International Journal of Engineering Science Invention (IJESI)
ISSN (Online): 2319 – 6734, ISSN (Print): 2319 – 6726
www.ijesi.org ||Volume 7 Issue 10 Ver III || Oct 2018 || PP 28-31

A Review Approach For Detecting Distracted Driver

Parth Sayani[1], Vyom Unadkat[1], Harshal Kapadia[2], Prasham Shah[3]
Harshal Dalvi[4]

[1,2,3](Student, Information Technology Engineering,Dwarkadas J. Sanghvi College of Engineering, Mumbai University, India)
[4](Assistant Professor, Information Technology Engineering,Dwarkadas J. Sanghvi College of Engineering, Mumbai University, India)
Corresponding Author:VyomUnadkat

Abstract : *Distracted driving causes a great loss of life as well as money. According to the National Highway Traffic Safety Administration (NHTSA), in 2015 alone, 3477 people were killed, and 391000 were injured in motor vehicle crashes involving distracted drivers. To address this problem various models of Deep Learning like Support Vector Machines (SVM), Convolutional Neural Networks (CNN), DeepCNN are used to detect the distracted driving. Out of these models using the VGG-16+INCEPTION V4+KNN architecture in basic CNN was found to have the highest accuracy. Here we have proposed a better model RNN which gives a better accuracy and also solves the problem that were faced by the traditional methods and CNN by providing images at regular intervals and not only specifically considering the image at a particular instance.*

Keywords *-Deep Learning, CNN, RNN, SVM, VGG-16 + InceptionV4 + KNN, TensorFlow.*

Date of Submission: 05-10-2018 Date of acceptance: 18-10-2018

I. INTRODUCTION

Since past few years, people are really concerned about safe driving and this concern has reached remarkable levels. Accidents that have been caused due to inattentive driving are regarded as one of the reasons for causing deaths. Distractions are divided into three types which are manual distractions, visual distractions and cognitive distractions. In manual distraction, the driver takes one or both the hands off the steering wheel. In visual distraction, the driver takes his eyes off the road. In cognitive distraction, the driver loses his focus from driving. In order to reduce these distractions, various education programs are conducted. Texting on phone, messaging are the different ways of distractions. Here the dataset which is used is collected from a competition conducted by State Farm on Kaggle [1].Our main task is to classify the images based on various classes as b0:safe driving, b1:texting with right hand, b2:talking with right hand on phone, b3-texting with left hand on phone, b4-texting with right hand on phone, b5-talking with right hand, b6-drinking, b7-reaching behind, b8-doing makeup, b9-talking to the person at the backseat. Here we are making a comparison of the accuracies obtained using various models where the feature vectors are extracted using traditional methods [2] and the advanced techniques [3].

II. DESCRIPTION OF MODELS

i. Traditional Methods

The traditional techniques consist of Histogram of Oriented Gradients (HOG) [4], Speeded up Robust Features (SURF) [5] and Scale Invariant Feature Transform (SIFT) [6] descriptors. HOG descriptor makes use of feature vectors for object detection using a sliding window detector whose representation is done in the form of histogram. These features that are extracted from HOG are passed to the Support Vector Machines (SVM) [8] using the Bag of Words (BoW) [12]. The Bag of Words model consists of two parts. First, we create BOF descriptors by obtaining set of bags for particular features. The second step is to create histograms with the help of the bags containing the set of features obtained from the first step and these histograms are used to classify an image and video.

ii. Basic CNN [7]

It consist of various layers like convolutional layer, max pooling layer and at the end, a softmax layer which is used for classification. The convolutional layer is used to extract various features whereas the max pooling is to reduce the size of image which in turn helps for faster execution. The problem of overfitting because of limited number of input images is solved by using dropout layer.

