e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:05/Issue:01/January-2023 Impact Factor- 6.752 www.irjmets.com

DRIVER'S BEHAVIORAL RECOGNITION USING CNN

Prasanna N[*1], Nandish N[*2], MJ Mohit Kumar[*3], Meka Sai Prasanna Kumar[*4], Monish M[*5]

[*1]Asst. Professor, Department Of Computer ScienceKSSEM Bangalore, India.

[*2,3,4,5]Student, Department Of Computer Science KSSEM Bangalore, India.

DOI : https://www.doi.org/10.56726/IRJMETS32794

ABSTRACT

This paper presents a smartphone-based system for the detection of drowsiness in automotive drivers. The proposed framework uses three-stage drowsiness detection. The first stage uses the percentage of eyelid closure (PERCLOS) obtained through images captured by the front camera with a modified eye state classification method. The system uses near infrared lighting for illuminating the face of the driver during night driving. The second step uses the voiced to the unvoiced ratio obtained from the speech data from the microphone, in the event PERCLOS crosses the threshold. A final verification stage is used as a touch response within a stipulated time to declare the driver as drowsy and subsequently sound an alarm. The device maintains a log file of the periodic events of the metrics along with the corresponding GPS coordinates. The system has three advantages over existing drowsiness detection systems. First, the three-stage verification process makes the system more reliable. The second advantage is its implementation on an Android smart-phone, which is readily available to most drivers or cab owners as compared to other general purpose embedded platforms. This network is a combination of feature extraction and classifier modules. The feature extraction module uses the advantages of the standard convolution layers, depth wise separable convolution layers, average pooling layers, and proposed adaptive connections to extract the feature maps.

Keywords: Attention Mechanism, Convolutional Neural Network, Driver Behavior Recognizer, Driver Warning System, Driver Drowsiness, Safety.

I. INTRODUCTION

Nowadays, road traffic systems have grown much in terms of quantity and complexity. Accordingly, the number of accidents also increased gradually. The statistics of the World Health Organization point out that about 1.35 billion people die and approximately 50 million road traffic collisions occur every year [1]. One of the common causes that leads to an increase in accidents is driver behavior. The statement above also mentioned that if the drivers really focused when driving, it could reduce the accident rate by four times. According to a statistic from the National Highway Transportation and Safety Administration (NHTSA) in the United States (US), about 2,895 people were killed in distracted driving accidents in 2019, accounting for 8.7% of all traffic accident deaths in that year. These reports show that from 2010 to 2019, the number of deaths and accidents caused by distracted driving still maintained a quite high rate between 8% and 10%, 14% The scientists focused on researching the problems, solutions of road traffic accidents, and giving some definitions of distracted driving. The authors in define distracted driving as driver behaviors that interrupt the focus from driving including operating cellphones (talking, texting), eating, drinking, and adjusting the entertainment system (radio, stereo). From another definition in, anything that distracts from paying attention to driving can be considered distracted driving. The main devices for measuring visual distraction are different sensors/cameras mounted on vehicles or directly attached to the driver's body to collect signals, then analyze and process. The proposed network takes advantage of the standard convolution layers, depth wise separable convolution layers, average pooling layers, and adaptive connections with the convolution block attention module (CBAM) to extract the feature maps then learn the outstanding features through the attention mechanism. Finally, the classifier module applies the global average pooling (GAP) layer and SoftMax function to compute ten probabilities of corresponding driver behaviors in the datasets.

II. RELATED WORK

This section will present several techniques applied to driver behavior recognition and their advantages and disadvantages. These methods are considered based on two respects: traditional machine learning and CNN-based methodology.

e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:05/Issue:01/January-2023 Impact Factor- 6.752 www.irjmets.com

A. TRADITIONAL MACHINE LEARNING METHODOLOGY

The first research focused on detecting cellphone usage during driving. It uses the Supervised Descent method, a Histogram of Oriented Gradients (HOG), and an Adaboost classifier to realize the actions of using a cellphone with the accuracy of 93.9%. This study is limited by the cell phone region extraction from facial landmark technique, illumination, and occlusion conditions. Other studies measure the relative distance between four components such as the face, mouth, hands, and cellphone using Hidden CRF [8] and Support Vector Machines (SVM) to classify cellphone use.

