Distracted Driver Detection*

Shyam Prasad Immadi
Praneeth Varma D B R S
Phani Chyavan DSK
Department of Computer Science
Amrita Vishwa vidyapeetham,
Amritapuri, India
shyamimmadi7@gmail.com

***Abstract*—There has been an increase in the number of road accidents worldwide over the past few years. Distracted driving is driving while doing an activity that diverts your attention fIn most car accident cases, the defendant driver is not accused of intentionally causing the accident, but rather of making a mistake or omission that led to an unexpected collision. There are many accidents caused by driver error. This paper proposes a solution for detecting distracted driving by drivers. For the classification of distracted drivers images in State Farm Distracted DriverDetection challenge on Kaggle we have used some pre-trained and simple architectures, namely: MobiNetv3, AlexNet, FNN. The Framework and languages used are Pytorch and python respectively.Our model recorded an accuracy of 98 percent with a loss of 0.021.**

***Index Terms*—Classification, Pytorch, CNN, Python, Transfer Learning, MobiNetv3.**

I. Introduction

In survey data from the World Health Organization (WHO), 1.3 million people worldwide die as a result of traffic accidents each year, which makes it the eighth leading cause of death. An additional 20-50 million are injured or disabled. As per the report of National Crime Research Bureau (NCRB), Govt. of India, Indian roads account for the highest fatalities in the world.

Today, Advanced Driver Assistance Systems (ADAS) are being developed to prevent accidents by alerting the driver to potential problems and keeping the driver and occupants of the car safe if an accident occurs.In case of an emergency, even the most modern autonomous vehicles require the driver to be alert and ready to take the wheel back.Tesla's autopilot slammed into a white truck-trailer in Williston, Florida in May 2016 which was the first fatal crash in autonomous vehicle testing. In 2018, an Uber self driving car with an emergency backup driver struck and killed a pedestrian in Arizona.The safety driver could have avoided both of these crashes, but evidence indicates that he was clearly distracted. This makes detecting distracted drivers essential for self-driving cars as well.In an effort to prevent road crashes, distracted drivers detection is deemed to be a vital component. If the vehicle could detect such distractions and warn the driver, the number of road accidents could be reduced.

The number of fatal crashes continues to rise despite improvements in road and vehicle design.Additionally, road traffic accidents cause large property damages, and distracted driving is a contributing factor to an increase in accidents.Although these accidents may not be completely avoided, these techniques can reduce accidents and warn drivers before it is too late. Input images are given to the model knowing driver distractions, which determines the output by labeling driver's actions with an output having a high probability of labeling.The model predicts the class of an image by giving as an output a probability for each class. Each image corresponds to one of the 10 classes defined in the dataset section.

II. Related Work

The following section includes a review of some of the relevant, significant work from literature on distracted driving detection. The most common source of manual distractions is the use of cell phones. Motivated by the same, some researchers investigated cell phone usage detection while driving.Zhang created a database based on a camera mounted above the dashboard and used Hidden Conditional Random Fields to detect cell phone usage. It basically works by observing facial, mouth, and hand features.Nikhil provided a dataset from 2015 that was used to detect hands in the automotive environment, and the average precision was 70.09 percent using the Aggregate Channel Features object detector.

Many of the earlier datasets focus on only a small set of distractions, and many of them aren't publicly available. However, StateFarm's distracted driver detection competition on Kaggle defined ten postures in April 2016 (Safe driving + nine distracted behaviors).State Farm hosted this challenge on Kaggle. Most solutions were based on SVM models to detect phone usage by drivers while driving. Others were based on facial and hand segmentation using CNN. In addition to handcrafted features such as HOG and DSIFT, there are a number of approaches based on Deep Neural Networks.

Some of them have not used ensembles and have only applied a single model to the dataset. Data augmentation is one of the technologies lacking in the previous studies. This technique adds more data to the dataset through zooming, rotating, and shearing. These techniques reduce overfitting.

