Survey on Distracted Driver Detection Systems

Saurabh Takle[1], Shubham Desai[2], Sahil Mirgal[3], Ichhanshu Jaiswal[4]

[1-3]Department of Information Technology, Vidyalankar Institute of Technology, Mumbai, Maharashtra, India

[4]Assistant Professor, Department of Information Technology, Vidyalankar Institute of Technology, Mumbai, Maharashtra, India

ABSTRACT

This paper describes all the existing papers associated with Distracted Driver Detection Systems. In the survey paper, we explain all the different approaches to the problem of driver's distraction also comparing them to find out the best possible approach.

Keywords : Distracted Driver Detection, InceptionV3, VGG16, Resnet50, Image Classification, Computer Vision, Ensembling

I. INTRODUCTION

According to the World Health Organization (WHO) survey [9], 1.3 million people worldwide die in traffic accidents each year, making them the eighth leading cause of death, and an additional 20-50 million are injured/ disabled. As per the report of the National Crime Research Bureau (NCRB), [10] Govt. of India, Indian roads account for the most fatalities within the world. There has been an endless increase in road crash deaths in India. According to the report compiled by the ministry Transport Research Wing of India, there has been a 3.2% rise in road fatalities which corresponds to the death of 1,50,785 people across the country in 2016. A total of 4,67,044 road accidents have been reported by States and UTs in the year 2018, taking 1,51,417 lives and injuring 4,69,418 persons," the report said and driver error is the most common cause behind these traffic accidents. The number of accidents due to distracted drivers has been increasing for few years. The National Highway Traffic Safety Administration [8] is

an agency of the U.S. federal government that describes distracted driving as "any activity that diverts the attention of the driver from the task of driving" which can be classified into Manual, Visual or Cognitive distraction. As per the definitions of the Center for Disease Control and Prevention (CDC), cognitive distraction is essential "driver's mind is off the driving". In other words, even though the driver is in a safe driving posture, he is mentally distracted from the task of driving. He might be lost in thoughts, daydreaming, etc. Distraction because of inattention (taking eyes off the road), sleepiness, fatigue, or drowsiness falls into visual distraction class where "drivers' eyes are off the road". Manual distractions are concerned with various activities where "driver's hands are off the wheel". Such distractions include talking or texting using mobile phones, eating, and drinking, lecture passengers within the vehicle, adjusting the radio, makeup, etc. Also, we cannot fully depend on Advanced Driver Assistance Systems (ADAS)since they are not 100% accurate. In this paper, we will study different existing models on

which researchers have worked upon with their unique approach, their drawbacks along with and their benefits.

II. LITERATURE SURVEY

[1] presented Maitree Leekha, Mononito Goswami proposed a system for distracted driving detection using foreground extraction and Convolutional Neural Network (CNN). Their model's design is inspired by the Occam's Razor principle which states that by choosing simple design alternatives, they can limit the size of a model and obtain robust performance the dataset was available from the State Farm Distracted Driver Detection dataset (SFD3) and the AUC Distracted Driver dataset (AUCD2) which gave test accuracy of 98.48% and 95.64% respectively. Their model also suggested that incorporating features such as posture through foreground extraction using Grab Cut improves performance. Grab Cut helped in eliminating background noise in the images and extract their foreground image as the input images were frames of the video and by setting the coordination of the bounding box, they were able to extract the driver and able to reduce the complexity of the architecture by eliminating the need of any localization algorithm. These images were then processed by CNN. Their ConvNet model had significantly fewer parameters (0.5M). The model has three convolutional blocks. Each block is composed of a convolutional layer, followed by ReLU nonlinear activation function, a max-pooling layer, and a dropout layer. All convolutional layers have 3*3 filters. The convolutional layer in the first block has 64 kernels, the second has 128, and the third has 256. The convolution blocks are followed by a dense layer and a SoftMax layer, both separated by a dropout layer. The dense layer has 192 neurons and ReLU as the activation function. Their VGG-16 models perform well but give significant delay compared to the ConvNet model surpasses any state-of-the-art

models by detecting real-time distractions without any parallel processing.

[2] written by Qingzhi Bu, Jun Qiu, Hao Wu, Chao Hu proposed the distracted driver detection system based on the histogram of oriented gradient (HOG) and support vector machine (SVM). HOG is used to extract different features of the driver's distracted behavior which is done first by Image preprocessing and then to feature extraction in which first if all gamma is normalized then image gradient is calculated then the histogram is constructed of oriented gradient and cause of large gradient changes block normalized to compress the light and shadow information which is connected to HOG descriptor which generates a feature vector which is input to SVM for training and recognition, and SVM classifier is to classify decentralized behavior. The paper included various algorithms with their accuracy like SVM linear, SVM RBF, LBP+SVM, HOG+SVM linear. The experimental results show that using HOG+SVM and optimizing SVM parameters has a good effect on the recognition rate of this dataset, with an average recognition rate of 93.33%. The paper also mentioned that there was a need to improve the training and recognition speed since LBP+SVM has the highest recognition rate for normal driving and lowest for drinking behavior while HOG + SVM has the highest recognition of normal driving and lowest on talking on the phone, so they concluded that HOG + SVM were more feasible and there was a need of a better algorithm and more samples were needed for training to get an insight to the behavior of the driver.

