IJARCCE

ISSN (Online) 2278-1021
ISSN (Print) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering
ISO 3297:2007 Certified
Vol. 7, Issue 4, April 2018

Distracted Driver Detection using CNN and Data Augmentation Techniques

Vasanti Sathe[1], Neha Prabhune[2], Anniruddha Humane[3]

Pune Institute of Computer Technology, Department of Computer Science, Pune, India[1,2,3]

Abstract: One of the critical problems prevailing in India is the deaths caused by road accidents. Almost 80% of the accidents are caused by the inattentiveness of the driver. Usage of mobile phones, talking to passengers, reaching behind to grab something and drinking while driving are some of the reasons due to which driver may lose attention. Distractions are of numerous types, out of which we focus on the manual distraction which is based on the posture of the driver. In this paper, we propose a system where we make use of Convolutional Neural Networks and data augmentation techniques. Data augmentation techniques are used to increase the variability of the dataset and decrease overfitting. We have used the first publicly available dataset as input for our model. Our aim is to categorize a test image into one of the nine distinct distracted states of the driver that we have considered. Conclusively, the experimental analysis has shown that applying data augmentation techniques, the proposed model gives better results.

Keywords: Convolutional neural networks, data augmentation techniques, deep learning methods, distracted driver.

I. INTRODUCTION

Distracted driving is any activity that distracts a driver from the task of safe driving. The activities include talking or texting on mobile phone, talking to a passenger, operating the radio or navigation system, reaching behind to grab something or drinking while driving. These tasks take driver's eyes off the road for a few seconds, but considering the speed at which they drive, they traverse a considerable distance. Consider a driver driving his car at the speed of 100 km/h. At this speed, he takes approximately 10 s to send a text and covers a distance of 280 m - blind. This is distance is equivalent to the distance covering the length of 12 tennis courts. Similarly, the distance covered by performing other such activities range from 120 m to 560 m. Also, the time span of these actions ranges from 4 s to 20 s. This seems very less but is enough for an accident to occur.

Distractions can be categorized into various types. Research on this subject is mainly categorized into three types, namely: manual, visual and cognitive. A situation where the driver takes his eyes off the road due to the presence of some visual representation away from the road is depicted by visual distractions. Studying these depends on the tracking of eyes and facial landmarks. Cognitive distraction means the driver is "mentally" distracted due to daydreaming or just lost in thought. Manual distractions take into consideration the posture of the driver and mostly, hand tracking.

Road accidents amount to a considerable percentage of deaths, most of which occur because of the distracted state of a driver. According to a survey conducted by Savelife Foundation, India in 2015, 400 lives are lost every day in road accidents. In 2015, 146,133 people were killed and 500,279 people have been injured. Analyzing these statistics, almost 17 people lose their lives per hour in these accidents. Also, these numbers are increasing at an alarming rate every year. Accident severity increased from 21.9% in 2015 to 31.4% in 2016. The above statistics prove that it is a critical social issue which should be addressed.

Classification of an image into its proper class is the structure of our problem. In image classification problems, an input image is taken for processing and output is given in the form of probability of the final classes .Processing is done by the convolutional neural network. In simple words, given an input image, a CNN finds basic features like edges and curves and builds to find more features through a series of convolutional layers. Considering all these features, it outputs the final class of the image. For training purpose, we are using images of particular drivers, which increase the chances of overfitting. To overcome this, we have implemented the concept of data augmentation. Apart from basic data augmentation techniques like a blur and sharpen, we have implemented concepts like parts-based and class-based augmentation. These concepts generalize the input dataset, and we get more generalized results.

In this paper, we demonstrate three different techniques:

1. Simple CNN on image,
2. CNN applied on parts-based augmented data,

IJARCCE
ISSN (Online) 2278-1021
ISSN (Print) 2319-5940
International Journal of Advanced Research in Computer and Communication Engineering
ISO 3297:2007 Certified
Vol. 7, Issue 4, April 2018

3. CNN applied on class-based augmented data.

Section 2 gives a brief about related work. Proposed model in Section 3 gives a detailed explanation of the developed model. In Section 4, the experimental analysis shows that the data-augmented images when trained through a CNN yield more precise results as compared to a simple CNN model. Section 5 and 6 gives the definitive conclusion and future scope of our project.

