DISTRACTED DRIVER BEHAVIOUR RECOGNITION WITH COMPUTER VISION USING DEEP CONVOLUTIONAL NEURAL NETWORK IN REAL WORLD APPLICATION

M. Ananthi [1] and Bharathram. P[2] Rahul Narayanan. L[3]

[1]AP,Department of Information Technology, Sri Sairam Engineering College.
[2]Department of Information Technology, Sri Sairam Computer Engineering College.
[3]Department of Information Technology, Sri Sairam Engineering College.

ABSTRACT

Distraction of drivers while driving on roadways results in the death of 1.2 million people, approximately every year around the globe. Even though there are several improvements made in road and vehicle design, the total number of accidents is higher. In 2015, 3,477 people were dead and 3,91,000 were injured during motor vehicle crashes associating distracted drivers. Our paper is aimed to prevent and reduce the rate of motor vehicle accidents that are caused by human errors and distraction during driving. We study the different postures of the driver by means of the hand localization, skin segmentation and facial data. In our paper, we propose a reliable deep-learning CNN model that attains 92% accuracy.

KEYWORDS

*Deep learning, Convolutional Neural Networks, Spatial pooling, Image processing, Transfer learning***.***

1. INTRODUCTION

According to the National Highway Traffic Safety Administration (NHTSA) report, 23% to 26% of road accidents are caused due to distraction of drivers. Driver distraction during driving can increase the risk of accidents by 4 to 6 times. It is noted that different distraction activities cause accidents with different rates of risk. Therefore, the study of driver distraction detection is of great magnitude. Due to the rapid development of deep learning neural networks, driver distraction detection has been an active research topic in the computer vision field in the recent years. Drivers do some other activities rather than focusing on driving which causes a lot of accidents.

The number of car accidents has increased because of distracted driving, since the last few years. 3,477 deaths and 391,000 injuries cases has been confirmed by NHTSA in 2015 due to distracted driving[9]. Presently, more companies are starting to work with a new system called (ADAS) Advanced Driver Assistance Systems by creating techniques to alert the driver when distraction activities occur. From this deep learning takes part of ADAS techniques.

D. Angeline, A. Arul et al.(Eds): ICICI-2021
pp. 13-21, 2021. IJCI – 2021 DOI:10.5121/ijci.2021.100203

In our paper, we explore the possibilities of training a system that can detect different postures of drivers being distracted. Drinking, texting, talking on phone, operating radio, etc are some of the common activities that the drivers indulge during driving. It is the prior responsibility of the automobile manufacturers to provide a system that prevents drivers from being distracted. Detecting distracted drivers from real time video feed from the dashboard camera itself is one of the key parts of this paper.

2. LITERATURE REVIEW

In this section the recent trends and the related works related to the problem are reviewed. [7] An SVM-based model was induced to detect the use of mobile phone while driving vehicles. The dataset used consists of frontal face images of the driver.[6] Another SVM based model which used the dataset of driver images taken from highways traffic light cameras. This model uses facial data only. [3] Osman proposes a bi-level hierarchical classification methodology using machine learning to identify the different types of secondary tasks drivers are engaged in using their driving behaviour parameters. This method uses only three parameters to detect the driver fatigue so the overall accuracy ranges from 55% to 79% only.

[2]Akshada proposed a system that detects the drowsiness of the driver while driving in the roadways. This system uses the state of the eye and the facial data which includes Principal Component Analysis (PCA) and Eigen face approach. [5]Simonyan. K investigates with the convolutional network depth and the effect on its accuracy with the large-scale image recognition setting. And evaluation of ConvNets for increasing depth using an architecture with very small (3×3) filters.[10]Kurylyak proposes a cascade of classifiers based on Haar-features to detect the eyes closure and opening.[12] Jia Deng illustrates the ImageNet by recognition of objects, classification of images and object clustering for 3.2 million images in total. [1] Abdul proposed a system with vanilla CNN which locates the facial landmarks like mouth corners, nose, eyes. The epochs used for this system is set to 5 which is very low and attains an accuracy of 75% in test run.

[8] Ralph Mbouna analysed the driver alertness by using the state of eye and head position. The images were taken from the dashboard camera. Eye index, eye pupil, head position was taken as visual parameters. [11] Murphy-Trivedi model focused on the ability to focus on fine head pose and coarse head pose that are well suited for unconstrained environments. [4]Koesdwiady proposed a structure with Convolutional Neural Network system VGG19 which has accomplished a test accuracy of 94%.

