

A Transfer Learning Approach for Identification of Distracted Driving

Ikromjanov Kobiljon Komil Ugli*, Ali Hussain*, Beom Su Kim*, Satyabrata Aich*, Hee-Cheol Kim**

*Institute of Digital Anti-Aging Healthcare, Inje University, South Korea

** Department of Computer Engineering/Institute of Digital Anti-Aging Healthcare/u-HARC, Inje University, South Korea

kobiljonikromjanov@gmail.com, alihussain.dream@gmail.com, beomsu91@naver.com, satyabrataaich@gmail.com, heeki@inje.ac.kr

Abstract: In recent years, the number of motor vehicle traffic crashes is rising around the world. According to the United States National Highway Traffic Safety Administration (NHTSA), the main cause of road fatalities and injuries is distracted drivers. Driver distraction is a specific sort of driver interference on the roadway. In this case, a deep learning-based system can detect and distinguish the source of distractions in real-time, to avoid traffic accidents and make better transport safety. This paper attempts to develop the system using transfer learning and fine-tuning methods with different model architectures. Various pre-trained weights were used with fine-tuning to improve accuracy and implemented using Mobile Net, VGG16, and ResNet50 models. Finally, the results illustrate that transfer learning on Mobile Net with frozen layers is the best model out of three models getting 99% (495 out of 500) of accuracy on the test dataset images.

Keywords—Distracted driver, deep convolutional neural networks, transfer learning, fine-tuning, pre-trained model.

I. INTRODUCTION

In our modern life, the rising number of transports on the roadways affected an increase in traffic crashes. According to the World Health Organization, every year nearly 1.3 million people die and 50 million are injured as a result of traffic accidents [1]. According to a recent study by the United States National Highway Traffic Safety Administration (NHTSA), one of the main causes of road fatalities and injuries is distracted drivers. Distracted driving is dangerous, claiming 2,841 human lives in 2018 alone. Among them: 1730 drivers, 605 passengers, 77 bicyclists, and 400 pedestrians on the US roads. These risky distractions lead to a danger to car owners, passengers as well as pedestrians, and pedal cyclists [2]. Distraction happens when people avert their attention from driving safely to focus on some other activities. According to the Centre for Disease Control and Prevention (CDC), these behaviors are mostly divided into 3 main groups: visual, manual, and cognitive distractions [3].

- Manual distraction: The driver takes his/her hands off the wheel, e.g. drinking, eating, etc.
- Visual distraction: The driver looks away from the road, e.g. reading, watching the phone, etc.
- Cognitive distraction: The driver's mind is not completely concentrated on the driving task, e.g. talking, thinking, etc.

Although driving distractions are categorized into 3 main different types they do not always occur separately. For example, in the event of texting on the phone, three types of distractions occur at the same time: manual, visual, and cognitive distraction. Sadly, most road crashes occur due to human error. Nowadays, some major companies like Intel are working to develop an Advanced Driver-Assistant System (ADAS) to tackle the issue by creating systems to warn the driver when detection activities are occurring. From this deep learning takes part in ADAS computer vision techniques.

Many recent studies focus on the detection of distracted drivers via visual data. Tabriz et al. analyzed a percentage of eye closure (PERCLOS) to detect driver drowsiness [4]. The frequency and duration of a driver's eye state for a minor task are used to make a total measurement of eyes off the roadway. Cray et al. introduced a hidden Markov model (HMM) based method to detect distracted driving behaviors that achieved 90% accuracy [5]. Their technique needs to detect the driver's face and right arm. In 2017, Abouelnaga et al. provided a dataset that is similar to the State Farm dataset [10] and introduced a real-time distraction detection approach using a weighted assembly of five Alex Net and five Google Net models with hand, face, and skin features [6]. The system achieved a better classification accuracy of 94.29%, but it is a bit complex for real-time detection. Baheti et al. tackled this problem of complexity and decreased the number of parameters remarkably and achieved an accuracy of 95.54% [7]. In 2019, Vlad et al focused on the activation functions on the VGG-16 base architecture and improved the test accuracy to 95.82% [8]. Basubeit et al. [9] proposed a transfer learning technique with 12 different models to compare pre-trained deep convolutional neural networks and achieved 96% accuracy result via the VGG16 model on 200 unseen test images.