iii. DeepCNN

The basic difference between CNN and DeepCNN is that DeepCNN has multiple layers of CNN and rest everything remains same. As the size of the dataset is very small we require fine tuning of various models like VGG-16, AlexNet, ResNet-152 [9]. For fine tuning AlexNet and VGG-16, last two FC layers were commenced with random weights. On the other hand, rest of the layers were set on the basis of pretrained AlexNet and VGG-16. As the number of classes is 10, the output depth of the last FC layer was set to 10. On the other hand, for fine tuning of Resnet-152, the FC layer was set to a depth of 10 and random weights in the FC layer were initialized, whereas the rest of the layers were prepared with the assistance of weights of the pretrained ResNet-152.

iv. Methods of Extracting Features

The feature extraction from AlexNet and VGG-16 was done with the help of activation functions like ReLu6 [11], ReLu7 [11], FC6 and others which are used to create non linearity, whereas for extracting features from ResNet-152, we make use of average-pooling layers which is used to reduce the number of parameters.

v. Inception-V4 [10]

It is a pretrained model used for image classification with the help of Tensorflow [21] as the backend. Previously, to get better performance CNN just tried to increase the number of layers in the hope to get better accuracy but now, it helps in the development of CNN classifiers.

vi. K-Nearest Neighbour (KNN) [14]

KNN is said to be the simplest algorithm in machine learning. The feature vectors are grouped together to form clusters based on the distance between them.

III. RESULTS AND DISCUSSION

i. Traditional Methods

After using various cell sizes to extract features using HOG, it was observed that the cell size 24x24 obtained the highest accuracy of 33.2%. Here is a table of various cell size and their respective accuracies:-

Cell size	Accuracy
8x8	27.2
16x16	30.5
24x24	33.2
32x32	29.9

A combination of HOG and BoW is used in which Principal Component Analysis (PCA) [13] is applied on the features resulting in 1287 features (a part of the extracted features). The accuracy using HOG and BoW obtained had a lower accuracy than only HOG, which concludes that it is not essential to use combined features as it yields a comparatively lower accuracy.

ii. Basic CNN

As you refer the graph of paper [2] of basic CNN, we can see that the training accuracy increased quickly to 99% which clearly shows the problem of overfitting. On the other hand, the validation accuracy remained same at 74%. Also when the confusion matrix in the paper [2] is observed, the performance of the classification model is known and hence it is concluded by observation that this model is very poor at predicting the c9 image (talking to passenger).

VGG-16 model

As seen in the paper [3], the problem of overfitting in basic CNN is solved by the VGG-16 model by making changes in the dropout rate. The training of the model is done by setting the value of learning rate, epoch and dropout probability as 0.001, 15 and 0.8 respectively. The dropout of 80% will reduce the 80% of neurons and thus reducing overfitting. The plot for both training and validation is done giving an accuracy of 85% in validation and loss of 0.45. Here when we look at the values in the confusion matrix from paper [3], it is observed that VGG is bad at predicting c0 class (safe driving). It is assumed that incorrect focus features is leading to this misclassification of features (e.g.-opening of driver's mouth). In order to reduce this problem, KNN can be used instead of focusing features incorrectly.

VGG-16+INCEPTION V4+KNN

When KNN is used, the previously faced problems are solved which increases its accuracy. So it is seen that accuracy increased to 88% and a loss of 0.45 was obtained. When the confusion matrix is analysed, it is learnt that the misclassification in the previous model has been reduced to a great extent.