3. PROPOSED SYSTEM

Our paper instruments a classifier based on Convolutional Neural Networks to detect driver distraction and also detect the cause for the distraction. The direct video feed from the dashboard camera is obtained and the image is being extracted frame by frame. And the image obtained from the video is feed into the neural network and the corresponding category is recognized. In this paper, we also make an attempt to reduce the training time of the Convolutional Neural Networks while making sure that we do not diminish the accuracy of the classifier. In our attempt to reduce the training time, we experiment with the pre-trained weights that were obtained by training the Convolutional Neural Networks model on the ImageNet dataset.Refer figure 1.

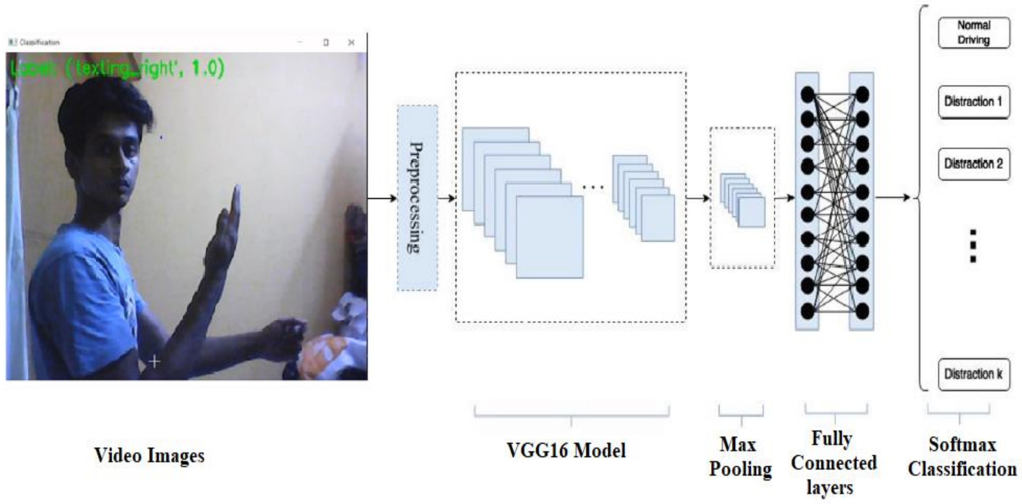

Figure.1 – Architecture Flow Diagram

A. Driver Behaviour Classification

In this paper we use the dataset provided by StateFarm [16]. The dataset consists of 22424 images of people classifying either driving safely or indulging in any of the nine below mentioned behaviours. The images in the dataset are RGB images of 640x480 pixels. The dataset was separated into 80% and 20% for training and testing respectively. The major advantage of this dataset is it consists of subjects from different ethnic backgrounds for example., Asian, Western, African, etc. The dataset consists of 10 classes of behaviours out of which the first one corresponds to the category of safe driving, while the remaining classes represents distraction while driving. These classes are some of the most common distractions while driving and can lead to accidents, loss of property and life.

B. Image Augmentation

The StateFarm image dataset [16] was used for our experiment. In the State Farm distracted drivers' dataset each of the image is a RGB coloured image of 640x480 pixels. The input layer of the model VGG16 takes an input of an image of size 3x224x224.Therefore the first step was to normalize and resize each image to the size of 224x224 pixels. Convolutional neural network requires a variety of diversified dataset to train the model effectively and to produce the results with high accuracy. To combat this challenge, we used the technique of data augmentation. For each epoch a new set of images were generated by generating various alterations on each image.

Table 1. Classification of distractions

CLASS	***DESCRIPTION***
Safe Driving	Safe driving is defined as the act of driving looking in the front with hands on the wheel while driving, as shown in Figure 1. And the driver should not be involved in any of the below activities.
Texting with Right hand when driving	This category defines the usage of mobile phone for texting using right hand of driver as shown in Figure 2.
Talking with Right hand on the phone	This category defines the act of using a mobile phone for making phone calls as shown in Figure 3.
Texting with Left hand when driving	This category defines the usage of mobile phone for texting using left hand of driver as shown in Figure 4.
Talking with Left hand on the phone	This category defines the act of using a mobile phone for making phone calls as shown in Figure 5.
Operating radio while driving	Operating radio is a category of distraction while driving. Figure 6 shows the mentioned act.
Drinking while driving	Drinking is also a category of distraction in which drivers frequently indulge in, as shown in Figure 7.
Reaching Behind while driving	In this category the driver reaches the back seat of the car by turning his/her back towards the steering wheel as shown in Figure 8.
Doing Hair or Makeup while driving	Correcting hair or doing makeup is another category of distraction. Figure 9 describes the above act.
Talking to passengers while driving	In this category the driver turns his face towards the passengers to speak with them as shown in Figure 10.