B. CNN-BASED METHODOLOGY

In recent years, convolutional neural networks have been widely applied in computer vision fields. Studies on human behavior in general, and driver behavior in particular, have also taken advantage of convolutional neural networks to build monitoring and warning applications. The application areas range from image detection and image segmentation to image classification. The work in uses a Faster R-CNN network as a detector to guess the hand movements on the steering wheel. The results show that this method achieves an accuracy of 92.4% and 91% respectively for cell phone usage and hands on the steering wheel cases. It applies the image segmentation method to localize the steering wheel, gear lever, and dashboard. After that, they propose a network architecture to detect the driver's hand position on previously segmented regions and achieve 74.3% of accuracy. This combined method can solve the problem of illumination changes but is computationally complex. In the image classification task, first proposed a dataset for distracted driving classification called the Southeast dataset with four classes: smoking or eating, talking on the phone, safe driving, and operating the gear lever. This dataset is used in the traditional machine learning methods, and applied several techniques with convolutional neural networks to classify these four classes with an overall accuracy of 99.78%. Later, proposed extended datasets for driving distraction with ten classes (the detailed descriptions are shown in the dataset subsection). Based on these datasets, many studies have used different CNN network architectures for training and evaluation. Also, in and, the authors propose an ensemble training method with five different CNN networks and result in an accuracy of 94.29% and 93.65%, respectively. Other typical classification neural networks such as VGG DenseNet, GoogleNet have also been exploited for driver behaviors classification with the accuracy form 95% to over 99%. For the purpose of reducing network parameters and deploying low-computation devices, [25], [26] proposed convolutional neural network architectures with depth wise separable convolution operation and a residual network to classify ten driver behaviors. These methods achieve very high accuracy (over 95%) and the small number of network parameters (less than 0.5M parameters). The above studies have high accuracy, but only focused on recognizing driver behavior with a limited set of classes (four classes) or only evaluated on individual datasets with a larger number of classes (ten classes). On the other hand, several proposed methods are heavyweight and difficult to apply in real-time systems. As a study on the strength of standard convolutional and depth wise separable convolution layers, and Inception and Residual networks, this work proposes a light-weight driver behavior classification convolutional neural network. It has just 0.43M parameters but the network guarantees high accuracy when compared with many other methods.

III. THE PROPOSED METHOD

1. THE PERCLOS COMPUTATION ALGORITHM

Stage I Compute PERCLOS from the images grabbed using the front camera of the smartphone.

Stage II Compute the voiced to the unvoiced ratio (VUR) in case the PERCLOS reaches a threshold.

Stage III Develop a reaction time test as a final stage verifier.

The system uses a three-stage approach. The first stage computes the PERCLOS using images captured from the front camera of the smartphone. On PERCLOS being higher than a preset threshold, the system asks the driver to say his full name. As a final stage verification, the driver is asked to touch the screen of the smartphone within 10s, once he is found to be fatigued by the earlier two stages, i.e., the PERCLOS and voice-based measures. Fig. 1 shows the overall framework of the proposed system.

PERCLOS is a drowsiness metric, based on eye closure rates. It has been authenticated in [14] as a significant marker of drowsiness. PERCLOS may be defined as the proportion of time in which the eyelids are at least 80% closed over the pupil [14]. Finally, the PERCLOS value is calculated as

e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:05/Issue:01/January-2023 Impact Factor- 6.752 www.irjmets.com

$$P = \frac{Ec}{Eo + Ec} \times 100\%$$

Here, Ec and Eo gives the counts of closed and open eyes respectively for a predefined interval. A higher value of P indicates higher drowsiness level and vice versa [1]. The steps involved in the computation of PERCLOS from an image sequence involve face detection followed by eye detection and eye state classification.

2. 3D CONDITIONAL GENERATIVE ADVERSARIAL NETWORK.

Since generative adversarial network (GAN) was originally proposed by Goodfellow et al. [28], varieties of improved GAN frameworks have been designed and applied in image translation [29], image generation [30], face recognition [31], [32], face aging [33], face inpainting [34] and etc. Motivated by the success of GAN in corresponding issues of face analysis, we design a 3D conditional generative adversarial network (3DcGAN) in this work, which differs from the aforementioned GAN-based network, and the main characteristics of the 3DcGAN

(1) 2D convolution is replaced by 3D convolution in both generator and discriminator. We aim to generate fake image sequences and enable the discriminator to learn short-term spatial-temporal features, instead of the combination of frame level features.