iii. Deep CNN

For deep CNN, various parameters like learning rate (η), weight decay (ƛ) and batch size (B) with the value of momentum as constant (9x10 -1) need to be set. In AlexNet, there was a serious issue of overfitting in the training set which led to a decrease in testing error [2]. Even an attempt was made to increase the batch size and the dropout probability but it failed to improve testing accuracy. The accuracy of 72.6% was observed to be highest with the value of learning rate $\eta = 9\times10{-}4$ for first 10 epochs, whereas for next 10 epochs to gain the highest accuracy, the learning rate was $7\times10{-}4$ and for the last 10 epochs, the learning rate was $5\times10{-}4$ with the weight decay $\lambda = 1\times10{-}5$ and the batch size B = 50 remaining constant throughout the experiment. In VGG-16 the batch size was fixed to 2. The best accuracy was obtained for a learning rate of $\eta = 5\times10{-}6$ for 30 epochs, $\eta = 3\times10{-}5$ for 15 epochs and $\eta = 1\times10{-}5$ for 10 epochs. Along with using these learning rates, the value of weight decay was chosen as $5\times10{-}5$ which resulted in an accuracy of 82.5%. The training and the validation's fine tuning graphs are shown in [2]. In ResNet-152, the batch size was reduced to increase the accuracy and it was found that the highest accuracy of 85% was obtained with batch size 2. With the batch size of 2, the learning rate of $\eta = 1\times10{-}3$ for the first 10 epochs, $\eta = 5\times10{-}4$ for the next 10 epochs, $\eta = 1\times10{-}4$ for next 5 epochs and $\eta = 5\times10{-}5$ for the next 5 epochs was found to have the lowest accuracy. The weight decay $\lambda = 5\times10{-}5$ helped to obtain better accuracy and on the other hand, fine tuning helped to reduce the error rate in both training as well as testing data.

Table 2. Results [2]

Model	Accuracy
AlexNet	72.6%
VGG-16	82.5%
ResNet-152	85%

iv. Accuracy Table

Here is a table of various architectures and their models along with their accuracy

Table 3. Results using all methods [2-3]

Architecture	**Model**	**Accuracy**
CNN	Basic CNN	74.83
AlexNet	Deep CNN	72.6
VGG-16	Deep CNN	82.5
ResNet-152	Deep CNN	85
AlexNet	SVM	70
VGG-16	SVM	78.5
ResNet	SVM	84.6
ResNet	Softmax as a classifier [15]	85
VGG-16	Softmax as a classifier	82.5
AlexNet	Softmax as a classifier	72.6

VGG+Inception-v4+KNN	Basic CNN	**88.65**

IV. Proposed Methods

If we consider a video that is being analysed where the driver is distracted, we can store all the frames in that video as a collection of images. We could use RNN for such use cases where if the action of the user is being repeated across a number of frames, we could classify it as a distraction. The main reason for using RNN is it can store data in its memory and helps in analysing sequential data. For instance, if the driver is not focused on driving for a long time then it could be classified as a distraction. We could set a threshold time for being distracted and if the time for which the driver is distracted is greater than the threshold time then this will be detected as a distraction. For e.g. if the threshold is 3 seconds and if the driver is just looking at the phone for a second, then it cannot be classified as a distraction but if the driver is looking for a time greater than or equal to 3 seconds, then it is classified as a distraction. Also, Machine Learning models can be used to improve the accuracy of traditional models [16-20].

V. Conclusion

In in paper [2], the highest accuracy of 85% was obtained by ResNet-152 architecture whereas in paper [3], it can be seen that VGG-16 when used with INCEPTION V4 and KNN gave an accuracy of 88.65% which is higher than the accuracy of the system using ResNet-152. The reason for this is KNN helps in removing the noise as well as outliers and as a result it increases the accuracy.

