DISTRACTED DRIVER DETECTION

Deep Ruparel
Shah & Anchor Kutchhi Engg College
Mumbai,India

Abhay Rajde
Shah & Anchor Kutchhi Engg College
Mumbai,India

Sahil Shah
Shah & Anchor Kutchhi Engg College
Mumbai,India

Prof..Manya Gidwani
Shah & Anchor Kutchhi Engg College
Mumbai,India.

***Abstract*—The rate of accidents have been increasing at an alarming rate in recent years all around the globe. When trying to find the main reason behind this increasing rate it was found that the majority of the drivers were distracted when driving which caused their accidents. The survey of the National Highway Traffic Safety Administration states that nearly one in five motor vehicle crashes is caused by distracted drivers. We intend to develop a model to successfully identify whether the driver is distracted while driving or is the driver driving safely. In this paper, we try to list down the various approaches of deep learning that were followed by us such as the vgg16 model, mobilenet, sequential model, pre-trained weights of vgg16 model and many more which helped us to develop a model which was able to provide us with promising accuracy and definite results.Experimental results show that our system achieves an accuracy of 93%.**

Keywords—distracted driving, vgg16 model, mobilenet, pretrained vgg16 weights, sequential model,Convolution Neural Networks, Deep learning.

1. INTRODUCTION

Driving a car is a complex task, and it requires complete attention. Distracted driving is any activity that takes away the driver's attention from the road.According to the World Health Organization (WHO), road crashes kill 1.2 million people and permanently disable another 50 million every year. India lost 1.3 million people to road crashes and another 5.3 million have been seriously injured. India has the highest number of road crash fatalities, with a crash occurring every minute and one death every four minutes. While it has just 1% of the world's vehicles, India accounts for over 10% of global road crash fatalities.Driving while being distracted has become a major problem around the world and will likely become worse before it gets better.According to the CDC motor vehicle safety division, one in five car accidents is caused by a distracted driver. Sadly, this translates to 425,000 people injured and 3,000 people killed by distracted driving every year.The report also states that the total number of deaths have risen to 1.46 lakhs in 2015 and driver error is the most common cause behind these traffic accidents. Studies have found that the driver distraction detection consists of data from multiple modality namely from visual distraction, manual distraction and cognitive distraction.

Visual Distraction involves taking eyes off the road while driving, Cognitive distraction involves taking your mind off while driving which in simple words translates that one might be present physically while driving but mentally one might not be paying attention while driving, this includes daydreaming, being lost in thoughts etc, whereas manual distraction is when the driver takes his/her hands off the wheel while driving to perform multifarious actions like drinking, talking or texting over the mobile, reaching behind, operating the radio or talking to passengers while driving.

Nowadays, Advanced Driver Assistance Systems (ADAS) are being developed to prevent accidents by offering technologies that alert the driver to potential problems and to keep the car's driver and occupants safe if an accident does occur. But even today's latest autonomous vehicles require the driver to be attentive and ready to take the control of the wheel back in case of emergency. Tesla autopilot's crash with the white truck-trailer in Williston, Florida in May 2016 was the first fatal crash in testing of autonomous vehicles. Recently in March 2018, Uber's self driving car with an emergency backup driver behind the wheel struck and killed a pedestrian in Arizona. In both of these fatalities, the safety driver could have avoided the crashes but evidence reveals that he was clearly distracted. This makes detection of distracted drivers an essential part of the self - driving cars as well.

The primary focus of this paper is to detect the manual distractions where the driver is not primarily focused on driving but performing various tasks except driving and accurately classify these tasks which lead to distraction of the driver.

