International Journal of Internet, Broadcasting and Communication Vol.14 No.2 119-128 (2022)
http://dx.doi.org/10.7236/IJIBC.2022.14.2.119

IJIBC 22-2-16

Efficient Driver Attention Monitoring Using Pre-Trained Deep Convolution Neural Network Models

JongBae Kim

Department of Software Engineering, Sejong Cyber University, Seoul, Korea,
jb.kim@sjcu.ac.kr

Abstract

Recently, due to the development of related technologies for autonomous vehicles, driving work is changing more safely. However, the development of support technologies for level 5 full autonomous driving is still insufficient. That is, even in the case of an autonomous vehicle, the driver needs to drive through forward attention while driving. In this paper, we propose a method to monitor driving tasks by recognizing driver behavior. The proposed method uses pre-trained deep convolutional neural network models to recognize whether the driver's face or body has unnecessary movement. The use of pre-trained Deep Convolitional Neural Network (DCNN) models enables high accuracy in relatively short time, and has the advantage of overcoming limitations in collecting a small number of driver behavior learning data. The proposed method can be applied to an intelligent vehicle safety driving support system, such as driver drowsy driving detection and abnormal driving detection.

Keywords: *Advanced Driving Assistance System, Intelligent Transportation System, Driver Monitoring System, Deep Learning, Transfer learning model*

1. Introduction

Vehicles are no longer just a means of transportation and are now becoming a living space. As such, safe driving of vehicles has become an essential element technology [1-3]. Element technologies for safe driving of vehicles include lane departure detection systems, forward collision detection systems, and rear vehicle detection systems, etc., [4-6]. Various technologies are being developed for the safety of vehicles, and even if they are installed in vehicles, drivers cannot drive without complete intervention in driving tasks. With the advent of autonomous vehicles, various vehicle driving support systems for intelligent vehicles are emerging. Various vehicle IT electronic products related to self-driving vehicle safety support are being commercially released and used to support safe driving [7-9]. An autonomous vehicle refers to a vehicle that can drive on its own without controlling a driver or passenger. Autonomous driving is classified into six stages according to the level of technology [10]. Currently, Level 2 (driver assistance) is being commercialized and mass-produced, and companies are continuously developing technologies that evolve beyond Level 3 (conditional automation). In order to achieve complete autonomous driving technology, artificial intelligence mounted on cars must be able to completely replace human driving ability.

Technology development is underway in various ways to realize complete self-driving vehicles without the intervention of drivers' driving tasks, and vehicle manufacturers around the world are competing to achieve

Manuscript Received: March. 20, 2022 / Revised: March. 23, 2022 / Accepted: March. 27, 2022
Corresponding Author: jb.kim@sjcu.ac.kr
Tel:+82-2-2204-8627, Fax: +82-2-2204-8111
Professor, Department of Software Engineering, Sejong Cyber University, Korea

complete self-driving vehicles. However, the realization of full autonomous driving of autonomous vehicles still has technical limitations, and currently, it is necessary to perform safe driving through indirect intervention of the driver while driving. As a result of analyzing the traffic accident status of some recently occurring autonomous vehicles, most accidents occurred without the driver's forward attention while activating the autonomous driving function. Self-driving vehicle manufacturers are also guiding that drivers should pay forward attention to driving tasks even if they activate self-driving functions in safety precautions. Traffic accidents of autonomous vehicles as shown in Figure 1 occur in cases such as misrecognition of the trucks ahead [11, 12].

Figure 1. Traffic accident due to misrecognition of self-driving vehicles

Figure 2 shows the accident rate according to the activation of the automatic driving function suggested by the Tesla manufacturer [13]. As shown in Figure 2, the actual accident frequency is slightly lower with the automatic driving function than the driver's manual driving. However, the data only compares the accident rate on the highway, and if the results of city road driving are included, the accident rate due to the automatic driving function will be high. Nevertheless, it can be seen that safe driving support is possible by activating the automatic driving function of the vehicle, but it can be seen that the driver must still have a duty of forward attention while driving.