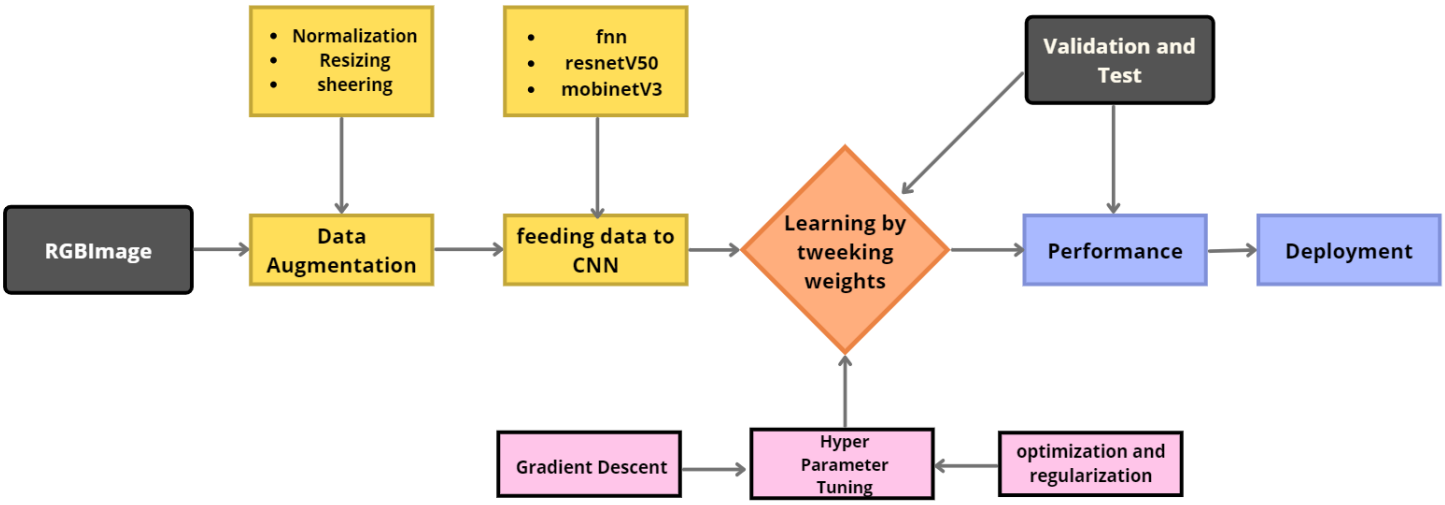


Fig. 1. Example of a figure caption.

III. Methodology

Typically, convolutional neural networks have a convolution layer, a pooling layer, and a full connection layer. The features are extracted through convolution layers and pooling layers. Then, all the feature maps from the last convolution layer are transformed into one-dimensional vectors for full connection. Finally, the output layer classifies the input images.All feature maps are transformed into one-dimensional vectors after the last convolution layer, which facilitates full connection of input and output images.

Each layer has three dimensions: width, height, and depth, where width and height refer to the size of the neurons, while depth corresponds to the number of channels in the input picture or the number of input feature maps.By using both depthwise and pointwise separable convolutions, MobileNet significantly reduces the number of parameters when compared to networks with regular convolutions of the same depth. This leads to lightweight deep neural networks.

This type of network is made from two operations.
1.Depthwise convolution, 2.Pointwise convolution.

In the MobileNet model, there are 27 convolution layers (Fig. 3.a) including 1 normal convolution, (Fig. 3.b) 13 depth-wise convolutions and 13 pointwise convolutions, 1 Average Pool layer, 1 Fully Connected layer, and 1 Softmax layer.

Among the convolution layers, there are the following: 13 3x3 depthwise convolutions 1 3x3 convolution 13 1x1 convolutions 95 percent of time is spent on 1x1 convolutions in MobileNet.

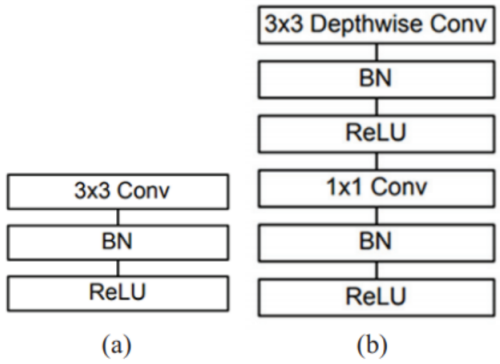


Table 1. MobileNet Body Architecture

Type / Stride	Filter Shape	Input Size
Conv / s2	$3 \times 3 \times 3 \times 32$	$224 \times 224 \times 3$
Conv dw / s1	$3 \times 3 \times 32$ dw	$112 \times 112 \times 32$
Conv / s1	$1 \times 1 \times 32 \times 64$	$112 \times 112 \times 32$
Conv dw / s2	$3 \times 3 \times 64$ dw	$112 \times 112 \times 64$
Conv / s1	$1 \times 1 \times 64 \times 128$	$56 \times 56 \times 64$
Conv dw / s1	$3 \times 3 \times 128$ dw	$56 \times 56 \times 128$
Conv / s1	$1 \times 1 \times 128 \times 128$	$56 \times 56 \times 128$
Conv dw / s2	$3 \times 3 \times 128$ dw	$56 \times 56 \times 128$
Conv / s1	$1 \times 1 \times 128 \times 256$	$28 \times 28 \times 128$
Conv dw / s1	$3 \times 3 \times 256$ dw	$28 \times 28 \times 256$
Conv / s1	$1 \times 1 \times 256 \times 256$	$28 \times 28 \times 256$
Conv dw / s2	$3 \times 3 \times 256$ dw	$28 \times 28 \times 256$
Conv / s1	$1 \times 1 \times 256 \times 512$	$14 \times 14 \times 256$
$5\times$ Conv dw / s1	$3 \times 3 \times 512$ dw	$14 \times 14 \times 512$
Conv / s1	$1 \times 1 \times 512 \times 512$	$14 \times 14 \times 512$
Conv dw / s2	$3 \times 3 \times 512$ dw	$14 \times 14 \times 512$
Conv / s1	$1 \times 1 \times 512 \times 1024$	$7 \times 7 \times 512$
Conv dw / s2	$3 \times 3 \times 1024$ dw	$7 \times 7 \times 1024$
Conv / s1	$1 \times 1 \times 1024 \times 1024$	$7 \times 7 \times 1024$
Avg Pool / s1	Pool 7×7	$7 \times 7 \times 1024$
FC / s1	1024×1000	$1 \times 1 \times 1024$
Softmax / s1	Classifier	$1 \times 1 \times 1000$

