification accuracy is 74.9% for the GMM-ResNet and 61.4% for the Raw image based ResNet.

Chapter 5

Conclusion

In this work, a driving-related activity recognition system based on the deep CNN model and transfer learning method is proposed. To increase the identification accuracy, the raw RGB images are first processed with a GMM-based segmentation algorithm, which can efficiently remove the irrelevant objects and identify the driver position from the background context. The classification results indicate that the segmentation contributes to a much more precise detection result than the model trained with the raw images. Another comparison is made between the transfer learning and other feature extraction methods. Finally, if using the CNN models as a binary classifier, the driver distraction detection rate can achieve 91% accuracy. In the future, the data will be further analyzed, and the model will be updated to increase the system robustness and detection accuracy. Meanwhile, the system will be tested and used for driver or passenger behavior analysis on the partially automated vehicles in the real world.

References

[1] Yang Xing, Chen Lv, and Huaji Wang., *"Driver Activity Recognition for Intelligent Vehicles: A Deep Learning Approach."*, IEEE Journal (2019), doi: https://doi.org/10.1109/TVT.2019.2908425.

[2] *"Gaussian Mixture Model"*, - https://brilliant.org/wiki/gaussian-mixture-model/.

[3] Zhaoxia Fu, Liming Wang *"Color Image Segmentation Using Gaussian Mixture Model and EM Algorithm"*, Springer, link.springer.com/chapter/10.1007/978-3-642-35286-7 9.