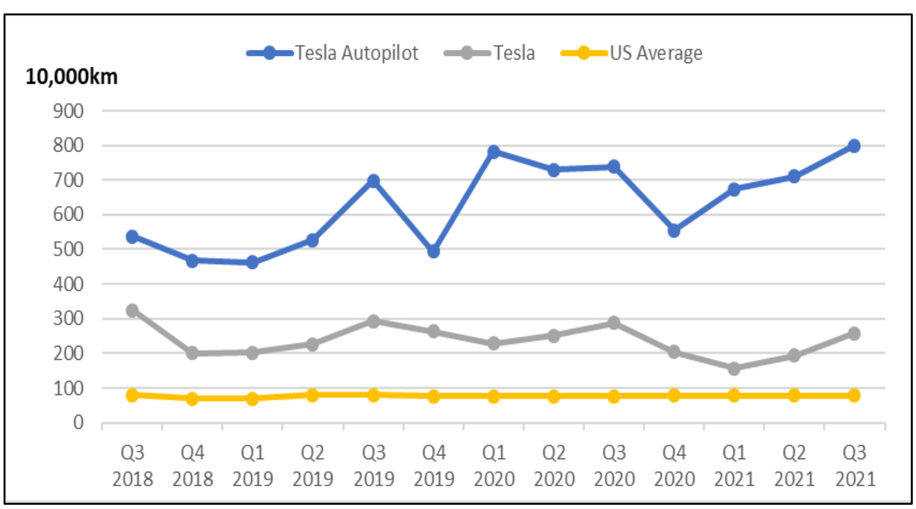

Figure 2. Comparison of the average km of one traffic accident in the United States

Due to the development of safe driving support technologies for vehicles and the expansion of traffic safety support infrastructure facilities worldwide, the incidence of vehicle traffic accidents is decreasing. Traffic accidents are also decreasing in Korea, but the number of deaths and injuries from traffic accidents still exceeds 200,000 every year. Figure 3 shows the current status of traffic accidents from 2018 to 2020 extracted from the traffic accident analysis system of the Korea Highway Traffic Authority and traffic accidents caused by violations of the law [14]. As shown in Figure 3, among the various causes of traffic accidents, the case caused by carelessness of car drivers was analyzed the highest. In Figure 3, the violation of the law, which has the highest frequency of accidents among traffic accidents, is the failure of the driver's duty to drive safely and securing the safety distance of the vehicle. Not all traffic accidents were caused by driver's negligence, but if the driver actively manipulated the car while driving, the frequency of traffic accidents could be reduced.

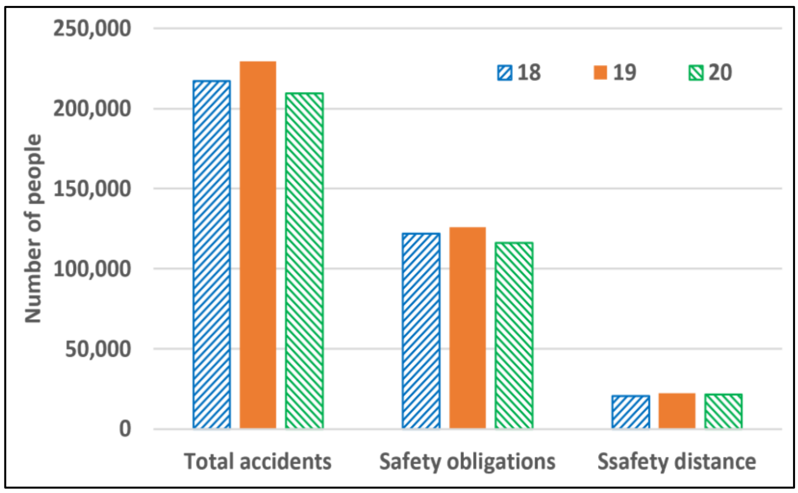

Figure 3. Number of traffic accidents by year and status of violation of laws (Korea traffic accident analysis system)