References

1. StateFarm, "State Farm Distracted Driver Detection". [Online]. Available:https://www.kaggle.com/c/state-farm-distracted-driver-detection [Accessed: 16-Aug-2018].
2. Murtadha D Hssayeni, Sagar Saxena, Raymond Ptucha, Andreas Savakis,"Distracted Driver Detection: Deep Learning vs. HandcraftedFeatures"[Online].Available:http://www.ingentaconnect.com/contentone/ist/ei/2017/00002017/00000010/art00004?crawler=true&mimetype=application/pdf [Accessed: 20 August 2018]
3. Liuming Zhao, Taiming Zhang, Lingzi Guo "Classification Models of Driving Distraction: Analysis and Comparison"[Online]. Available: http://cs229.stanford.edu/proj2017/final-reports/5241763.pdf [Accessed: 20th August 2018]
4. Histogram of Oriented Gradients. [Online]. Available:https://en.wikipedia.org/wiki/Histogram_of_oriented_gradients [Accessed: 8-Sept-2018].
5. Speeded Up Robust Features [Online]. Available:https://en.wikipedia.org/wiki/Speeded_up_robust_features [Accessed: 8-Sept-2018]
6. Thao Nguyen, Eun-Ae Park, Jiho Han, Dong-Chul Park, and Soo-Young Min "Object Detection Using Scale Invariant Feature Transform" Advances in Intelligent Systems and Computing book series (AISC, volume 238).
7. Samer Hijazi, Rishi Kumar, and Chris Rowen"Using Convolutional Neural Networks for Image Recognition." Cadence Design Systems, Inc.
8. Sunil Ray (13th september 2017) Understanding Support Vector Machine algorithm from examples (along with code). [Online]. Available:https://www.analyticsvidhya.com/blog/2017/09/understaing-support-vector-machine-example-code/
9. Siddharth Das (16th November 2017)CNN Architectures [Online]. Available: https://medium.com/@sidereal/cnns-architectures-lenet-alexnet-vgg-googlenet-resnet-and-more-666091488df5 [Accessed: 8-Sept-2018].
10. Christian Szegedy, Sergey Ioffe, Vincent Vanhoucke. 23 August 2016. Inception-v4, Inception-ResNet and the Impact of Residual Connections on Learning. [Online]. Available:https://arxiv.org/pdf/1602.07261.pdf
11. Abien Fred M. Agarap "Deep Learning using Rectified Linear Units (ReLU)" arXiv: 1803.08375v1 [cs.NE] 22 Mar 2018
12. Jason Brownlee (9th October 2017)A gentle introduction to the Bag of Words model [Online]. Available: https://machinelearningmastery.com/gentle-introduction-bag-words-model/
13. Ian T. Jolliffe and Jorge Cadima (2015) "Principal component analysis: a review and recent developments" [Online]. Available: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4792409/
14. Gongde Guo, Hui Wang, David.A.Bell, Yaxin Bi "KNN Model-Based Approach in Classification" Researchgate.
15. Haihan Lan (13th November 2017)"The Softmax Function, Neural Net Outputs as Probabilities, and Ensemble Classifiers" [Online]. Available:https://towardsdatascience.com/the-softmax-function-neural-net-outputs-as-probabilities-and-ensemble-classifiers-9bd94d75932
16. P.Kanani, V.Kaul, K.Shah," Hybrid PKDS in 4G using Securred DCC", International Conference on Signal Propagation and Computer Technology (ICSPCT) 2014, pp- 323-328.
17. M. Padole, P. Kanani, "Textimage Ciphering", 2nd International Conference for Convergence in Technology (I2CT), 2017, pp - 926-930.
18. Y. Doshi, A. Sangani, P. Kanani, M. Padole, "An Insight into CAPTCHA", International Journal Of Advanced Studies In Computer Science And Engineering IJASCSE Volume 6, Issue 9, pp. 19-27. 2017.
19. T. Reddy, J. Sanghvi, D. Vora, P. Kanani, "Wanderlust: A Personalised Travel Itinerary Recommender", International Journal of Engineering Development and Research IJEDR, Volume 6, issue 3, pp.78-83. 2018
20. P. Kanani, K. Srivastava, J. Gandhi, D. Parekh, M. Gala, "Obfuscation: maze of code**",** Proceedings of the 2nd International Conference on Communication Systems, Computing and IT Applications (CSCITA), IEEE (2017), pp. 11-16.
21. Tensorflow, "Tensorflow Guide". [Online]. Available: https://www.tensorflow.org/guide/. [Accessed: 10-Sept-2018]

VyomUnadkat"A Review Approach For Detecting Distracted Driver "International Journal of Engineering Science Invention (IJESI), vol. 07, no. 10, 2018, pp 28-31