2. RELATED WORK

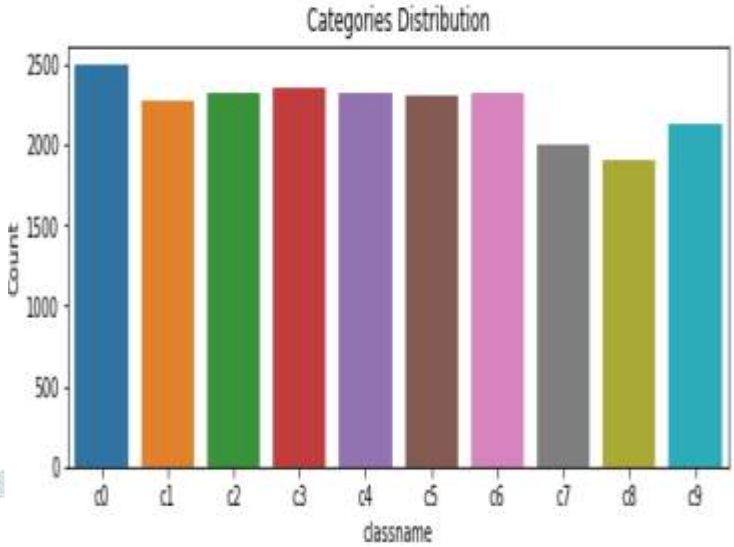
This section tries to summarize review of some of the relevant and significant work already performed from literature for distracted driver detection.

Reference[1] provides a solution by using a genetically weighted ensemble of convolutional neural networks.These networks were trained on raw images , face images , skin segmented images,hand images and face plus hand images.On these five image sources , the model was trained and benchmarked an AlexNet network , an InceptionV3 network, a ResNet network having 50 layers and a VGG-16 network.A pre-trained ImageNet model was fine tuned for the above mentioned networks.Then,by using a generic algorithm,they evaluated a weighted sum of all networks outputs , with the help of which they obtained the final class distribution.This final class distribution is formulated by the weighted sum of all softmax layers.They had yielded an accuracy of 76% using VGG-16 model and an accuracy of 82% by using Resnet50.Their proposed solution had a high computation time.

Reference[2] provides a solution by using an algorithm which consists of various parts.The first part was acquisition, which consists of the driver image capture system.The second part and one of the most integral part of this solution was pre-processing.In this part, the face of driver was located by using three detectors and the regions that were not useful were cropped out,thus obtaining the regions that would be the main focus.By doing so,the width of the region is increased by 40% i.e. 20% to left and 20% to the right ,thus enabling us to detect

hands of the driver,which otherwise would have been cropped out.The third part is segmentation, which is primarily based on locating pixel of the drive's skin present in the pre-processed images.It was done by firstly converting the
image into two different color spaces namely,HSV and YCrCb.The last part was extraction of features,which included two features which are part of the driver's classification with or without cell phone namely,Percentage of the Hand (PH) and Moment of Inertia (MI).PH was obtained by calculating skin pixels in two regions of the image whereas, MI was calculated using inertia movement.The support vector machines (SVM) was used as a classifier in this approach.They received an accuracy of 77.33% for "with phone" and 88.97% for "no phone" and an overall accuracy of 83%.

Reference[3] created their own dataset similar to StateFarm's dataset for distracted driver detection. They proposed a solution that consists of a genetically-weighted ensemble of convolutional neural networks. The convolutional neural networks train on raw images, face images, hands images, and "face+hands" images. They trained an AlexNet Krizhevsky and an InceptionV3 Szegedy on those four image sources. In the InceptionV3 network, they fine-tuned a pre-trained ImageNet model (i.e. transfer learning). Then, they evaluated a weighted sum of all networks' outputs yielding the final class distribution. The weights were evaluated using a genetic algorithm.Although they had an accuracy of 95%, the implementation of this system in real-time was a very complex and tedious task.



DATASET DESCRIPTION

In this paper, we have taken the dataset from kaggle and made use of it.This dataset is provided on kaggle by Statefarm. This

Figure 1: 10 Classes of driver postures from the dataset

dataset has ten classes namely safe driving(c0), texting on mobile phones using right hand(c1), texting on mobile phones using left hand(c2), talking on mobile phones using right hand(c3), talking on mobile phones using left hand(c4), adjusting radio(c5),eating or drinking(c6), hair and makeup(c7), reaching behind(c8) and talking to passengers(c9).Fig 1 provides a snapshot of each class which are mentioned above. The dataset consists of a total of 102150 images. Out of these, 21,569 images are taken for training the model. The training dataset is partitioned into ten different classes according to the categories of the distracted driver and each class consists of almost 2200 images each as shown in Figure 2. The training data is then divided once again with about 17,423 images which are used for learning and 4,146 images which have been utilised for validation testing.The rest of the images have been used for testing of the model developed.