II. RELATED WORK

Formerly, researchers have done some work in the problem domain of distracted driver detection. Most of these researches are limited to cognitive and visual distractions. Cognitive distraction focuses on the mental instability of the driver while visual distraction focuses on the "eyes off the road" behaviour of the driver.

In [2], Yuang Liang et. al. have conducted a study on detecting cognitive distractions in real time. They have utilized the concept of Support Vector Machines(SVM). SVM is a very popular and efficient data mining method. The classification was based on specifications like eye movement and driving behaviour. Yuan Liang et. al. in [3] have studied the same using Bayesian Networks. It is a data mining method based on a probabilistic framework. The history of drivers behaviour and eye movements were the features.

In this paper, we have presented a model which works on manual distraction. A similar study was conducted by Arief Koesdwiady et. al.[4]. Their work compares two frameworks, namely: VGG-19 model and XGBoost. They have utilized the concept of transfer learning. Features were extracted by the use of a pre-trained model named VGG-19. XGBoost is a state-of-the-art framework. In [5], Yehya Aboulnga et. al. have demonstrated a weighted ensemble of Convolutional Neural Networks(CNN) for detecting manual distractions. They have trained their model on five different sets of input. Transfer learning is implemented by employing the trained model of ImageNet. The obtained five classifiers are weighted and combined to give the final output.

The research most closely to our work is done by Jaco Cronje [1]. They have used the concept of data augmentation [6][7] and no pre-trained model was used for execution. In this paper, we have implemented some basic data augmentation techniques like shift, scale, blur as well as some advanced techniques. Our aim is to compare the performance of a simple CNN on raw images and CNN model on augmented images. Data augmentation is introduced to generalize the data to avoid overfitting. D.T.Mane et. al. in [8] have presented a deep summary about convolutional neural networks(CNN) and its various applications in domains of Computer vision, Object detection and natural language processing.

III. PROPOSED MODEL

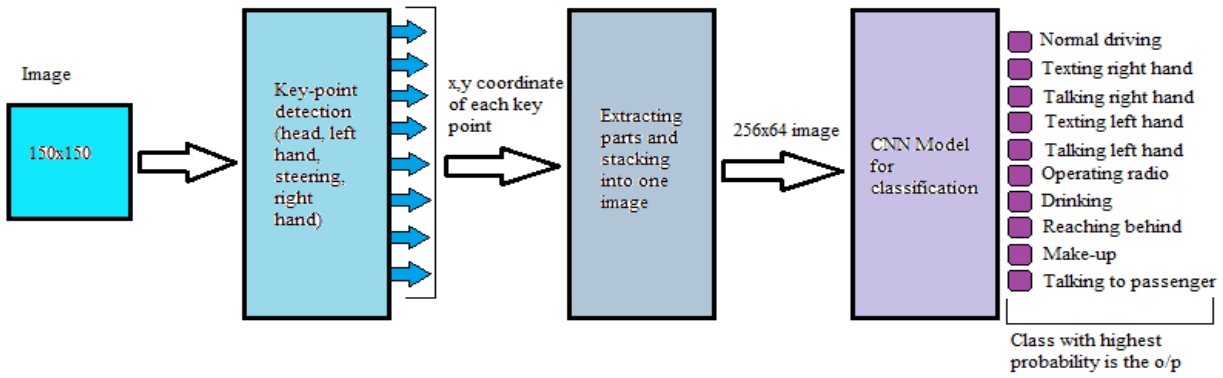

Fig 1.1 Architecture of the proposed model

The proposed model is represented in the form of a block diagram. This work is composed of three main parts, namely: key point detection, parts-based data augmentation and class-based data augmentation.