Therefore, this paper proposes a method for monitoring the presence or absence of abnormal behavior other than safe driving behavior of drivers while driving. The driver's driving work images are collected from the camera installed in the vehicle, labelled manually, and then learned using a deep convolution neural network (DCNN) [15,16]. The driver's seat of the vehicle is a narrow space, and the driver's movement while driving is relatively limited. Therefore, although the size of the driver's behavior is small, there are a wide variety of movements. For this reason, there is a difficulty in learning the driver's behavior normally and abnormally. When a large amount of driving manipulation images of various drivers was collected and used for DCNN learning, the accuracy of recognizing abnormalities in driver behavior may be increased. However, in reality, there is a limit to collecting a large amount of learning data. To overcome these limitations, DCNN models learned in advance in this paper are utilized.

The concept of pre-training is not limited to the NLP field, but is used in various fields such as computer vision. The concept that emerges with pre-training is transfer learning, which focuses on storing knowledge while solving a problem, and utilizes it to solve different but related problems. If a neural network model has learned to recognize faces, it can be used to recognize eyes with it. In other words, it is used to pre-trained model that solves face recognition problems, and to solve eye recognition problems that are different but similar problems with pre-trained models. Neural networks have learning data and learning proceeds. The neural network extracts knowledge from the learning data, and the knowledge is stored as weights of the neural network. The weights of these neural networks can be extracted and transferred to other neural networks. So instead of learning new neural networks, it transfers learned characteristics. In this paper, there is also a limitation in collecting a large amount of driving work learning images of vehicle drivers. Thus, a large amount of learning data is constructed to derive driver behavior recognition results from a relatively small amount of driver task learning data using pre-trained models.

Previous researches [7-9] have proposed methods such as image-based driver face detection and eye detection, driver biometric analysis-based driver fatigue monitoring, and intensive monitoring of driving with eye tracking[17-19]. However, these methods are methods of analyzing the current state based on the extraction of the feature information of the image, and there is a limit to applying only at a specific point in time and a specific motion, so there is a limit to general application. In order to solve this problem, instead of analyzing feature information at a specific point in time, a method of extracting the most appropriate feature information using artificial intelligence technology in various scenarios using deep neural networks has recently been studied [20].

DCNN can be said to be a feature information extractor optimized from the output results provided for learning by combining functions such as multi-step feature information extraction, dimension reduction, and optimal feature information selection. In this paper, we obtain driving task images of driving drivers and label them for abnormal behavior using the ground-truth method. Then, the labelled driver driving task images are transfer-learned using the pre-trained DCNN models. The driver driving operation image is input to the DCNN models derived through such transfer learning to recognize whether the driver has abnormal behavior. Each recognition result output from each DCNN model is combined to derive a final recognition result. Safe vehicle driving will be realized by applying the proposed method to intelligent transportation systems.

2. Proposed Method

Figure 4 shows the processing flow of the learning process and the actual application process in the vehicle driving work images of the proposed method. For the learning of driving task images of pre-trained DCNN models, driving task images of a total of 4 drivers in two vehicles and driver monitoring learning images obtained from AI Hub [21] were used. CCD cameras were used for directly acquired learning images, and near-infrared cameras were used for AI Hub learning images. Several pre-trained DCNN models are used in the proposed method to learn whether or not the driver movement is abnormal in the images. Each pre-learned DCNN model is designed to have effectively recognizable neural network weight values for each class of objects to be distinguished from a large number of learning data. In our work, we leverage the advantages of neural networks of pre-trained DCNN models. As shown in Figure 4, a total of 8 pre-trained DCNN models are used to learn vehicle driver driving task images. The learning process computes the loss of the predictive labelling results output from the pre-trained DCNN model with the actual learning labels, as shown in Figure 4(a). Then, we modify the weights of the pre-DCNNs while repeating the learning to reduce the loss predicted through the learning algorithm. In this way, DCNN networks learned with driver driving task learning images are generated. Predictive label results are derived by inputting to DCNN networks learned from the input image. As shown in Figure 4(b), the predictive labels output from each DCNN and the accuracy of the verification data in the DCNN learning process are calculated as weights and recognized as driver abnormal behavior when above the threshold.

