2021 The 21st International Conference on Control, Automation and Systems (ICCAS 2021)
Ramada Plaza Hotel, Jeju, Korea, Oct. 12∼15, 2021

Distracted Driver Recognizer with Simple and Efficient Convolutional Neural Network for Real-time System

Duy-Linh Nguyen[1], Muhamad Dwisnanto Putro[2] and Kang-Hyun Jo[3*]

[1,2,3]Department of Electrical, Electronic and Computer Engineering, University of Ulsan, Ulsan, Korea
(ndlinh301@mail.ulsan.ac.kr, dputro@mail.ulsan.ac.kr, acejo@ulsan.ac.kr) * Corresponding author

Abstract: The traffic accident is a big problem in the world and it is happening every day. One of the main causes is distracted driving. Those are the actions of the driver when they are not focusing on driving on the road such as using the cellphone, drinking, makeup, talking to others, etc. For the purpose of warning drivers, this paper proposes a distracted driver recognizer with a simple and efficient Convolutional Neural Network (CNN). The evaluation results on the State Farm Distracted Driver Detection dataset with ten activities achieved an accuracy of 99.51% and on video with the latency allowed for deployment in the real-time system based on a low-computation device.

Keywords: Convolutional Neural Network, Distracted driver recognizer, Driver Surveillance system, Image classification.

1. INTRODUCTION

According to the World Health Organization, every year about 1.3 billion people die and 30 to 50 million people are injured in traffic accidents [1]. There are many reasons leading to this situation, but one of the main causes is distracted driving. This report also outlines several types of driver distraction, particularly the impact of technology devices such as cell phones, radio sound systems and navigation devices. Drivers who do not use technology devices while driving can reduce the rate of traffic accidents by approximately four times. There are several definitions of distracted driving. Distracted driving is any activity that diverts attention from driving, including talking or texting on the phone, eating and drinking, talking to people in vehicle, operating with the stereo, entertainment or navigation system [2]. In another definition, anything that takes attention away from driving can be a distraction. Accordingly, distracted driving is divided into three groups: Visual (taking the eyes off the road), Manual (taking the hands off the wheel), Cognitive (taking the mind off driving) [3]. From the above observations, the governments of many countries have tightened the supervision and penalty of distracted driving behaviors. In addition, researchers have focused on developing devices used to warn of distracted driving and built into modern cars [4], [5], [6]. In the group of visual devices, the devices analyze the state of the eyes and the posture of the driver's head. It uses sensor systems mounted in the vehicle or mounted on the driver's body. In addition, it also uses a system of information cameras to predict the driver's condition. With the manual group, the devices estimate the actions of the hands on the steering wheel. This group uses hands tracking and estimating the driver's posture. Regarding cognitive group, the main method is to predict the psycho-physiological state of the driver. These methods require expensive, complex devices with high precision and are embedded in wearable devices. This can cause inconvenience to the driver when operating. In order to simplify the use and save the installation cost but still maintain the high accuracy of the warning, this paper focuses on developing the driver activities recognizer based on a simple and efficient convolutional neural network. This network exploits basic features such as Convolutional layers, pooling layers, optimization methods, and especially the Global Average Pooling technique to replace Fully Connected layers to reduce computation cost. As the high accuracy and the allowable delay in video testing, this network can be applied to the real-time driver distraction alarm system.

The main contributions of this paper are as follows:
- Proposed a simple and efficient Convolutional Neural Network for driver behavior classification.
- Developed a driver distraction warning system that can be deployed on low-profile personal computer systems and low-computing devices.

2. RELATED WORK

In this section, the paper will introduce some methodologies applied to driver distraction warning systems, focusing on detecting driver's behavior through images. These methodologies are divided into classical image processing and convolutional neural network-based (CNN-based) methodologies.