Key-point detection

After looking at the image, it was observed that there are some attributes of the image, which are important for detecting the drivers' behaviour. The key regions to be focused on are the head, the left hand, the steering wheel and the right hand. Training is performed on 18000 images. The location of these four attributes is labelled manually using an OpenCV python code. A list of image names and the eight coordinates are maintained in a text file. Using the images

IJARCCE
ISSN (Online) 2278-1021
ISSN (Print) 2319-5940
International Journal of Advanced Research in Computer and Communication Engineering
ISO 3297:2007 Certified
Vol. 7, Issue 4, April 2018

and the text file, a CNN model is trained to detect these four key points. It is not necessary for the predictions to be highly accurate because an error of few pixels will still work for extracting the required region. The CNN architecture follows. To begin with, a 3 channel input image is reduced to 150x150 pixels. One convolutional layer follows with a filter size of 3x3 and 32 filters. To add nonlinearity to this network, a rectified linear unit activation function is applied. Next, three convolutional layers follow with each having a filter size of 3x3 and 64 filters. Finally, three fully connected layers are added; first with 64 units, second with 64 units and final output with 8 units.

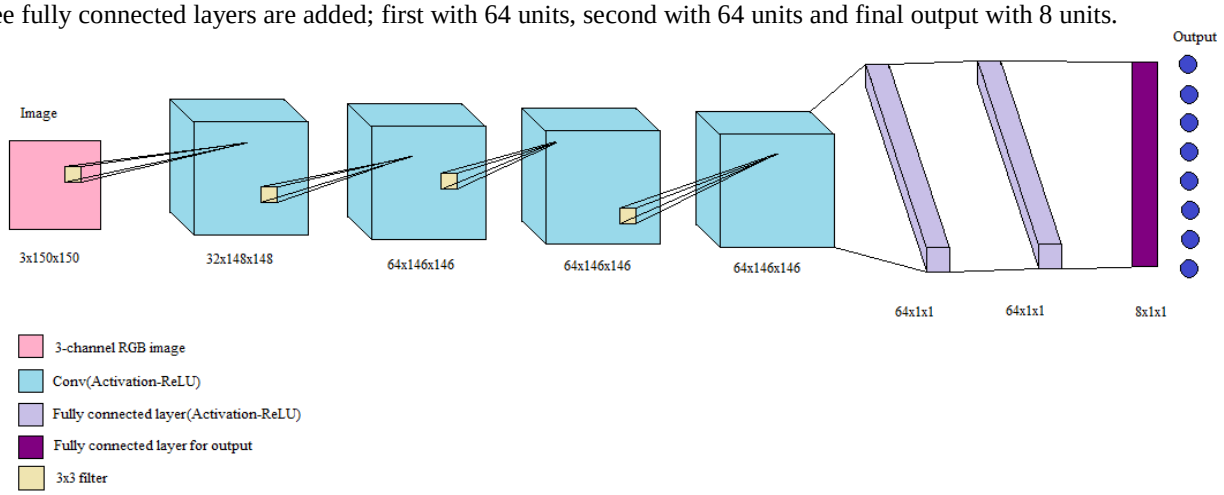

Fig 1.2 CNN Architecture for key-point detection

These outputs are the x-y coordinates of each of the attributes. Each output is in the range [0, 1] and indicates the position of the point. Training was performed for 50 epochs with a batch size of 16. Initial learning rate was 0.001 which was later reduced to 0.0001.

Parts-based data augmentation

The output from the previous subsection is taken as an input for this part. Centred on the coordinates of each point in the output, a 64x64 key region is extracted from the original 640x480 image. This is done so that CNN makes a decision only on the required features of the image. We obtain such four 64x64 regions around each attribute. Stacking all the four regions, a 256x 64 image is formed.

Fig 1.4 shows the newly formed augmented image. A CNN is again trained on this newly formed dataset of 18000 images. The name parts-based indicates that the original image is distorted and combined into one image, partwise. This helps in removing the regions comprising window, seat and shirt of the driver which are not essential to classify the image.