In this research, Mobile Net, VGG16, and ResNet50 deep convolutional network are used as a transfer learning framework where they come pre-trained on the ImageNet dataset. Besides, three different fine-tuning strategies were applied to freeze some of their layers for comparing and increasing the accuracy. State Farm dataset [10] is used for the experiment to classify images into ten classes. The classification accuracy of the second model of Mobil Net will be visualized using the non-normalized and normalized

confusion matrix that showed better accuracy results among them with 495 correct predictions out of 500 image testbeds.

II. DATASET DESCRIPTION



Figure 1: Sample images of each class of distracted driving

In this experiment, State Farm Distracted Driver Detection is used which is a publically available image dataset on Kaggle.com. In 2016, State Farm started a challenge to detect distracted drivers' behaviors based on a provided dataset. It consists of 22,424 training and 79,727 testing color images that showed drivers either captured in distracting activities or driving safely [10]. There is a total of ten classes to predict. Table 1 illustrates the ten different behaviors, while Figure 1 shows some of the sample images for each class.

When it comes to training the model, 50 images are taken in each class for testing (500 unseen images), 80 percentage for training (17,544), and 20 percentage for validation (4,380) respectively among 22,424 total images.

Table. 1 List of distracted driving activities

Class	Behaviour
C0	Safe driving
C1	Texting using right hand
C2	Talking on the phone using the right hand
C3	Texting using left hand
C4	Talking on the phone using the left hand
C5	Operating (Adjusting) the radio
C6	Drinking
C7	Reaching behind
C8	Hair and makeup
C9	Talking to a passenger

III. DEEP LEARNING MODEL

A. Transfer Learning

Transfer learning is a deep learning technique that stores the knowledge gained while solving one issue and use it to a new but related problem. Nine models based on Mobile Net, VGG16 and ResNet50 models were used in this research. They are originally trained on the ImageNet database which can classify

images into thousands of object categories. Due to these advantages, three models' architectures and weights are used on top layers of the model to train the State Farm dataset. Table 2 demonstrates nine models, hidden layers, classification layers, and overall, their depth. The first three models have been applied 87 Mobile Net layers and its trained weights and three additional layers added on classification part. While the other three models were pre-trained 23 layers of VGG16 but ten categories on the output layer.

Moreover, the remaining models built-in 175 layers of ResNet50 and GlobalAveragePooling2D (), Dense (150), and Dense (10) layers.

Table. 2 List of Models

Architectures	Models	Hidden Layers	Classification Layers	Depth
MobileNet	model_1	87	GlobalAveragePooling2D() Dense(150) Dense(10)	90
	model_2			
	model_3			
VGG16	model_4	19	Flatten() Dense(4096) Dense(4096) Dense(10)	23
	model_5			
	model_6			
ResNet50	model_7	175	GlobalAveragePooling2D() Dense(150) Dense(10)	178
	model_8			
	model_9			

B. Fine-Tuning

Fine-tuning is a way of utilizing or applying transfer learning. There is a different kind of fine-tuning technique, truncating the last layer, using a smaller learning rate, and freezing the weights of the layers are among these methods. In this paper, the freezing technique is used to compare each model. Table 3 represents that model_1, model_4, and model_7 layers are frozen till hidden layers, while model_3, model_6, and model_9 layers are all trainable (or the last one has frozen VGG16). When it comes to the remaining three models, different approaches are used. They are frozen 80, 21, and 160 accordingly to achieve average trainable parameters.

Table. 3 List of frozen layers and trainable parameters

Models	Depth	Frozen Layers	Total params	Trainable params
model_1	90	87	3,384,124	155,260
model_2		80		1,217,148
model_3		No Frozen		3,362,236
model_4	23	19	134,301,514	119,586,826
model_5		21		16,822,282
model_6		22		40,970
model_7	178	175	23,896,572	308,860
model_8		160		5,829,244
model_9		No Frozen		23,843,452

IV. RESULT AND DISCUSSION

In this section, we want to demonstrate the result of the figures and tables for getting clear comprehension of the comparison of all nine pre-trained models used different techniques. Looking at figures 2 and 3 provide that re-trained models with only 15 epochs represent a significant increase in accuracy and decrease in loss values. Especially, fine-tuned models that have average trainable parameters, model_2, model_5, and model_8 showing better performance in early epochs. Table 4 highlights the first and the last fifteen epochs with train accuracy, validation accuracy, train loss, and validation loss values of nine models. Comparing all re-trained models, model_2 used Mobile Net architecture, and pre-trained weights with 80 frozen layers yield the highest accuracy of 99.33% and 99.22% and the lowest loss of 2.32% and 2.4% rate on train and validation sets respectively.