3. METHODS / IMPLEMENTATION

1. The first approach used by us was to use a sequential model with the help of Keras.The Sequential model is a linear stack of layers.It allows us to create models layer-by-layer. It is limited in that , as it does not allow us to create models that share layers or have multiple inputs or outputs.In our approach we used the different layers present in the sequential model, such as Conv2d, max pooling with pool size of (2,2) and relu was used for activation . Flatten and dense was also used along with the mentioned layers.The optimiser used was Adam as well as SGD , with categorical_crossentropy and accuracy was used as a metric to evaluate the result.It was found that the Adam optimiser was performing better than the SGD optimiser as the learning curve of Adam had a better growth curve when juxtaposed with SGD.The overall accuracy of this method was about 12%. The shortcomings of this approach were that it wasn't able to identify the images with the desired accuracy.Then , we tried performing permutation and combinations,by changing the values in the above specified layers , with the help of nested loops.The results of these were plotted on a graph and with the help of these plotted graphs, we were able to find the best combination which when used would provide a better accuracy , as the learning curve of this combination was comparatively better than the other combinations. Surprisingly, even after using this combination, with a better learning curve, it was of no use as the accuracy increased only by 8%, i.e. we got an accuracy of 20% .This accuracy was nowhere near our expected goal.

The second approach used by us was making use of CNN.
CNNs have shown impressive progress in various tasks like object detection, action recognition,image classification, natural language processing and many more,since the last few years. The basic architecture of a CNN based system includes Convolutional filters/ layers, Activation functions, Pooling layer and Fully Connected (FC) layer. A CNN is formed by stacking these layers one after the other. Various architectures like AlexNet, ZFNet, VGGNet, GoogLeNet, ResNet have established benchmarks in computer vision. In this paper, for our project, we are using the VGG-16 architecture and modifying it for the task of distracted driver detection.
In our approach ,we first split the training data for validation and training. A set of drivers were selected and all the images pertaining to those drivers were selected for the purpose of validation testing. Of the total 21,569 images ,9327 were selected for validation testing. VGG-16 is a simple architecture model which always makes use of 3x3 filters with stride of one in convolution layer and uses the same padding in pooling layers 2x2 with stride of two.On top of these layers , we added Flatten and Dense layers . The dense layer used helps us to classify the image based on the ten postures which are mentioned above.We use "softmax" as an activation in the dense layer and SGD as an optimiser. The accuracy we got using this approach increased significantly when compared to the previous approach where we had just received an accuracy of 20% the accuracy here rose to 80% and when tested on the images, it had a success rate of 80%. This approach did solve the shortcomings of previous approaches where the model couldn't properly identify the images. This model had a few drawbacks of its own as it was not able to differentiate between left and right hands while identifying. It was able to correctly classify that the driver was distracted and talking or texting over the phone but when it comes to identifying which hand was used, this model was failing. The results of this model are present in Table 1. Another drawback is that this model was very heavy in weight and thus couldn't be used easily in mobile applications as it would increase the size of application.