Fig 1.3 Original 640x480 image

Class-based data augmentation

The CNN model on parts-based augmented images reduces overfitting. However, there is still a chance that the CNN may learn to detect a specific driver. Accordingly, we made use of the concept of class-based data augmentation. This technique interchanges or shuffles the 64x64 regions within the same class of images. In this way, a new dataset is formed. Fig 1.5 depicts the class-based augmented image. This technique allows CNN to perceive the action of the driver rather than the driver itself.

IJARCCE
ISSN (Online) 2278-1021
ISSN (Print) 2319-5940
International Journal of Advanced Research in Computer and Communication Engineering
ISO 3297:2007 Certified
Vol. 7, Issue 4, April 2018

Fig 1.5 Class-based augmented image

Fig 1.4 Parts-based augmented image

This work does not focus on the structure of CNN model. It focuses on demonstrating It focuses on demonstrating that data augmentation techniques used reduces overfitting and increases the accuracy of classification. The simple CNN model to train these images follows. The input is a 3 channel 256x64 image. Next, come two convolutional layers of filter size 3x3 and 32 filters. A max-pooling layer follows with a stride of 2x2. Two convolutional layers follow of filter size 3x3 and 64 filters. Again, a max-pooling layer follows with a stride of 2x2. Finally, two fully connected layers are added; one with 64 units followed by a dropout of 50 percent and the final output layer of 10 units. Each entry in these 10 units gives the probability of the image belonging to the respective class. The network is trained for 15 epochs with a fixed learning rate of 0.01 and batch size 32.

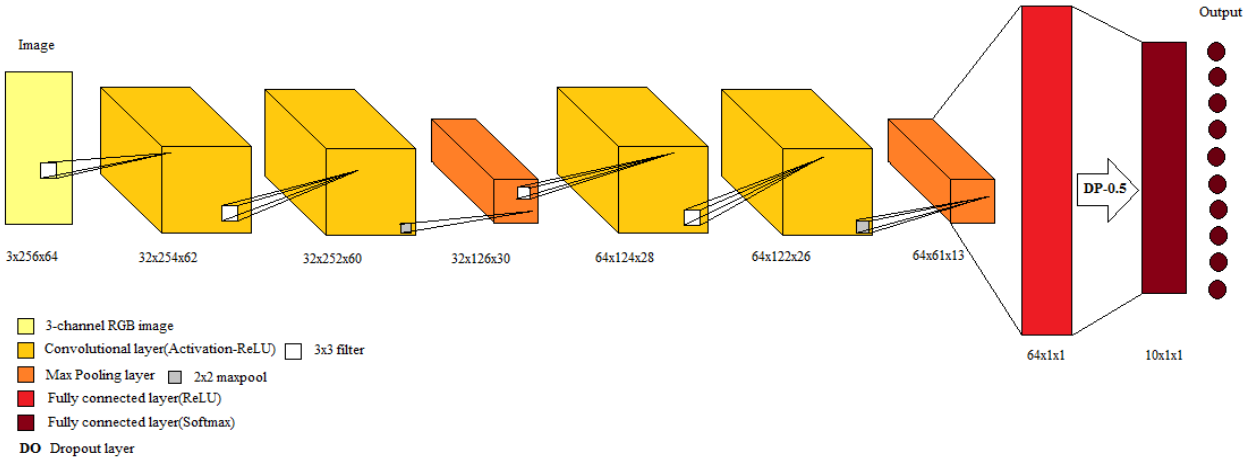

Fig 1.6 Simple CNN architecture

IV. EXPERIMENTAL RESULTS

Original dataset contains 22,000 images of different drivers. Out of those 22,000, 18,000 images were used for training our models while the remaining 4000 images were used for testing it. Our CNN was trained for 10,000 iterations (8 epochs), keeping the batch size 32 and a learning rate of 0.01. CNN model trained on parts-based augmented images and class-based augmented images were tested on these images. Fig 1.7 and Fig 1.8 graphically depict the loss values while training the models. It is very clear from the graphs that training and validation loss is lowest for the CNN model trained on class-based augmented images. This fact makes it clear that the model is being trained well and is able to generalize the data.