Furthermore, Table 5 illustrates predictions of 500 unseen test data and their accuracy. On the testbed, model_2 predicted 495 images correctly and only making 5 mistakes out of 500 test images, scoring 99% as overall accuracy and 0.04169 multiclass logarithmic loss. Figures 4 and 5 show the non-normalized and normalized confusion matrix.

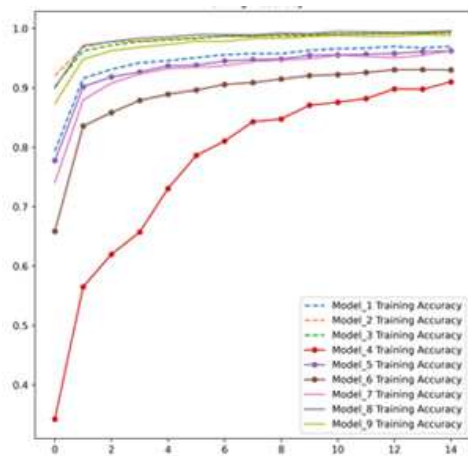


Figure 2. Training Accuracy

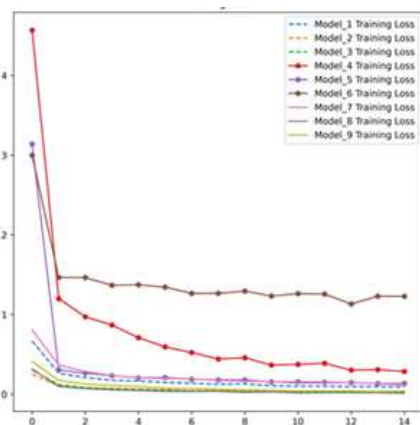


Figure 3. Training Loss

Table. 4 First and Fifteenth Epoch values of all models

Models	Training Accuracy		Validation Accuracy		Training Loss		Validation Loss	
	epoch_1	epoch_15	epoch_1	epoch_15	epoch_1	epoch_15	epoch_1	epoch_15
model_1	0.7934	0.9699	0.9221	0.9667	0.6686	0.0923	0.2621	0.0951
model_2	0.9203	0.9933	0.9676	0.9922	0.2544	0.0232	0.1086	0.024
model_3	0.9023	0.9924	0.9358	0.9856	0.3052	0.0234	0.1928	0.0634
model_4	0.3422	0.9101	0.4868	0.9112	4.5661	0.2862	1.2673	0.2955
model_5	0.7776	0.9624	0.9217	0.9731	3.1387	0.1356	0.2715	0.1226
model_6	0.6586	0.9302	0.8644	0.9594	2.9989	1.2299	1.3217	0.8119
model_7	0.7405	0.9628	0.8078	0.9596	0.8094	0.1152	0.5304	0.1284
model_8	0.9	0.9928	0.9632	0.991	0.3177	0.025	0.1176	0.026
model_9	0.8739	0.9881	0.9034	0.989	0.414	0.041	0.3142	0.042

Table. 5 Test set accuracy

Models	Samples	Correct	Incorrect	Accuracy %
model_1	500	476	24	95,2
model_2	500	495	5	99
model_3	500	487	13	97,4
model_4	500	466	34	93,2
model_5	500	493	7	98,6
model_6	500	485	15	97
model_7	500	475	25	95
model_8	500	492	8	98,4
model_9	500	486	14	97,2