[3] written by Ketan Ramesh Dhakate and Ratnakar Dash they proposed two models, where they have used ensembling of different CNN models to predict the distraction. With the help of Transfer-learning, they have used five different pre-trained CNN models: VGG-16, VGG-19, InceptionV3, Resnet-50, and Xception which were trained by recent researchers, and each having notable accuracy as

initialization which accelerated the training and increased the performance of the model. Since the prediction of a single classifier is not optimal also all models have a different way to look into the images and predict the class for a particular image. So, they did was they used Stack ensembling in which they trained all models as define earlier removed the final layer, and treated the output of each model as features, they connected them to a meta-classifier or meta-regressor. They have used Categorical Cross Entropy as loss function, SoftMax as activation function, Adam optimizers as an optimization parameter. They have proposed two stack ensemble models in one feature vector of Xecption and InceptionV3 is concatenated while in other, the feature vector of Xception, ResNet-50, VGG-19, and InceptionV3 is used which gives better results than the first one. The different postures resulted in an incorrect prediction also model often gets confused due to slight change in body movement also and if the driver is driving in the night and the car cabin is in the dark, then this model may not work properly.

[4] written by Bhakti Baheti, Suhas Gajre, and Sanjay Talbar proposed a system based on an architecture proposed by Simonyan and Zisserman. They propose a thinned version of VGG-16 with just 15M parameters and still achieving satisfactory classification accuracy compared to the older version of VGG-16 with 140 parameters. The system processes 42 images per second on average. Performance of the system is significantly improved with the addition of dropout, L2 weight regularization, and batch normalization which results in 96.31% accuracy on the test set.

[5] written by Chen Huang, Xiao Chen Wang, Jiannong Cao, Shihui Wang, and Yan Zhang proposed a hybrid CNN framework (HCF) consisting of three modules: a cooperative CNN module where they have pre-trained ResNet50, Inception V3, and Xception which are trained on the ImageNet dataset, and then transfer learning is applied to fine-tune the system then they are combined to extract multiscale behavior features in parallel, a feature concatenation module in which the deep behavior features extracted by the cooperative CNN module are deeply fused into a set of one-dimensional vectors then the fused features contain high-level semantics and lay the foundation for the subsequent image classification, and a feature classification module in which the neurons in the fully connected layer of the HCF capture the key elements of the fused feature vectors and use them as a basis for classification. They have used momentum-based training rate optimization since it is more stable and has a faster convergence speed. They have also fine-tuned the pre-trained model using weight regularization and have from some layers to extract more features out of each model. They have conducted 10 groups of experiments for each algorithm and have randomly selected 100 images from the dataset in each experiment group, the average time of 42 ms is taken to classify one image. The fused features are classified into different distracted driving behaviors HCF can achieve high performance for classification accuracy of 96.74% concerning detection accuracy, average image processing time, and robustness for ten types of typical distracted driving behaviors.

[6] written by Pramila Chawann, Shreyas Sataedekar, Dharmin Shah, Rohit Badugu they have proposed an ensemble of 3 models viz VGG -16, VGG-19and Inception with a log loss value of 0.795. They have used a pre-trained model from ImageNet performed data augmentation. Then fine-tuned the models by removing the last layer and adding a Global Average Pooling layer and fed output to a dense layer of 256 output with relu activation function for VGG models and used SGD optimizer to yield better results from the Inception model. They got validation accuracy of 73% but further proposed to improve the model by using KNN, trying different models, and suggested getting rid of noise from the images.

[7] written by Hesham M. Eraqi, Yehya Abouelnaga, Mohamed H. Saad, and Mohamed N. Moustafa they have proposed a thinned version of their ensemble that achieves 84.64% classification accuracy and can operate in a real-time environment. They have created a new database using Asus ZenFone and DS325 Sony DepthSense camera collected data in video format of 44 participants across 7 countries. They have proposed a solution consisting of a genetically weighted ensemble of CNN. The system first extracts features like face, hands, and skin. Then with the help of skin segmentation and face and hand detection from the pre-trained model, trained the features on a fine-tuned CNN with 30 epoche with 50 images each. The model utilizes a genetically weighted ensemble of convolutional neural networks to achieve a 90% classification accuracy. The number of parameters the deep models adopted and benchmarked is VGG16, AlexNet, ResNet50, and InceptionV3 have 134.3M, 58.3M, 23.6M, and 21.8M parameters, respectively. Since it is still a large number for a real-time system. So, they proposed a version of NasNet Mobile that reduces the number of parameters further to only 4.3M so they used an ensemble of two NasNet Mobile models to produce a satisfactory classification accuracy of 84.64%.