The third approach used by us was making use of the Mobilenet model which is another CNN. Mobilenet is a very simple but efficient and not computationally intensive convoluted neural network. Mobilenet has been widely used for object detection, fine-grained classifications, face attributes, and localization. Since it is lightweight, it can be used for fine grained classifications. It could easily overcome the drawbacks of the previous approach. The Mobilent has a total of 28 layers which contains convolution layers with each convolution layer followed by batch normalization and Relu. The architecture also comprises average global pooling layer and softmax as a classifier in the end part of architecture. Like in the previous approach we have added Flatten layer and dense layers to the traditional architecture. The SGD optimizer has been used by us in this approach because of its better learning curve than Adam optimizer. The major advantage of this Mobilenet architecture was that it consists of resolution multiplier and width multiplier which allows better image classification. Like in the previous approach, we selected a set of drivers whose images were used for validation testing and the rest were used for training of the model.A total of 9,237 from the 21,569 images were used for validation testing and the rest were used for training of the mobilenet model. When epochs were run for training the model ,we received a learning accuracy of 89% and a validation accuracy of 86%. On performing the testing on test images provided in the dataset , we got an accuracy of 86% which was a great increase when compared to previous approaches. It was easily noticeable that this model was able to differentiate between the left and right hands unlike the previous model.But the accuracy wasn't that good when differentiating between hands though better than the previous model still not the desired goal we wished to have. It also had some problems in classifying images in which the driver was performing hair and make up or was reaching behind. The accuracy for these classes were very low and nowhere near our desired mark.To overcome this we tried to save the trained model using a checkpoint method where the model with the best accuracy while training is saved. Once this model was saved we tried to train the saved model once again. This was done by running more epochs on the saved model with a view to increase the training accuracy and most importantly increasing the validation accuracy. This method is performed by running about 60-70 epochs on the already saved model. The learning accuracy by this approach increased to 95% but the validation accuracy did not show such a surge, it increased by only 1% i.e. the validation accuracy which we hoped will increase exponentially was increased to only 87% and when manual testing was performed the accuracy of this model was almost the same with only few classes showing a minor increase in their accuracy.The results of this testing are shown in Table 2 below.

The fourth approach used by us was the popular technique of transfer learning. In simple words, transfer learning is the method in machine learning that will focus on storing the knowledge which is gained by solving one problem and using this stored knowledge on some other problem. In recent years transfer learning has become very popular and has yielded very successful results. In our research it was found that the method of transfer learning was effective in reducing the confusion which the earlier approaches had while differentiating between left and right hands. Another major advantage of transfer learning is that it saves a lot of training time and even provides better performance of neural networks. We are able to utilize the knowledge that the model already has and apply it onto our problem. It also helps us to reduce the overall time that is taken when we try to develop a model from scratch as the model will utilize the knowledge already present. A lot of pre-trained model weights by several people are available over kaggle. We used the vgg-16 pre-trained weights for our approach. Rather than following the same methodology in our previous approach, this time we split the data as 80% of the total training data was used for training of the model and the rest 20% was used for validation testing. The optimizer used was SGD because of the previous mentioned reasons and when the model was trained by running just 7 epochs we were able to achieve a training accuracy of 97% and a validation accuracy of 98% which was a significant increase when compared to the previous approaches. The model was tested by providing test data from the dataset and we were able to receive an accuracy of 93%. This model which was developed by transfer learning increases the accuracy of those classes which prior to this approach were not providing good results. This model was able to clearly differentiate between left and right hand and even the accuracy of hair and makeup, reaching behind saw an exponential increase. The results of this approach are shown in Table 3.

4. RESULTS AND DISCUSSION

The result of the first approach using the sequential model was not upto the mark and this model was guessing almost all of the images incorrectly. Table 1 below depicts the result of the vgg-16 model.

Class	Total Samples	Correct Prediction	Incorrect Prediction	Accuracy
Safe Driving	39	24	15	61
Testing-right hand	40	34	6	85
Texting-left hand	37	29	8	81
Talking-right hand	32	32	0	100
Talking-left hand	37	32	5	90
Adjusting Radio	40	34	6	86
Drinking	32	24	8	75
Reaching Behind	50	42	8	84
Hair & Makeup	42	36	6	85
Talking to Passenger	29	17	12	58

Table 1: Results using Vgg-16 model

This model as mentioned above provided an exponential increase in accuracy with properly identifying many of the images correctly which were prior identified incorrectly. The average accuracy of the testing on 387 images was about 80%.

Table 2 below shows the results of the Mobilenet Model in which we saved the weights and trained those previously saved weights .This model provided a better validation accuracy as well as learning accuracy but when we performed the manual testing of the 387 images, the accuracy was 85%.

