Distracted Driver Detection with MobileNet-v2

Sayali More, Heet Detroja, Austin White

UNC CHARLOTTE

Motivation

- According to the National Highway Traffic Safety Administration (NHTSA) texting is the most alarming distraction. Any non-driving activity you engage in is a potential distraction and increases the risk of crashing. In 2017 alone, 3166 people were killed in vehicle crashes involving distracted drivers.
- Distracted driving accounts for approximately 25% of all motor vehicle crash fatalities.

Background

Industry Solutions

TESLA MOTORS

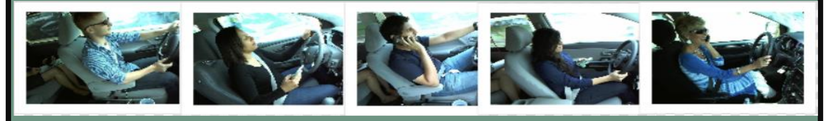
Academic Literature

- **Detection of Distracted Driver using Convolutional Neural Network** looks into modified VGG-16 Network pretrained on ImageNet and had an accuracy of 95.54%. They we able to process 42 images per second through an Nvidia P5000.
- **Real-time detection of distracted driving based on deep learning** implements 4 different models among which ResNet gave the best accuracy of 92%. They used Nvidia Jetson TX1 and NanoPi M3 embedded systems for computation.

Future Enhancements

- Real time processing of video or continuous images.
- Running model in augmented environment with alert system for the drivers.
- Trying other models and reducing processing speed further.
- Currently, the similarities in postures of different behaviours result in incorrect classifications. The behaviours that have more misclassifications are ‘drinking’, ‘hair and makeup’ and ‘texting on the phone – left’. In the future, we need to solve this misclassification issue.

Dataset



The dataset is from Kaggle State Farm Distracted Driver Detection Challenge. It consist of labelled images divided into 10 classes to predict. The images are divided into training and testing set with 79726 images for the test set. The train and test data are split on the drivers, such that one driver can only appear on either train or test

Approach

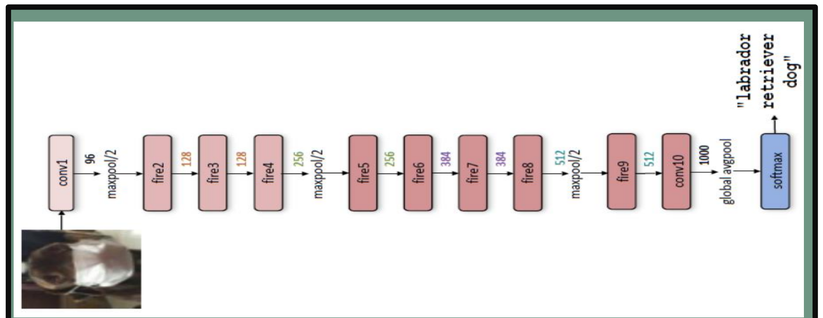


Fig. 01: SqueezeNet begins with a standalone convolution layer (conv1), followed by 8 Fire modules (fire2-9), ending with a final conv layer (conv10). A Fire module is comprised of: a squeeze convolution layer (which has only 1x1 filters), feeding into an expand layer that has a mix of 1x1 and 3x3 convolution filters.

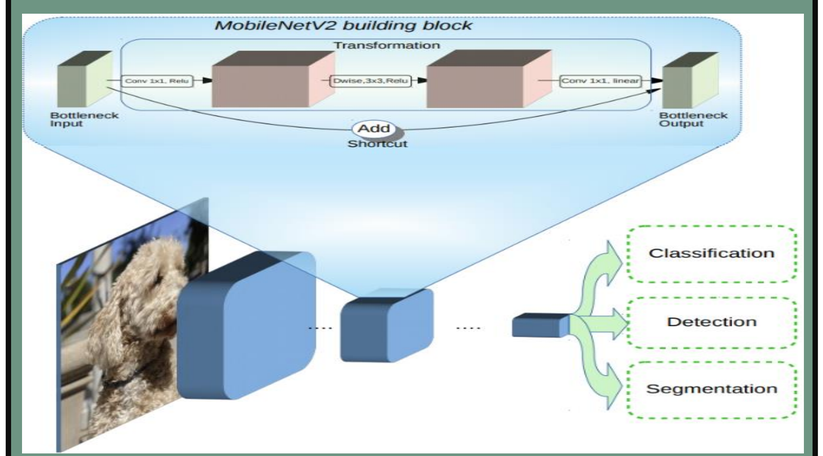


Fig. 02: MobileNetV2 contains the initial fully convolution layer with 32 filters, followed by 19 residual bottleneck layers with kernel size of 3 × 3 as is standard for modern networks, and utilize dropout and batch normalization.

Results

As there already exists a lot of good models with higher accuracies as a solution to this problem but there processing time is too high, hence our focus was to get faster results along with keeping same or very little loss in the accuracies and following are the results we have achieved.

Model	Frames Per Second Processed	Accuracy
Squeezenet	22	98.5
MobileNetV_2	10	99.1
VGG16	2	99.2

Below are the top losses or misclassified images. Most of the misclassifications were observed between the classes like Safe driving or Talking to passenger or talking to phone and drinking or eating.

References

[1] State Farm Distracted Driver Detection.https://www.kaggle.com/c/state-farm-distracted-driver-detection/overview .

[2] Baheti, Bhakti & Gajre, Suhas & Talbar, Sanjay. (2018). Detection of Distracted Driver Using Convolutional Neural Network. 1145-11456. 10.1109/CVPRW.2018.00150.

[3] Sheng, Weihua & Tran, Duy & Do, Ha & Bai, he & Chowdhary, Girish. (2018). Real-time Detection of Distracted Driving based on Deep Learning. IET Intelligent Transport Systems. 12. 10.1049/iet-its.2018.5172.

[4] Abouelnaga, Yehya & Eraqi, Hesham & Moustafa, Mohamed. (2018). Real-time Distracted Driver Posture Classification.