2.1 Classical image processing methodologies

Several classical methods commonly used to classify driver activity patterns are Support vector machines (SVM), AdaBoost, Hidden Markov Models, and Hidden CRF. In [7] uses Support vector machines to predict mobile phone use while driving in image dataset collected from cameras on highways and traffic lights. This method is also used to classify a driver's mobile phone use based on relative hand, mobile device and face position [8]. [9] combines AdaBoost classifier and Hidden Markov Models to classify images obtained from

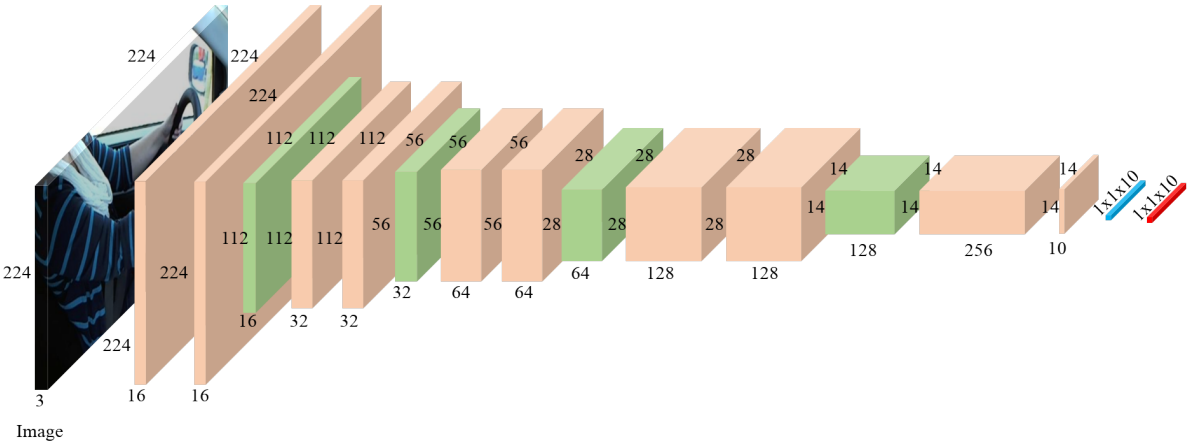

Fig. 1. The architecture of proposed network.

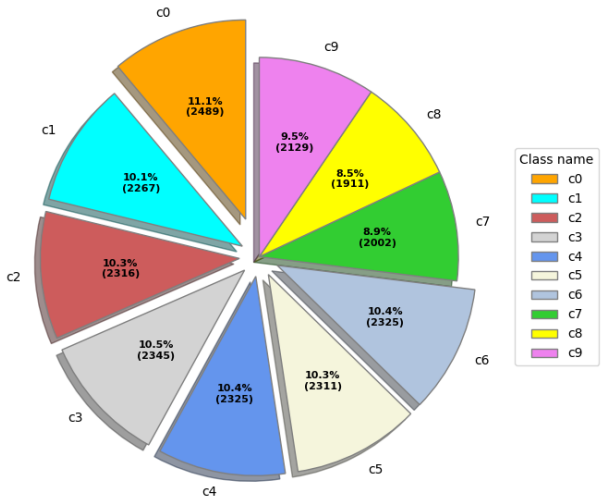

Fig. 2. Dataset distribution.

Kinect's RGB-D sensors with actions that simulate cell phone use while driving. In addition, [10] recognizes phone usage through facial, mouth and hand features based on the Hidden Conditional Random Fields (Hidden CRF) method. For multi-action classification, [11] proposes Contourlet Transform combined with Random Forests classification. Then, the authors evaluated and compared with other methods such as K-Nearest Neighbors (KNN) and Multilayer Perceptron (MLP).

2.2 CNN-based methodologies

The explosive growth of machine learning and the popularity of convolutional neural networks accelerate the development of applications in the field of computer vision. Driver behavior monitoring and alerting systems are also built based on convolutional neural networks for widespread deployment and improved accuracy. For feature extraction and classification of driver behaviors, popular convolutional neural networks can be used such as LeNet [12], VGG [13], ResNet [14], Inception [15], Xception [16], DenseNet [17], MobileNet [18], NASNet-Mobile [19] depending on the purpose and dataset. Besides, object detection and segmentation methods are also exploited or combining different methods to improve feature extraction and action detection. With object detection, [20] is interested in predicting cell-phone usage and hands on steering wheel detection based on Faster-RCNN network. In [21], the image is segmented into three areas: the steering wheel, gear, and dashboard. Next, a classification network is used to predict the appearance of the driver's hand on segmented image regions. In another approach, [22] has proposed light-weight convolutional neural networks to detect and classify eye states. From there, apply to the driver's drowsiness warning system. Other research has regarded 3D-CNN to extract the head postures and use gradient boosting machine for classification to distinguish head postures while driving [23].