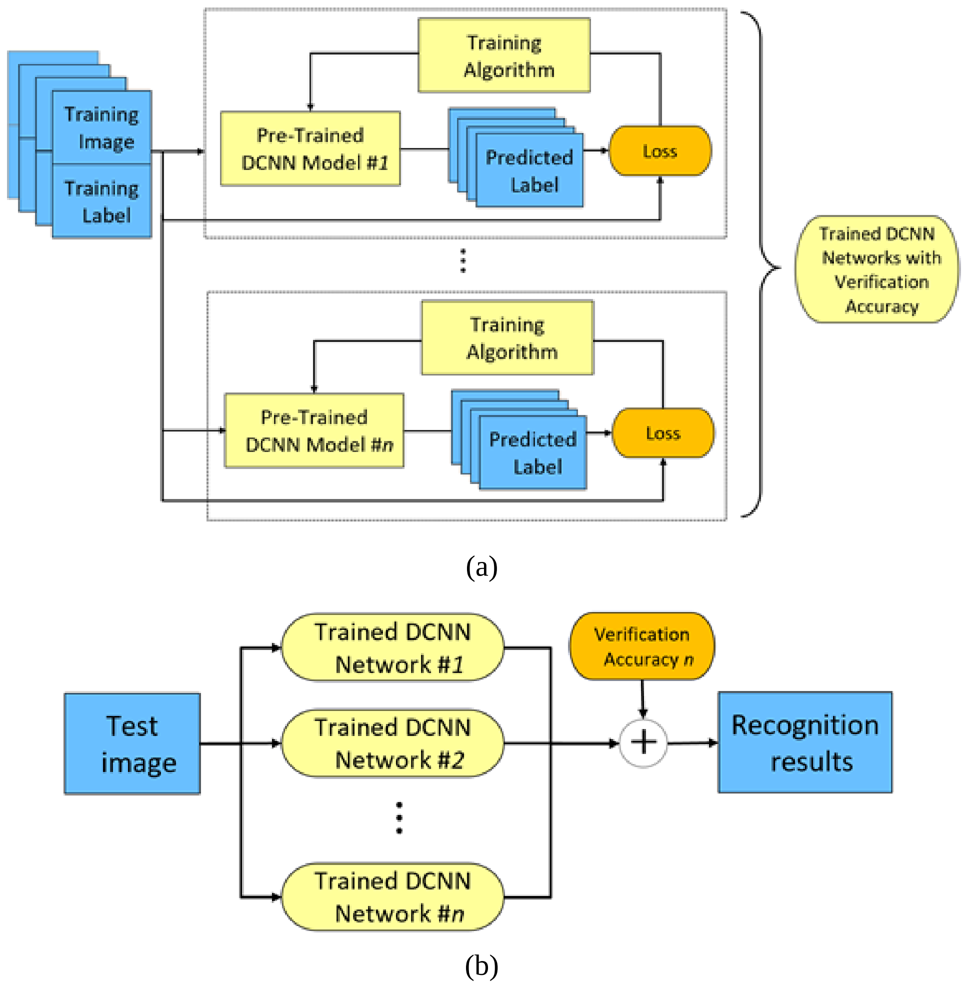

Figure 4. Flow of the proposed learning process and experimental process

In order to monitor the abnormal behavior of the driver, a large amount of driving task learning images is required. It is possible to extract and analyses unique feature information on driver behavior during driving from learning images, and set criteria for determining abnormal behavior patterns and normal behavior patterns. It is difficult to collect a large amount of learning images while driving various drivers. In addition, insufficient training data required for learning in DCNN creates an overfitting problem for the learning data given by the learning process. To solve this problem, we apply a pre-trained DCNN model. The transfer learning is a model built using a large amount of learning data, and has the advantage of being able to effectively express features even if a small amount of learning data is used. In the proposed method, we use a transfer learning generated through a pre-trained model, and then fine-tune the learning parameters through an error back-propagation method. Effective fine tuning is possible by providing better initial values than newly initialized DCNN learning with the use of pre-trained DCNN models.