Figure. 2. Classification of Driver postures

C. Convolutional Neural Network

The idea proposed by LeCun of using Convolutional neural network has made a serious impact on the field of image classification and image detection. The Convolutional neural network introduces a number of hidden layers added to the model, thereby reducing the dimensions of the image and enabling the model to extract the image features in reduced dimensional images. The conventional neural models consist of 2 modules namely feature extraction of images and classification module based on the classes. Due to the different separate modules of the model the extractor module is only able to extract a certain significant feature based on the algorithm and is not efficient to extract the discriminating and trivial features of the image under different categories. The classification module then uses these features to classify the images. Convolutional neural network possesses a multistage processing layers which are able to extract these discriminating and significant features and then these features are transferred to the classification module to get the category of classification based on the classes. The CNN models are implemented using Keras API with TensorFlow in the backend.

D. Image Pre-processing

The mean value of the RGB pixels over all pixels were subtracted from each pixel value. Excluding the mean value of the dataset serves to centre the data. The subtracting of mean is executed so that the training model is involved in multiple weights and the process of adding weights to the initial inputs in order to fire activation which then activates the back propagation with gradients to train the model. Each feature has a similar range that of the feature, to prevent the gradients from accessing out of range. Convolutional neural networks also involve in sharing the parameters; hence the input is scaled in similar range, avoiding which the sharing may not occur. It is due to the fact of each image having a large value of weights focussed and while the other part is filled with lesser weights.

E. Transfer Learning

VGG16 is a Deep Convolutional Networks for Large-Scale Image Recognition. The architecture of VGG16 can be seen in Figure 3. The input to the ConvNet is an image of size 224 × 224 and is a RGB image. The image is then passed through layers of convolutional layers, where filters with a very less receptive field of 3 × 3 are used. The convolutional stride is fixed to 1 pixel of image. The spatial padding of convolution layer input is 1 pixel for 3 × 3 thus preserving the spatial resolution. Five max-pooling layers carry out the spatial pooling, which then follow some of the convolutional layers and not passing through all the convolutional layers.

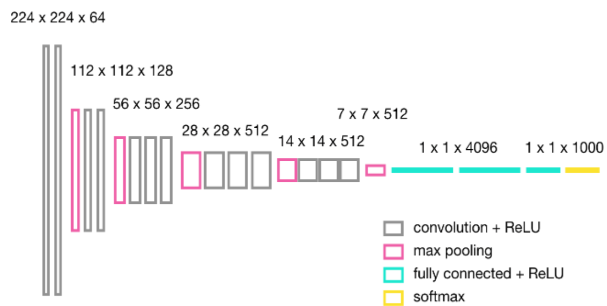

Figure.3. VGG16 Layered Net

4. EXPERIMENT RESULTS

Once the training was completed, the test dataset was run through both the models and the results obtained were recorded. A detailed analysis of the results obtained is explained below in Table II.

A window size of 2 x 2 pixels is used for Max-pooling and stride size of 2 is used. The convolutional layers are fully connected by three layers, the first two layers have 4096 channels and the last layer executes 1000-way classification. All hidden layers are equipped with the rectification nonlinearity and the final layer is the SoftMax layer. Pre-trained weights which were obtained by training the VGG16 model on the ImageNet [10] database is used to initialize the weights in one of the models. Pre-trained weights that were obtained by training the VGG16 model on the ImageNet were used to initialize the weights of one of the models, in the other model, random initialization of weights is used. Hence the final layer was neglected and replaced with a fully connected SoftMax layer with 10 channels to perform the 10-way classifications. The remaining model was retained as it is. Although the original model had 1000 channels signifying the 1000 categories which were aimed to classify. Here only 10 classes are being targeted.

Table 2. Depicts the accuracy obtained for each class

Scenario	Total Samples	Correct Predictions	Incorrect Predictions	Accuracy (%)
Safe Driving	622	619	3	99.52
Texting using Right Hand	565	565	0	100
Talking on phone using Right hand	579	578	1	99.83
Texting using Left Hand	587	580	1	100
Talking on phone using Left hand	581	572	6	99.83
Drinking	578	577	4	99.86
Operating the radio	501	500	1	99.80
Reaching Behind	581	577	4	99.31
Talking to Passenger	534	528	6	99.88
Hair and Makeup	478	476	2	99.58

The model used with pre-trained weights required only very few epochs to converge, despite the model being a very deep model. ImageNet weights can also be used because of the large size which nearly contains around 1. 2 million images.