Table 1. Table of Comparison

Author	Year	Approach	Description
Chen Huang, Xiao Chen Wang, Jiannong Cao, Shihui Wang, and Yan Zhang	2020	HCF	They proposed a hybrid CNN framework (HCF) consisting of three modules that are trained on the ImageNet dataset, and then transfer learning is applied to fine-tune the system.
Ketan Ramesh Dhakate and Ratnakar Dash	2020	Model Ensembling	They proposed two models, where they have used an ensemble of different CNN models to predict the distraction.
Hesham M. Eraqi, Yehya Abouelnaga, Mohamed H. Saad, and Mohamed N. Moustafa	2019	Model Ensembling	They proposed a thinned version of their ensemble which can operate in a real-time environment withVGG16, AlexNet, ResNet50, and InceptionV3.
Maitree Leekha, Mononito Goswami, Yifang Yin, and Roger Zimmermann	2019	Convolutional Neural Network	They proposed a system for distracted driving detection using foreground extraction and Convolutional Neural Network (CNN).
Qingzhi Bu, Jun Qiu, Hao Wu, Chao Hu	2019	HOG & SVM	They proposed the distracted driver detection system based on the histogram of oriented gradient (HOG) and support vector machine (SVM).
Bhakti Baheti, Suhas Gajre, and Sanjay Talbar	2018	VGG-16	They proposed a system based on an architecture proposed by

				Simonyan and Zisserman where they have proposed a thinned version of VGG-16 with just 15M parameters.
ramila Chawann, Shreyas Sataedekar, Dharmin Shah and Rohit Badugu	2018	Model Ensembling		They proposed an ensemble of 3 models viz VGG -16, VGG-19, and Inception.

III. CONCLUSION

In the conclusion of this paper, we have studied how distractions can be identified, detected, and classified using the actions performed by the drives. And in the Literature review of all the systems mentioned in the paper, we have studied many different approaches, the problem faced the solution to the problem and suggestions which we can apply while making our model. The main aim of this paper is to identify different methods and many approaches to get acceptable results.

IV. PROPOSED SYSTEM

There is an urgent need for robust real-time distracted driver detection systems which can raise a warning message whenever a driver is distracted. To this end, we propose a CNN-based system to detect distracted driving. It will analyze all the actions performed by the driver and categorize them as safe and not safe driving according to the dataset.

V. REFERENCES

[1]. M. Leekha, M. Goswami, R. R. Shah, Y. Yin, and R. Zimmermann, "Are You Paying Attention? Detecting Distracted Driving in Real-Time," 2019 IEEE Fifth International Conference on Multimedia Big Data (BigMM), Singapore, 2019, pp. 171-180, doi: 10.1109/BigMM.2019.00-28.

[2]. Q. Bu, J. Qiu, H. Wu and C. Hu, "Research on Driver's Distracted Behavior Detection Method Based on Multiclass Classification and SVM," 2019 IEEE International Conference on Robotics and Biomimetics (ROBIO), Dali, China, 2019, pp. 444-448, doi: 10.1109/ROBIO49542.2019.8961551.5

[3]. K. R. Dhakate and R. Dash, "Distracted Driver Detection using Stacking Ensemble," 2020 IEEE International Students' Conference on Electrical,Electronics and Computer Science (SCEECS), Bhopal, India, 2020, pp. 1-5, doi: 10.1109/SCEECS48394.2020.184.

[4]. B. Baheti, S. Gajre and S. Talbar, "Detection of Distracted Driver Using Convolutional Neural Network," 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), Salt Lake City, UT, USA, 2018, pp. 1145-11456, doi: 10.1109/CVPRW.2018.00150.

[5]. C. Huang, X. Wang, J. Cao, S. Wang and Y. Zhang, "HCF: A Hybrid CNN Framework for Behavior Detection of Distracted Drivers," in IEEE Access, vol. 8, pp. 109335-109349, 2020, doi: 10.1109/ACCESS.2020.3001159.

[6]. Chawan, Pramila & Pawar, Prof. (2018). Distracted Driver Detection and Classification. 10.9790/9622-0804036064.

[7]. Eraqi, Hesham & Abouelnaga, Yehya & Saad, Mohamed & Moustafa, Mohamed. (2019). Driver Distraction Identification with an Ensemble of Convolutional Neural Networks. Journal of Advanced Transportation. 2019. 1-12. 10.1155/2019/4125865.

[8]. US Department of Transportation. "Distracted Driver." nhtsa.gov. https://www.nhtsa.gov/risky-

driving/distracted-driving (accessed Mar. 31, 2021).

[9]. Social Determinants of Health. "Road Traffic Injuries." who.int.

[10]. https://www.who.int/news-room/fact-sheets/detail/road-traffic-injuries(accessed Mar. 31, 2021).

[11]. Prachi Savle." Poor enforcement, training: The reasons why there are so many road accidents in India." indiaspend.com.

[12]. https://scroll.in/article/944201/poor-enforcement-training-the-reasons-why-there-are-so-many-road-accidents-in-india (accessed Mar. 31, 2021)

Cite this article as :