Figure 5 shows the processing performed for learning driving task images. Learning images are used by normal and abnormal labels, and an image augmentation process is applied to increase the number of learning data. Augmentation of the learning image generates a larger number of learning images through color change, rotation, and pixel movement of the learning image and is used for learning. When labelling learning images, the driver's abnormal behavior is when the driver's gaze is not looking at the steering wheel or the front, but looking left, right, up and down. And, when the head, shoulders, and hands moved excessively, it was determined as abnormal behavior. However, the illumination inside the vehicle is very low in the night driving image, so there is a limit to recognizing behavior in the CCD image.

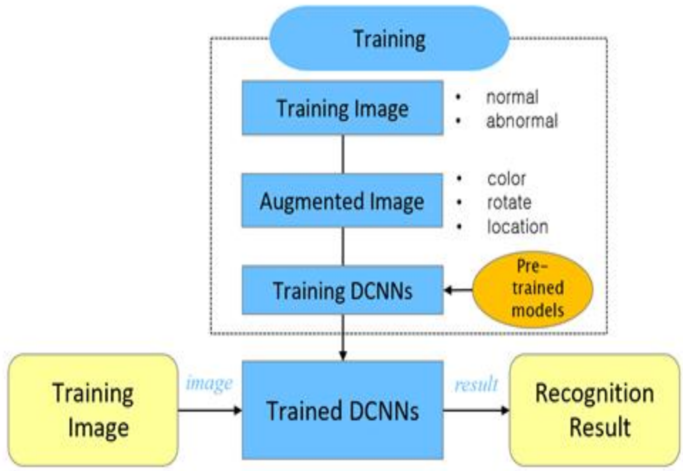

Figure 5. Flowchart of the proposed driver abnormal behavior monitoring method

3. Experimental Results

In order to obtain an image containing a driver while driving, an indoor driver image (QHD 2560×1440) was obtained using a total of two vehicle black boxes (fine view X3000 channels). At this time, the automatic color and brightness correction function of the black box was set to off. In addition, driver state information images provided by AI Hub (AI Hub, 2022) were also used as learning data. The learning images used in the experiment were collected by adjusting various angles and positions in vehicle dashboards, windows, and instrument panels. However, the image acquisition camera was acquired at an angle and position facing the driver. The PC environment used in the experiment was performed with Matlab in the MS Windows 10 OS environment (3.2GHz Hexa Core, 128GB, GPU 3060 ti 8G). Table 1 shows the driving task images of the vehicle driver used for learning the proposed method. If the steering wheel portion or the driver's hand and arm overlap more than 50% in the image collected due to steering wheel manipulation while driving the vehicle, it was excluded from the learning data labelling.

Table 1. Learning images for experiment

Experiment sets	No. of image (normal/abnormal)	Obtained Times	Locations
#1	500/300	Morning	Highway
#2	500/200	Afternoon	City
#3	500/150	Sunset	Highway
AI Hub [21]	2473/2064	-	-

Figure 6 shows images of normal and abnormal behavior during driver driving tasks used in the experiment. The experimental images used for learning were images acquired while driving the experimental driver and driver monitoring data shared by AI Hub [21].

(a) normal behavior | (b) abnormal behavior

Figure 6. Sample images used for training

In the proposed method, in order to select a transfer learning model that offers high classification accuracy and low loss rate, an experiment was conducted by combining various learning parameters for various pre-training models. There are a total of 8 transfer learning models used in the experiment, and squeezenet, googlenet, resnet18, mobilenetv2, resnet50, resnet101, inceptionv3, and inceptionresnetv2 are used [22,23]. Table 2 is the size of the input image used by the pre-trained model.