3. METHODOLOGY

From the above analysis, this paper proposes a simple and efficient convolutional neural network to classify ten driver activities in State Farm Distracted Driver Detection Dataset.

3.1 Network Architecture

The detailed description of the proposed network architecture is shown in Fig. 1. Like other classifier, this network is built based on sequential layers such as convolution layer, average pooling layer, BatchNormalization (BN), Dropout technique, followed by Global Average Pooling layer. The output is calculated by the Softmax function to get ten probabilities corresponding to ten actions of the driver in the dataset.

Specifically, the network uses four convolutional blocks. Each block uses two convolutions with kernel sizes of 7×7, 5×5, 3×3 respectively, followed by an Average Pooling layer and a ReLU activation function. In which, the last two convolution blocks use a kernel size of 3×3. The use of convolution operation with a large kernel size increases the reception field, helping the network capture the best information. The feature extractor ends with a convolution layer with a kernel size of 3×3. From the input image of size 224×224, going through the feature extractor to obtain a feature map of size 14×14.

Then, the Global Average Pooling layer further reduces the feature map size to 1×1 with ten channels corresponding to ten class names. The difference of this network from other popular classification networks is that it makes use of the Global Average Pooling layer as an alternative to the Fully Connected Layer. This significantly reduces network parameters but still ensures the preservation of important information. On the other hand, it also increases the ability to prevent the overfitting issue. Finally, the network uses the Softmax activation function to calculate the predicted probabilities for the ten classes. Table 1 shows the detailed description of proposed architecture in each block.

3.2 Loss Function

During training, the network uses the Cross-Entropy loss function to calculate the loss of the entire network.

4. EXPERIMENTAL RESULTS

4.1 Dataset

The proposed network was trained and evaluated on the State Farm Distracted Driver Detection dataset from the competition of the Kaggle website [24]. This dataset contains 22,424 color images with the resolution of 640×480 and they are divided into ten classes corresponding to folders marked c0 to c9. The ten classes in this dataset are safe driving (c0), texting - right (c1), talking on the phone - right (c2), texting - left (c3), talking on the phone - left (c4), operating the radio (c5), drinking (c6), reaching behind (c7), hair and makeup (c8), talking to passenger (c9). Fig. 2. shows the State Farm Distracted Driver Detection dataset distribution. The dataset is split into 80% (17,939 images) for the training set and 20% (4,485 images) for the evaluation set.

Table 1. Detailed description of proposed architecture.

Layer (type)	Output Shape
Block1, $7 \times 7 \times 16$	$224 \times 224 \times 16$
Pool1, $2 \times 2 \times 16$	$112 \times 112 \times 16$
Block2, $5 \times 5 \times 32$	$112 \times 112 \times 32$
Pool2, $2 \times 2 \times 32$	$56 \times 56 \times 32$
Block3, $3 \times 3 \times 64$	$56 \times 56 \times 64$
Pool3, $2 \times 2 \times 64$	$28 \times 28 \times 64$
Block4, $3 \times 3 \times 128$	$28 \times 28 \times 128$
Pool4, $2 \times 2 \times 128$	$14 \times 14 \times 128$
Conv9, $3 \times 3 \times 256$	$14 \times 14 \times 256$
Conv10, $3 \times 3 \times 10$	$14 \times 14 \times 10$
GlobalAveragePooling	$1 \times 1 \times 10$
Softmax	$1 \times 1 \times 10$

4.2 Experimental setup

In order to train the classification network, this experiment uses a GeForce GTX 1080Ti GPU. In addition, this GPU is also used for real-time video evaluation and another Intel Core I7-4770 CPU @ 3.40 GHz, 8GB of RAM (PC). The network is trained with several basic configurations such as Adam optimization method, the learning rate is 10^{-4}, batch size of 16, and the number of epoch is 200. On the other hand, this paper also uses data augmentation methods to improve accuracy and prevent overfitting. These methods include flip, shift, random rotation, random brightness, random zoom.