IJARCCE
ISSN (Online) 2278-1021
ISSN (Print) 2319-5940
International Journal of Advanced Research in Computer and Communication Engineering
ISO 3297:2007 Certified
Vol. 7, Issue 4, April 2018

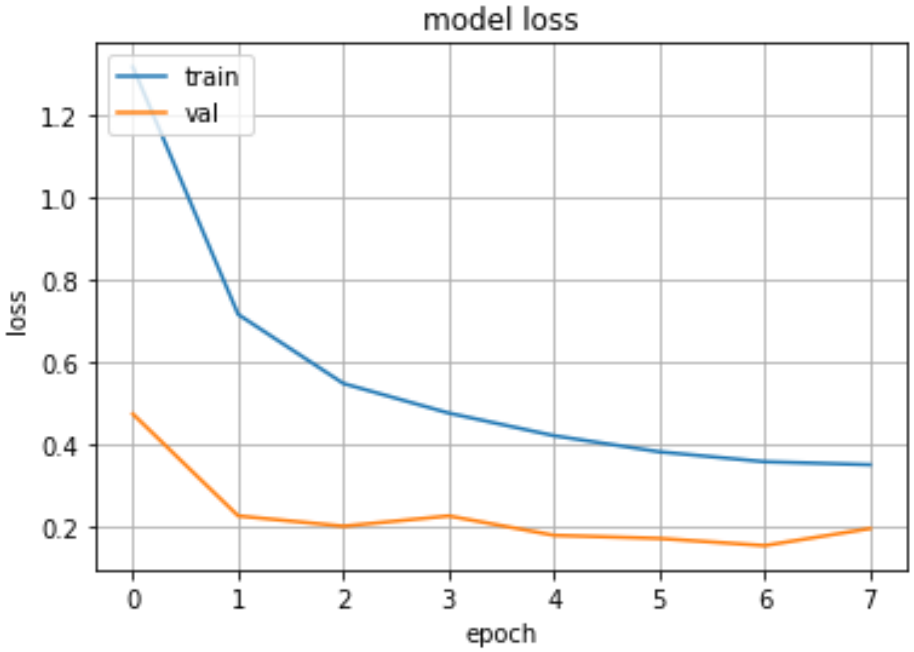

Fig 1.7 Graph of loss vs num_of_epochs for training performed on parts-based augmented data

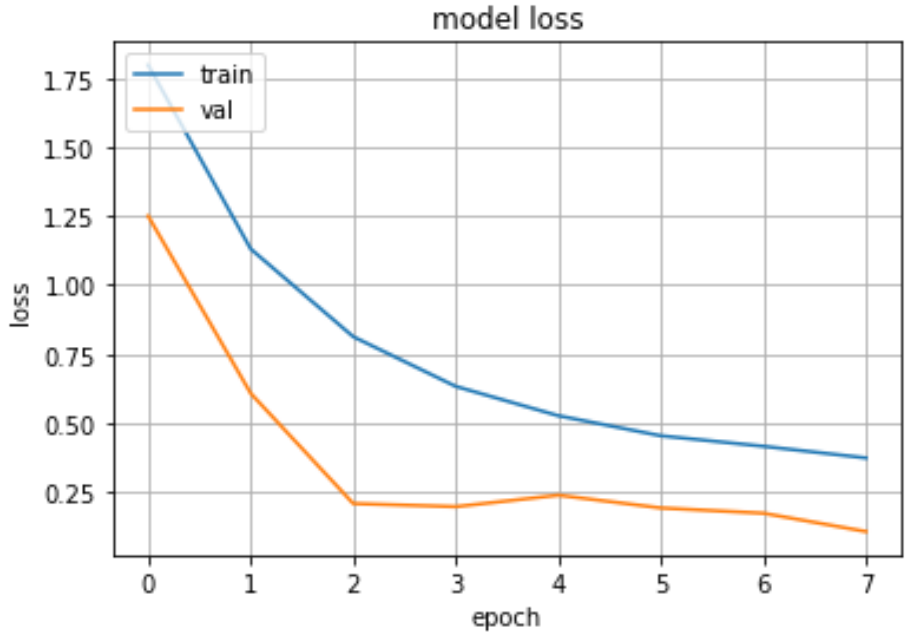

Fig 1.8 Graph of loss vs num_of_epochs for training performed on class-based augmented data