(2) Borrowing the structure of the image-to-image translation, generator is composed of a 3D encoder-decoder network and can translate the original image sequences to fake image sequences. The pixel-wise regression loss in generator ensures the quality of fake samples and improves training stability.

(3) Some auxiliary information, such as eye condition, mouth condition or illumination condition is annotated and added to the generator, aiming at encouraging the discriminator to learn drowsiness-related representation.

3. FEATURE EXTRACTOR MODULE

Most of the popular convolutional neural networks can extract high-level feature maps from raw pixels without any manual processing steps. Therefore, later tasks such as image classification, object detection, and image segmentation will be easily applied and achieve high precision. Meanwhile, traditional machine learning methods rely heavily on image preprocessing and feature extraction, so the received precision is unstable. This study focuses to design the feature extractor based on many novel techniques to obtain the most effective feature maps. The feature extractor includes four convolutional blocks (CBs), four ACs, one CBAM, and two depth wise separable convolution layers. The CBs have two different architectures. The first architecture is built based on two standard convolution layers, a batch normalization (BN) layer, a rectified linear unit (ReLU) activation function, and an average pooling layer (in CB1, CB3). The other architecture uses a standard convolution layer, a depth wise separable convolution layer, a BN layer, a ReLU activation function, and an average pooling layer.

4. CLASSIFIER MODULE

Traditionally, common classification networks have widely used fully connected layers at the end of the classification network. However, this technique significantly increases the network parameters, thus increasing the computational burden on the network and reducing the processing speed when applied on low computing devices. This study proposes a method to replace all fully connected layers in the popular classification networks with only one GAP layer. For this technique, the spatial features are extracted along each channel and the 14 × 14 × 10 feature map from the extractor will quickly reduce the dimensions to 1 × 1 × 10, saving a lot of network parameters. Finally, a SoftMax function is applied to calculate the probability of each object class appearing in the input image. For simplicity, the categorical cross-entropy loss function is used to calculate the difference between the predicted value and the target value during training. It is defined as follows:

$$L_{cls} = -\sum_{i=0}^{9} = p_i^* \log(p\, i),$$

5. VIDEO TESTING SYSTEM

The system consists of input, the trained model, and the output. In which, the input is a set of videos with different resolutions including VGA, HD, FHD. The model is trained on the State Farm dataset and the stored weight file. The output is message text signals on the screen including prediction class, accuracy, and speed in

FPS. This system can flexibly replace the input with a conventional camera, and the output can install audio signals to the speaker to alert the driver. This is the structure of the real-time driver warning system.

IV. CONCLUSION

This paper introduces a driver behavior recognizer based on a light-weight convolutional neural network and attention mechanism. This architecture exploits the advantages of standard convolution, depth wise separable convolution operation, and proposed adaptive connections to extract feature maps. Then, the network uses the CBAM attention mechanism to make the network focus on learning the most salient features. Finally, the classifier is applied to recognize ten driver behaviors. This work applied several techniques for reducing the number of network parameters and increasing the accuracy. The proposed network used all three benchmarks to train, evaluate, and report the results in the accuracy metric. On the other hand, it was also tested on different resolution videos with good processing speeds. In the future, this approach continues to develop based on a two-stage driver behavior warning system. The proposed network will be integrated into this system as the second stage after the driver body detection stage. By extracting the driver body positions first and then classifying it is possible to greatly increase the behaviors classification accuracy, especially with the real-time applications.

V. REFERENCES

[1] Road Traffic Injuries. Accessed: Jan. 25, 2022. [Online]. Available: https://www.who.int/news-room/fact-sheets/detail/road-traffic-injuries

[2] Distracted Driving. Accessed: Jan. 25, 2022. [Online]. Available: https://www.nhtsa.gov/risky-driving/distracted-driving

[3] Distracted Driving. Accessed: Jan. 25, 2022. [Online]. Available: https://www.cdc.gov/transportationsafety/distracted-driving

[4] A. Eriksson and N. A. Stanton, ‘‘Takeover time in highly automated vehicles: Noncritical transitions to and from manual control,’’ Hum. Factors, J. Hum. Factors Ergonom. Soc., vol. 59, no. 4, pp. 689–705, 2017.