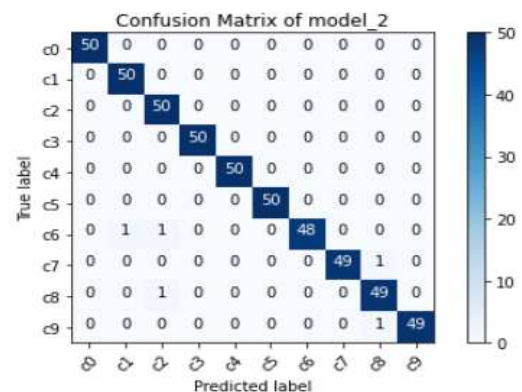


Figure 4. Non-normalized Confusion Matrix of model_2

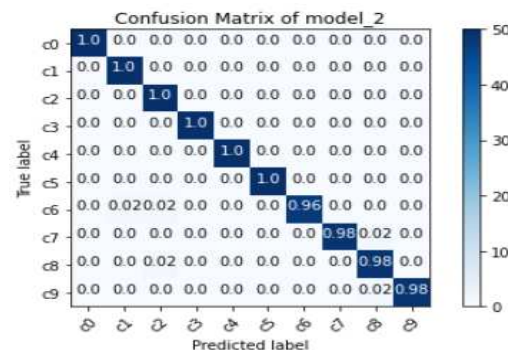


Figure 5. Normalized Confusion Matrix of model_2

V. CONCLUSIONS

In this paper, it was found that pre-trained deep learning models on the ImageNet dataset can be fine-tuned to improve the accuracy to classify State Farm distracted driver dataset. Three architectures and their weights were used and applied with three different freezing techniques on the layers making them overall nine models. After re-training models with distracted drivers image dataset, they are evaluated by predicting 500 unseen test sets. As a result, Mobile Net model_2 showed the best with 99% accuracy. It is recommended that this model could be used in real-time environments.

ACKNOWLEDGMENT

This research was supported by the Ministry of Trade, Industry, and Energy (MOTIE), KOREA, through the Education Program for Creative and Industrial Convergence (Grant Number N0000717)

REFERENCES

- [1] WHO Library Cataloguing-in-Publication Data, "Mobile phone use: a growing problem of driver distraction" 2011 Available:



Ikromjanov Kobiljon Komil Ugli is a master student at the Institute of Digital Anti-Aging Healthcare, Inje University. He has received his Bachelor degree in computer engineering from South Korea. He has worked in many real-life projects related to artificial intelligence. His research interests are artificial intelligence, Machine learning, big data, machine vision, Internet of Things.



Ali Hussain is a master student at the Institute of Digital Anti-Aging Healthcare, Inje University. He has received his Bachelor degree in computer engineering from Pakistan. He has worked in many real-life projects related to artificial intelligence. His research interests are artificial intelligence, Machine learning, big data, machine vision.



Beomsu Kim is a master student at the Institute of Digital Anti-Aging Healthcare, Inje University. He has worked in many real-life projects related to artificial intelligence. His research interests are artificial intelligence, Machine learning, big data, text mining

- [2] https://www.who.int/violence_injury_prevention/publications/road_traffic/distracted_driving_en.pdf
- [3] US Department of Transportation – National Highway Traffic Safety Administration: 'Traffic safety facts', April, 2020. Available : <https://crashstats.nhtsa.dot.gov/Api/Public/ViewPublication/812926>
- [4] Centre for Disease Control and Prevention (CDC), "Distracted driving." Retrieved from October 26, 2020. Available at: https://www.cdc.gov/transportationsafety/distracted_driving/index.html
- [5] Tabrizi, P.R. and Zoroofi, R.A., 2009, September. Drowsiness detection based on brightness and numeral features of eye image. In 2009 Fifth International Conference on Intelligent Information Hiding and Multimedia Signal Processing (pp. 1310-1313). IEEE.
- [6] Craye, C. and Karray, F., 2015. Driver distraction detection and recognition using RGB-D sensor. arXiv preprint arXiv:1502.00250.
- [7] Abouelnaga, Y., Eraqi, H.M. and Moustafa, M.N., 2017. Real-time distracted driver posture classification. arXiv preprint arXiv:1706.09498.
- [8] Baheti, B., Gajre, S. and Talbar, S., 2018. Detection of distracted driver using convolutional neural network. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops (pp. 1032-1038).
- [9] Tamas, V. and Maties, V., 2019. Real-Time Distracted Drivers Detection Using Deep Learning. American Journal of Artificial Intelligence, 3(1), pp.1-8.
- [10] Basubeit, O.G., How, D.N.T., Hou, Y.C. and Sahari, K.S.M., Distracted Driver Detection with Deep Convolutional Neural Network.
- [11] Kaggle: State Farm Distracted Driver Detection. Available at: <https://www.kaggle.com/c/state-farm-distracted-driver-detection/data>



Satyabrata Aich, is working as a Lecture at the Institute of Digital Anti-Aging Healthcare, Inje University. He has published many research papers in journals and conferences in the realms of Machine learning, Text mining, and Artificial intelligence. His research interests are natural language processing, Machine learning, supply chain management, Text mining, Bio informatics, and Blockchain.



Hee-Cheol Kim, received his BSc at Department of Mathematics, MSc at Department of Computer Science in SoGang University in Korea, and PhD at Numerical Analysis and Computing Science, Stockholm University in Sweden in 2001. He is Professor at Department of Computer Engineering and Head of the Institute of Digital Anti-aging Healthcare, Inje University in Korea. His research interests include Machine learning, Text mining, Bio Informatics.