Class	Total Samples	Correct Prediction	Incorrect Prediction	Accuracy
Safe Driving	39	35	5	75
Texting-right hand	40	39	5	87
Texting-left hand	37	30	7	81
Talking-right hand	32	32	0	100
Talking-left hand	37	33	4	89
Adjusting Radio	40	35	5	87
Drinking	32	26	6	81
Reaching Behind	50	42	8	84
Hair & Makeup	42	36	6	85
Talking to Passenger	29	22	7	81

Table 2: Results using Mobilenet model

The Table 3 below depicts the results of the method where transfer learning was performed. The results of this approach were the best among all the approaches that were performed by us. The validation accuracy increased to a whooping 98% and the accuracy when manual testing was performed on 387 images, the accuracy was 93% which translates to 360 of the 387 images predicted right. Of those which were predicted wrong had minor errors because in many cases the cup was perceived as mobile or vice versa by our model.

Class	Total Samples	Correct Prediction	Incorrect Prediction	Accuracy
Safe Driving	39	36	4	90
Texting-right hand	40	39	1	97.5
Texting-left hand	37	34	3	92
Talking-right hand	32	28	4	88
Talking-left hand	37	34	3	92
Adjusting Radio	40	38	2	96.6
Drinking	42	39	3	93
Reaching Behind	50	44	8	88
Hair & Makeup	42	38	6	90
Talking to Passenger	29	23	7	80

Table 3: Results using Pre-trained Vgg-16 Model(Transfer Learning)

CONCLUSION AND FUTURE SCOPE

Distracted driving is a very serious problem that we face so an effort is made to develop a model to successfully identify what the driver is doing while driving which makes him distracted putting his life at risk.Thus an attempt to develop procedure and implement it using different methods and algorithms is done. Here, we present a robust Convolutional Neural Network based system to detect distracted drivers. We use the pretrained VGG-16 architecture weights for this task and apply the Dense and Flatten layers to this model along with SGD as an optimizer to prevent overfitting to the training data. Experimental results show that the model predicts the distracted drivers with an accuracy of 93%.

As an extension of the work, in order to increase the accuracy of the model we can use layers which will identify the face images, face+hand images and hand images in order to accurately detect what activity the driver is performing while driving.In the near future, we can also develop models which along with identifying manual distraction can also identify the cognitive and visual distraction thus leading to minimisation of accidents due to distracted driving. An attempt to incorporate IOT(Internet of Things) can also be done where the camera will capture real time images and send it over a cloud architecture where this model is deployed thus detecting whether the driver is distracted and if so, notifying the driver which will save his/her life.

REFERENCES

[1] Y. Abouelnaga, H. M. Eraqi, and M. N. Moustafa. Real-time distracted driver posture classification. CoRR, abs/1706.09498, 2017.

https://arxiv.org/pdf/1706.09498.pdf

[2]Driver Distraction Identification with an Ensemble of Convolutional Neural Networks.

https://www.hindawi.com/journals/jat/2019/4125865/

[3] R. A. Berri, A. G. Silva, R. S. Parpinelli, E. GirardandR. Arthur, "A pattern recognition system for detecting use of mobile phones while driving," in Proceedings of the 9th International Conference on Computer Vision Theory and Applications, VISAPP 2014, vol. 2, pp. 411–418, IEEE, Portugal, January 2014.

[4]Distracted Driving In India "A Study On MobilePhone Usage,Pattern &Behavior"Tns India Pvt. Limited (Kantar Public India)For Savelife Foundationan independent, non-governmental organization in April 2017.

[5]The report, 'Distracted Driving in India: a study on mobile phone usage, pattern & behaviour', :https://www.edriving.com/three60/distracted-driving-in-india-report-released/

[6]https://www.kaggle.com/c/state-farm-distracted-driver-detection/overview

[7]https://www.nhtsa.gov/risky-driving/distracted-driving

[8]Intel AI Developer "Combating Distracted Driving Behavior"https://software.intel.com/en-us/articles/ai-developer-project-part-2-combating-distracted-driver-behavior"