Table 2. Input image sizes of pre-trained models

DCNN Models	Input Size
Squeezent	$227 \times 227 \times 3$
Googlenet, Resnet18, Mobilenetv2, Resnet50, Resnet101	$224 \times 224 \times 3$
Inceptionv3, Inceptionresnetv2	$299 \times 299 \times 3$

Table 3 shows the results of experiments in which learning was conducted using various learning parameters in a pre-learned model. To minimize the loss of the predicted labels from the pre-trained DCNN model, we use the Stochastic Gradient Descent with Momentum optimization function in our learning algorithm to modify the weights of the DCNNs. It is necessary to set parameters of a learning algorithm for generating a neural network model for driver monitoring recognition using pre-trained DCNN models. The initial learning rate was 0.001, the learning cycle was 50, and the mini-batch size was 16 and 32, respectively. For learning using a pre-learned model, learning was conducted by setting the ratio of learning data and verification data at a ratio of 8:2.

As a result of learning, it may be possible to see a difference in verification accuracy according to the size of the mini-batch among the learning parameters. Experiments are conducted by selecting a pre-trained DCNN model with high verification accuracy for driver behavior recognition. In addition, in the case of Inceptionresnetv2 pre-trained model, an error occurred as a result of the experiment by setting the training parameter with a mini-batch size of 32. The experiment could not be conducted due to the memory excess of the GPU used in the experiment.

Table 3. Training results using pre-trained DCNN models

DCNN Models	MiniBatch Sizes	Training Loss	Validation Accuracy(%)	Validation Loss	Training Time(hh:mm:ss)
Squeezenet	16	1.7366e-5	84.4493	2.8857	42:16
Googlenet	16	2.9292e-5	85.3304	2.5611	43:50
Resnet18	16	0.0001	81.9163	2.2113	58:25
Mobilenetv2	16	1.8633e-5	82.3568	1.9090	59:52
Resnet50	16	8.7917e-5	75.5286	2.4685	58:37
Resnet101	16	5.0514e-6	71.3436	3,6183	01:09:23
Inceptionv3	16	2.7845e-5	90.7269	1.4114	01:07:43
Inceptionresnetv2	16	0.0003	80.9251	1.4983	03:34:11
Squeezenet	32	1.1365e-5	79.9339	4.52987	24:06
Googlenet	32	1.1875e-5	79.6035	3.8039	29:25
Resnet18	32	7.0778e-6	77.6211	2.8593	24:40
Mobilenetv2	32	0.0002	83.0176	2.0420	43:04
Resnet50	32	6.4593e-6	79.2731	2.3729	43:20
Resnet101	32	0.0004	83.3480	1.3845	57:44
Inceptionv3	32	0.0001	84.8899	1.6556	01:12: 29
Inceptionresnetv2	32	NA	NA	NA	NA

In the learning results of the DCNN models of Table 3, verification accuracy is used as a weight when calculating the final recognition result. In the proposed method, multi DCNNs are used to finally recognize whether the driver's driving task is abnormal behavior. As a result of the DCNN learned from the input image, a prediction label and a score are output. This score is a measure of how close the input image is to the prediction label. Therefore, each score of the driver abnormal behavior label result of each DCNN model in the input image is multiplied by the DCNN verification accuracy and normalized to a value between [0 1]. And, if the average of each normalized result is 0.5 or more, it is finally recognized as an abnormal behavior image. Figure 6 shows the results of verification accuracy and loss per learning cycle of inceptionv3 DCNN, which has the highest verification accuracy among pre-learned DNCC models.