4.3 Experimental result

First, the network was trained and evaluated on a GeForce GTX 1080Ti GPU. As a result, the network achieves 99.51% of the accuracy on the evaluation dataset with a very small number of network parameters, it is only 651,418. This result can compete with other popular classification networks in both accuracy and network parameters. Especially when choosing a solution to build a real-time warning system and deploy it on low-computing devices. The table shows the comparison result of the proposed network with popular classification networks on the State Farm Distracted Driver Detection Dataset. The qualitative result of network evaluation on State Farm Distracted Driver Detection Dataset shows as in Fig. 3. In other experiments, the network training results are evaluated with real-time video on both GPU and CPU with a resolution of 640×480. The evaluation results show that, with the network GPU, it can reach over 60 frames per second (FPS) and over 7 FPS with the CPU. With its light-weight and low latency, this network is capable to apply on mobile devices and low computing devices. With the results displayed on the confusion matrix in Fig. 4., it can be seen that the network is equally predictable with the classes in the dataset. For actions that do not involve devices such as safe driving, hair and makeup and talking to passenger, this network predicts weaker than other actions but with a small difference.

Table 2. The comparison result of the proposed network with popular classification networks on State Farm Distracted Driver Detection Dataset.

Network	Accuracy (%)	Paramenters
Proposed	99.51	651,418
MobileNet	99.67	4,288,714
VGG13	99.82	5,228,618
NASNetMobile	99.62	5,362,334
DenseNet121	99.62	8,097,354
VGG16	99.97	15,250,250
VGG19	99.24	20,559,946
Xception	99.02	22,969,906
ResNet	99.84	23,772,042
InceptionV3	99.89	23,911,210
LeNet	99.79	78,432,080

5. CONCLUSION

This study proposes a distracted driver recognizer based on a simple, light-weight, and efficient network.

Fig. 3. The qualitative result of network testing on State Farm Distracted Driver Detection Dataset.

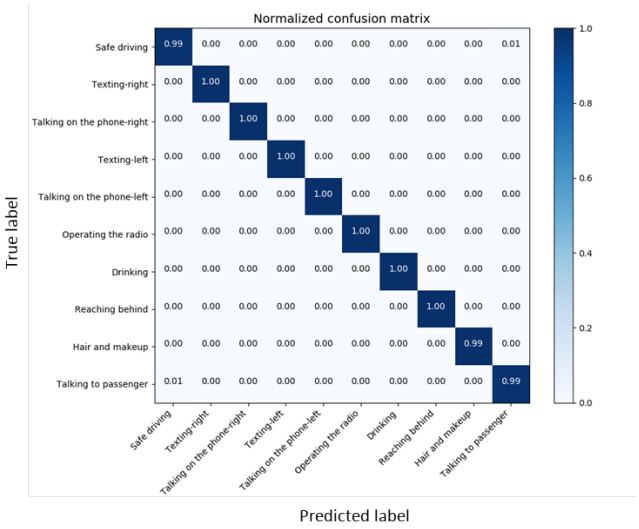

Fig. 4. Confusion matrix.

This network is built with basic layers like Convolution layer, Average Pooling layer, Global Average Pooling layer, and ends with Softmax function to calculate predicted probabilities for ten action classes. With the simplicity and minimalism of the network parameter, it can be deployed on low computing devices to build a real-time alarm system. In the future, the research continues to be improved with a human body detector to improve accuracy and reduce latency when operating in real-time.

ACKNOWLEDGEMENT

This work was supported by the National Research Foundation of Korea(NRF) grant funded by the government(MSIT).(No.2020R1A2C200897212)

REFERENCES

[1] "Road traffic injuries." https://www.who.int/news-room/fact-sheets/detail/road-traffic-injuries. Accessed: 2021-07-14.