5. Conclusion

Deep learning using Convolutional Neural Networks has been extensively used in image classification, image detection, etc. In this paper, we use VGG16 for detecting distracted drivers and also identifying the cause of their distraction using Convolutional Neural Networks. Thus, the results suggests that our system can be used to detect the driver state. The model not only identifies the basic distraction but also their cause of distraction automatically within the mentioned 10 classes. The mentioned system was shown to be efficient and workable with an accuracy of more than 99%. The proposed system can be used to effectively monitor the posture of the driver while he is driving. These systems when installed are useful in trying to prevent any accidents due to distraction from the driver by raising warnings whenever the driver gets distracted. A significant reduction in training time was achieved by diminishing the accuracy of our classification models. In future work as an extension, more categories of distraction classes can be brought in to the model. Even considering certain specific scenarios such as detecting drowsiness among drivers may also provide an opportunity to widen the scale of the work which were not targeted in the present work.

REFERENCES

[1] Abdul Jamsheed. V, Janet. B, U. Srinivalsu, “Real Time Detection of driver distraction using CNN”, Proceedings of the Third International Conference on Smart Systems and Inventive Technology, (ICSSIT - 2020).

[2] C. Akshada, D. Pramila, D. Ankita, et. al,” Driver Drowsiness Detection and Alarm System”, International Journal of Research in Electronics and Computer Engineering, IJRECE Vol. 7, Dec 2019.

[3] A. Osman, Mustafa Hajij, Sherif Ishak, SogandKarbalaieali, “A Hierarchical Machine LearningClassification Approach for Secondary Task Identification from Observed Driving Behavior Data”, Accident Analysis & Prevention – Elsevier. Vol-123, 2018.

[4] Koesdwiady. A, Bedawi S.M, Karray. F, “End-to-End Deep Learning for Driver Distraction Recognition”, Publishedat International Conference on Image Analysis and Recognition 2017.

[5] Simonyan. K, and A. Zisserman,” Very deep convolutional networks for large-scale image recognition”, Published at ICLR 2015.

[6] Y. Artan, R. P. Loce, O. Bulan, and P. Paul, “Driver cell phone usage detection from HOV/HOT NIR images”, Proceedings of the 2014 IEEE Conference on Computer Vision and Pattern Recognition Workshops - CVPRW, June 2014.

[7] R. A. Berri, A. G. Silva, R. S. Parpinelli, and R. Arthur, “A pattern recognition system for detecting use of mobile phones while driving”, Proceedings of the 9th International Conference on Computer Vision Theory and Applications, VISAPP 2014, IEEE.

[8] Ralph Mbouna, Seong G. Kong, "Visual analysis of eye state and head pose for driver alertness monitoring." IEEE transactions on intelligent transportation systems, 2013.

[9] Hirshkowitz. M, “Fatigue, sleepiness, and safety: Definitions, assessment, methodology”. Sleep Medicine Clinics, 2013, Vol.8(2):183–189.

[10] Y. Kurylyak, F. Lamonaca, and G. Mirabelli, “Detection of the eye blinks for human’s fatigue monitoring”, IEEE International Symposium on Medical Measurements and Applications Proceedings - MeMeA, 2012.

[11] Murphy-Chutorian, E., & Trivedi, M. M, “Head pose estimation in computer vision: A survey”, IEEE transactions on pattern analysis and machine intelligence, 2009.

[12]Jia Deng, Wei Dong, Richard. S and Li Fei-Fei, “ImageNet: A Large-Scale Hierarchical Image Database”, IEEE Conference on Computer Vision and Pattern Recognition, 2009.

[13] LeCun, et al. "Gradient-based learning applied to document recognition. " Proceedings of the IEEE 86, Vol no. 11 2278-2324, 1998.

[14] Chollet F, others. Keras [Internet]. GitHub; 2021. Available from: https://github. com/fchollet/keras

[15] TensorFlow GitHub; 2021. Available from: https://github. com/tensorflow/tensorflow

[16] State Farm Distracted Drivers Dataset, https://www.kaggle.com/c/state-farm-distracteddriver-detection/data

AUTHORS

M. Ananthi
AP,Department of Information Technology,
Sri Sairam Engineering College.
ananthi.it@sairam.edu.in

Bharathram. P
Department of Information Technology,
Sri Sairam Computer Engineering College.
bharathpadmanaban2000@gmail.com

Rahul Narayanan. L
Department of Information Technology,
Sri Sairam Engineering College.
rahulatoff@gmail.com