[5] H. M. Eraqi, M. N. Moustafa, and J. Honer, ‘‘End-to-end deep learning for steering autonomous vehicles considering temporal dependencies,’’ CoRR, vol. abs/1710.03804, Dec. 2017, doi: 10.48550/arXiv.1710.03804.

[6] H. M. Eraqi, J. Honer, and S. Zuther, ‘‘Static free space detection with laser scanner using occupancy grid maps,’’ CoRR, vol. abs/1801.00600, Jan. 2018, doi: 10.48550/arXiv.1801.00600.

[7] K. Seshadri, F. Juefei-Xu, D. K. Pal, M. Savvides, and C. P. Thor, ‘‘Driver cell phone usage detection on strategic highway research program (SHRP2) face view videos,’’ in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW), Jun. 2015, pp. 35–43.

[8] T. Hong and H. Qin, “Drivers drowsiness detection in embedded system,” in Proc. IEEE Int. Conf. Veh. Electron. Saf(ICVES), Dec. 2007, pp. 1–5.

[9] L. Lang and H. Qi, “The study of driver fatigue monitor algorithm combined PERCLOS and AECS,” in Proc. Int. Conf. Comput. Sci. Softw. Eng., vol. 1, Dec. 2008, pp. 349–352.

[10] W. Qing, S. BingXi, X. Bin, and Z. Junjie, “A PERCLOS-based driver fatigue recognition application for smart vehicle space,” in Proc. 3rd Int. Symp. Inf. Process. (ISIP), Oct. 2010, pp. 437–441.

[11] B.-G. Lee and W.-Y. Chung, “Driver alertness monitoring using fusion of facial features and bio-signals,” IEEE Sensors J., vol. 12, no. 7, pp. 2416–2422, Jul. 2012.

[12] Z. Wan, J. He, and A. Voisine, “An attention level monitoring and alarming system for the driver fatigue in the pervasive environment,” in Brain and Health Informatics. Springer, 2013, pp. 287–296.

[13] C.-W. You et al., “CarSafe app: Alerting drowsy and distracted drivers using dual cameras on smartphones,” in Proc. 11th Annu. Int. Conf. Mobile Syst., Appl., Services, Taipei, Taiwan, 2013, pp. 13–26.

[14] D. F. Dinges, M. M. Mallis, G. Maislin, and J. W. Powell, “Evaluation of techniques for ocular measurement as an index of fatigue and the basis for alertness management,” Nat. Highway Traffic Saf. Admin., Washington, DC, USA, Tech. Rep. DOT HS 808 762, 1998.

[15] K. Zuiderveld, "Contrast limited adaptive histogram equalization," in Graphics Gems IV. New York, NY, USA: Academic, 1994, pp. 474–485.

[16] M. A. Hearst, S. T. Dumais, E. Osuna, J. Platt, and B. Scholkopf, "Support vector machines," IEEE Intelligent Systems and their Applications, vol. 13, no. 4, pp. 18–28, July 1998.

[17] F. Schroff, D. Kalenichenko, and J. Philbin, "Facenet: A unified embedding for face recognition and clustering," in 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), June 2015, pp. 815–823.

[18] Z. Liu, P. Luo, X. Wang, and X. Tang, "Deep learning face attributes in the wild," in 2015 IEEE International Conference on Computer Vision (ICCV), Dec 2015, pp. 3730–3738.

[19] Y. Taigman, M. Yang, M. Ranzato, and L. Wolf, "Deepface: Closing the gap to human-level performance in face verification," in 2014 IEEE Conference on Computer Vision and Pattern Recognition, June 2014, pp. 1701–1708.

[20] S. Ren, K. He, R. Girshick, and J. Sun, "Faster r-cnn: Towards real-time object detection with region proposal networks," in Proceedings of the 28th International Conference on Neural Information Processing Systems - Volume 1, ser. NIPS'15. Cambridge, MA, USA: MIT Press, 2015, pp. 91–99. [Online].Available:http://dl.acm.org/citation.cfm?id=2969239.2969250