[2] "Distracted driving." https://www.nhtsa.gov/risky-driving/distracted-driving. Accessed: 2021-07-14.

[3] "Distracted driving." https://www.cdc.gov/transportationsafety/distracted-driving. Accessed: 2021-07-14.

[4] A. Eriksson and N. A. Stanton, "Takeover time in highly automated vehicles: Noncritical transitions to and from manual control," *Human Factors*, vol. 59, no. 4, pp. 689–705, 2017. PMID: 28124573.

[5] H. M. Eraqi, M. N. Moustafa, and J. Honer, "End-to-end deep learning for steering autonomous vehicles considering temporal dependencies," *CoRR*, vol. abs/1710.03804, 2017.

[6] "Static free space detection with laser scanner using occupancy grid maps," *CoRR*, vol. abs/1801.00600, 2018. Withdrawn.

[7] Y. Artan, O. Bulan, R. P. Loce, and P. Paul, "Driver cell phone usage detection from hov/hot nir images," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, June 2014.

[8] R. A. Berri, A. G. Silva, R. S. Parpinelli, E. Girardi, and R. Arthur, "A pattern recognition system for detecting use of mobile phones while driving," in *2014 International Conference on Computer Vision Theory and Applications (VISAPP)*, vol. 2, pp. 411–418, IEEE, 2014.

[9] C. Craye and F. Karray, "Driver distraction detection and recognition using rgb-d sensor," *arXiv preprint arXiv:1502.00250*, 2015.

[10] X. Zhang, N. Zheng, F. Wang, and Y. He, "Visual recognition of driver hand-held cell phone use based on hidden crf," in *Proceedings of 2011 IEEE international conference on vehicular electronics and safety*, pp. 248–251, IEEE, 2011.

[11] C. Zhao, B. Zhang, J. He, and J. Lian, "Recognition of driving postures by contourlet transform and random forests," *IET Intelligent Transport Systems*, vol. 6, no. 2, pp. 161–168, 2012.

[12] Y. Lecun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document

recognition," *Proceedings of the IEEE*, vol. 86, no. 11, pp. 2278–2324, 1998.
[13] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," 2015.
[14] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," 2015.
[15] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, "Rethinking the inception architecture for computer vision," 2015.
[16] F. Chollet, "Xception: Deep learning with depthwise separable convolutions," 2017.
[17] G. Huang, Z. Liu, L. van der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," 2018.
[18] A. G. Howard, M. Zhu, B. Chen, D. Kalenichenko, W. Wang, T. Weyand, M. Andreetto, and H. Adam, "Mobilenets: Efficient convolutional neural networks for mobile vision applications," 2017.
[19] B. Zoph, V. Vasudevan, J. Shlens, and Q. Le, "Learning transferable architectures for scalable image recognition," pp. 8697–8710, 06 2018.
[20] T. Hoang Ngan Le, Y. Zheng, C. Zhu, K. Luu, and M. Savvides, "Multiple scale faster-rcnn approach to driver's cell-phone usage and hands on steering wheel detection," in *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*, pp. 46–53, 2016.
[21] E. Ohn-Bar, S. Martin, and M. Trivedi, "Driver hand activity analysis in naturalistic driving studies: challenges, algorithms, and experimental studies," *Journal of Electronic Imaging*, vol. 22, no. 4, p. 041119, 2013.
[22] D.-L. Nguyen, M. D. Putro, and K.-H. Jo, "Eye state recognizer using light-weight architecture for drowsiness warning," in *Intelligent Information and Database Systems* (N. T. Nguyen, S. Chittayasothorn, D. Niyato, and B. Trawiński, eds.), (Cham), pp. 518–530, Springer International Publishing, 2021.
[23] X.-P. Huynh, S.-M. Park, and Y.-G. Kim, "Detection of driver drowsiness using 3d deep neural network and semi-supervised gradient boosting machine," in *Asian Conference on Computer Vision*, pp. 134–145, Springer, 2016.
[24] "State farm distracted driver detection." https://www.kaggle.com/c/state-farm-distracted-driver-detection/data. Accessed: 2021-07-14.