Fig 1.9 gives a brief about the accuracy of the three models in percentage. The values in the testing column show a gradual increase in the accuracy from a simple CNN implementation to CNN on class-based augmented images. It can be noted that the main purpose of generalizing the model when a large dataset is not provided has been achieved. The CNN model which is implemented is a very small and rather simple. These results can also be improved by using a rather complex implementation of the network.

IJARCCE
ISSN (Online) 2278-1021
ISSN (Print) 2319-5940
International Journal of Advanced Research in Computer and Communication Engineering
ISO 3297:2007 Certified
Vol. 7, Issue 4, April 2018

Method	Training accuracy	Testing accuracy
Simple CNN	86.71%	91.05%
CNN with part based augmentation	88.65%	94.84%
CNN with class based augmentation	87.47%	96.72%

Fig 1.9 Table representing training and testing accuracies on three CNN models

V. CONCLUSION AND FUTURE SCOPE

This study was performed to understand the concepts of deep learning as well as the data augmentation techniques. Experimental results clearly depicts that applying these techniques has increased the accuracy of the model by decreasing overfitting. This approach is mostly useful when a dataset with low variability has to be processed through CNNs. Also, these types of augmentations can be used for other datasets where the classification can be made based on certain regions in an image.

Distracted driver detection systems can be installed in cars. A camera and dashboard can be installed in the car to monitor the drivers' driving. It could also act as an alert system by generating a beep sound whenever the driver is found distracted. These kinds of systems would be very useful for cab service providers like OLA and UBER. They can monitor their drivers in real time and have some statistics about his actions. Car rental service providers like Zoomcar can also use these systems to keep a check on their customers. We would like to make use of our model in real time system and deploy it. If possible, we would also like to create an Indian dataset and see how our model is compatible with that dataset.

REFERENCES

1. Jaco Cronje, Andries P. Engelbrecht: Training Convolutional Neural Networks with Class Based Data Augmentation for Detecting Distracted Drivers. Proceedings of the 9th International Conference on Computer and Automation Engineering (2017)
2. Y. Liang, M. L. Reyes, J. D. Lee.: Real-time detection of driver cognitive distraction using support vector machines. IEEE transactions on intelligent transportation systems, vol. 8, no. 2, pp. 340-350 (2007)
3. Y. Liang, J. Lee, M. Reyes.: Nonintrusive detection of driver cognitive distraction in real time using bayesian networks. Traansportation Research Record: Journal of the Transportation Research Board, no. 2018, pp. 1-8 (2007)
4. Arief Koesdwiady,Safaa M. Bedawi,Chaojie Ou,Fakhri Karray.: End-to-End Deep Learning for Driver Distraction Recognition. ICIAR 2017: Image Analysis and Recognition pp 11-18 (2017)
5. Yehya Abouelnaga, Hesham M. Eraqi, and Mohamed N. Moustafa.: Real-time distracted driver posture classification. arXiv:1706.09498 (2017)
6. I. Sato, H. Nishimura, K. Yokoi.: Apac: Augmented pattern classification with neutral networks. arXiv preprint arXiv: 1505.03229 (2015)
7. P. Y. Simard, D. Steinkraus, J. C. Platt. Best practices for convolutional neural networks applied to visual document analysis. In ICDAR, vol 3, pp. 958-962 (2003)
8. D. T. Mane, U. V. Kulkarni.: A Survey on Supervised Convolutional Neural Network and Its Major Applications. International Journal of Rough Sets and Data Analysis (IJRSDA), (2017)