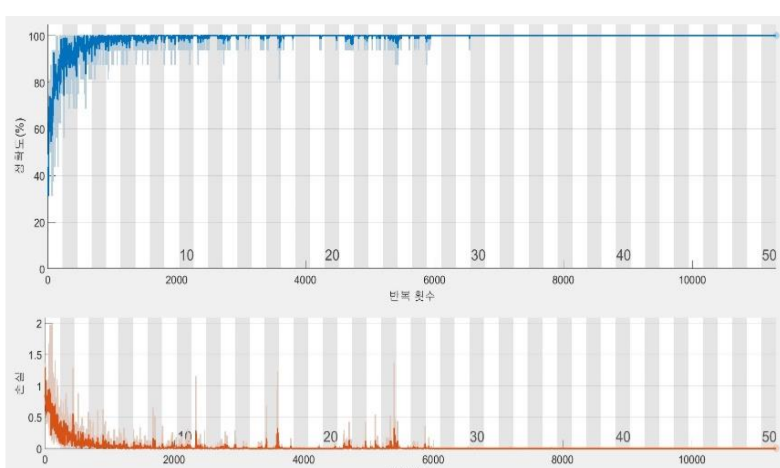

Figure 6. Training progress plots of accuracy and loss rate per epoch of inceptionv3 DCNN model

Table 4 is a confusion matrix that shows the recognition results of driver driving task images using verification data using the learned inceptionv3 DCNN model. In the case of normal behavior of the driver, recognition accuracy was high because most of them stared straight ahead. However, in the case of abnormal behavior of the driver (phone call, yawning, hand movement, excessive head movement, etc.), the recognition accuracy was relatively low. The cause is a lack of learning data for various driver behaviors.

Table 4. Confusion matrix results of driver behavior recognition using Inceptionv3 DCNN

Class	Normal	Abnormal
Normal	92%	8%
Abnormal	13%	87%

As a result of experimenting with the proposed method, on average, about 86%. The above accuracy was presented, and the recognition speed took about 1.2 seconds. Figure 7 shows the results of determining whether there is an abnormal behavior in the driver's driving image acquired in a new environment. In the case of misdiagnosed case, the forward gaze was focused, but the situation was not used for learning, such as errors due to cell phone gaze or spectacle light reflection, and misidentified as normal behavior.

Figure 7. Results of error in recognizing the behavior of the driver in the proposed method. (a) It's normal, but it's recognized as abnormal, (b) It's abnormal, but it's recognized as normal

4. Conclusions

In this paper, we propose a monitoring method to determine whether the driver's attention is focused on driving. In the proposed method, the transfer leaning of normal and abnormal driver behavior is performed in advance, and a driver behavior discrimination model is generated using a deep neural network. In order to select the optimal transfer learning model, various learning parameters were applied to the experiment to select a transfer learning model with the accuracy of determining the abnormal behavior of the driver and minimizing the learning loss. If the driver's behavior while driving the vehicle occurs within a limited range, there was a case of misinterpretation of abnormal behavior due to the difference in illuminance due to the image acquired from the vehicle inside. In future research, we intend to study a method to additionally utilize the driver's posture estimation characteristic information in the input image to improve accuracy.

Acknowledgement

An extended version of this paper has been published at the ICESI 2022 conference[24], and this work was supported by the NRF grant funded by the Korea government (MSIT)(NRF-2020R1F1A106890011)

References

[1] Y. H. Kim, "V2X Technology Trends for Next-Generation Mobility," *International Journal of Internet, Broadcasting and Communication*, vol. 12, no. 1, pp.7-13, 2020.

[2] B. C. Juan, D. C. Rufas, et al., " Resource-Constrained Machine Learning for ADAS: A Systematic Review," *IEEE Access*, vol. 8, pp.40573-40598, 2020.

[3] V. K. Kukkala, J. Tunnell, S. Pasricha, and T. Bradley, "Advanced Driver-Assistance Systems: A Path Toward Autonomous Vehicles," *IEEE Consumer Electronics Magazine*, vol. 7, no. 5, pp.18-25, 2018.