Fig. 2. sample images of 10 classes

IV. Dataset

The dataset used in the study is taken from a public

Kaggle challenge by State Farm.State Farm is a large group of insurance and financial services companies throughout the United States. They released their dataset of 2D dashboard camera images for a Kaggle Challenge. The dataset consists of 22,400 training and 79,727 validation labelled images of driver behaviours. Resolution was 640 x 480 pixels. There are a total of ten classes of behaviours provided in the dataset.

- c0: normal driving
- c1: texting - right hand
- c2: talking on the phone - right hand
- c3: texting - left hand
- c4: talking on the phone - left hand
- c5: operating the radio
- c6: drinking
- c7: reaching behind
- c8: hair and makeup
- c9: talking to passenger

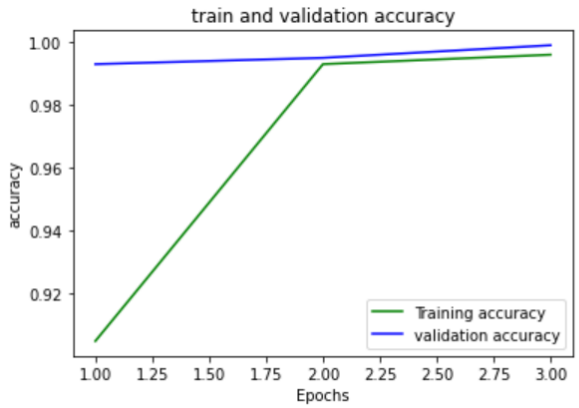
V. Data Availability

data set used is a public data set that can be downloaded online.

VI. Results

A Convolutional Neural Network system is designed for distracted driver detection using the pre-trained MobiNetv3 model for weight initialisation and the transfer learning concept. Weights of all the layers of the network are updated based on the dataset. We have fine tuned all the hyperparameters after extensive testing. Training is carried out with Stochastic Gradient Descent using a learning rate of 0.001, a momentum value of 0.9 and batch size and number of epochs set to 3 and 10 respectively. Then we extracted loss for 5 batches, it recorded results 2.343, 0.154, 0.006 for 3 epochs. The framework is developed using pytorch. When original MobiNetv3 is used as it is for the task of distracted driver detection, it produces 100 percent on the training set and 98 percent accuracy on the test set.

This MobiNetv3model worked better compared to the Alexnet architecture which got a testing accuracy of 94 percent and Restnet architecture which got a testing accuracy of 95 percent.



A. Discussion

During the extension of this work, we will be working on improving the efficiency and reducing the number of parameters. We hope that incorporating temporal context may assist in enhancing classification accuracy and reducing misclassification errors.

Furthermore, we wish to develop a system that detects visual and cognitive distractions as well as manual distractions in the future.

VII. CONCLUSION

Following an experiment with CNN models, the standard MobileNet,Alexnet,VGG models have 4.2M,60M,138M parameters respectively. The use of MobileNet gives us computational freedom and reliability. Our best ensemble was created using MobiNetv3 probabilities.Our final log loss was 0.021. We were only using the GPU provided by the Google Colab Platform for the project. If we had had access to more computing power, we might have achieved better results.

[2] [1]

REFERENCES

[1] Bhakti Baheti, Suhas Gajre, and Sanjay Talbar. Detection of distracted driver using convolutional neural network. In *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*, pages 1032–1038, 2018.

[2] Andrew G Howard, Menglong Zhu, Bo Chen, Dmitry Kalenichenko, Weijun Wang, Tobias Weyand, Marco Andreetto, and Hartwig Adam. Mobilenets: Efficient convolutional neural networks for mobile vision applications. *arXiv preprint arXiv:1704.04861*, 2017.