[4] S. Raviteja and R. Shanmughasundaram, "Advanced Driver Assistance System (ADAS)," *International Conference on Intelligent Computing and Control Systems*, pp.737-740, 2018.

[5] S. S. Sarikan and A. M. Ozbayoglu, "Anomaly Detection in Vehicle Traffic with Image Processing and Machine Learning," *Procedia Computer Science*, vol. 140, pp.64-69, 2018.

[6] N. Albarella, F. Masuccio, L. Novella, M. Tufo, and G. Fiengo, "A Forward-Collision Warning System for Electric Vehicles: Experimental Validation in Virtual and Real Environment," *energies*, vol. 14, no. 4872, pp.1-12, 2021.

[7] J. B. Carazo, D. C. Rufas, and E. Biemp, "Resource-Constrained Machine Learning for ADAS: A Systematic Review," *IEEE Access*, vol. 8, pp.40573-40598, 2020.

[8] C. Guo, J. Meguro, Y. Kojima, and T. Naito, "A Multimodal ADAS System for Unmarked Urban Scenarios Based on Road Context Understanding," *IEEE Trans on ITS*., vol. 16, no. 4, pp.1690-1704, 2015.

[9] Z. Sun, G. Bebis, and R. Miller, "On-road vehicle detection: A review," *IEEE Trans. Pattern Anal. Mach. Intell*., vol. 28, no. 5, pp.694-711, 2006.

[10] https://www.sae.org/blog/sae-j3016-update

[11] https://taiwanenglishnews.com/tesla-on-autopilot-crashes-into-overturned-truck/

[12] https://globalnews.ca/news/8286090/tesla-home-crash-ottawa/

[13] https://www.wheelsjoint.com/tesla-autopilot-accident-rate-drops-by-half-in-first-quarter-of-2020/

[14] http://taas.koroad.or.kr/

[15] Y. Cui, R. Chen, W. Chu, L. Chen, D. Tian, Y. Li, and D. Cao, "Deep Learning for Image and Point Cloud Fusion in Autonomous Driving: A Review," *IEEE Transactions on Intelligent Transportation Systems*, vol. 23, no. 2, pp.722-739, 2022.

[16] T. W. Kwon, J. Y. Lee, and K. D. Jung, "Design of Falling Recognition Application System using Deep Learning," *International Journal of Internet, Broadcasting and Communication*, vol. 12, no.2, pp.120-126, 2020.

[17] B. Alshaqaqi, et al., Vision based system for driver drowsiness detection," *Int. Symp. on Programming and Systems*, pp.103-108, 2013.

[18] W. J. Chang, L. B. Chen, and Y. Z. Chiou, "Design and Implementation of a Drowsiness-Fatigue-Detection System Based on Wearable Smart Glasses to Increase Road Safety," *IEEE Trans. on CE*., vol. 64, no. 4, pp.461-469, 2018.

[19] Y. L. Yang, F. Chang, and Y. Lu, "Driver Gaze Zone Estimation via Head Pose Fusion Assisted Supervision and Eye Region Weighted Encoding," *IEEE Trans. on. CE*., vol. 67, no. 4, pp.275-284, 2021.

[20] R. Girshick, J. Donahue, T. Darrell, and J. Malik, "Rich Feature Hierarchies for Accurate Object Detection and Semantic Segmentation," *Proc. of the IEEE Conf. on CVPR*., pp.580-587, 2014.

[21] https://aihub.or.kr

[22] O. Russakovsky, J. Deng, H. Su, et al., “ImageNet Large Scale Visual Recognition Challenge,” *International Journal of Computer Vision*, vol. 115, no. 3, pp.211–252, 2015.

[23] B. Zhou, A. Lapedriza, A. Torralbam, and A. Oliva, "Places: An image database for deep scene understanding," arXiv preprint arXiv:1610.02055, 2016.

[24] J. B. Kim, “Deep learning-based driver attention recognition for driver's abnormal behavior monitoring,” *Int. conf. on electric-vehicle, smart gird, and information technology*, p02-1, 2022.