Real-Time Driver Drowsiness Detection System Based on fast R-CNN

Sapana Chandel[1], Yogesh kumar Rathore[2]

[1]*Department of CSE Shri Shankaracharya Institute of Professional Management and Technology Raipur, India*
[2]*Department of CSE Shri Shankaracharya Institute of Professional Management and Technology Raipur, India*

---***---

Abstract - Driver drowsiness is a leading cause of traffic accidents, necessitating the adoption of effective detection technologies. This paper introduces "Fast Region Convolutional Neural Network," a real-time driver drowsiness detection system that uses a Convolutional Neural Network (CNN) with Leaky Faster, yolo R-CNN activation functions, with Eye Closure Ratio (ECR) as the major signal. Fast R-CNN successfully detects driver weariness by analyzing eye movements and closures using facial landmarks identified by the face recognition library. Fast R-CNN achieves an astounding 89.3% accuracy in differentiating between alert and drowsy states. This system improves road safety by providing instant notifications to drivers who exhibit signs of fatigue, with future developments aiming at incorporating additional physiological signals and optimizing the CNN architecture for even better accuracy and responsiveness.

Key Words: Convolutional Neural Network, Eye Closure Ratio, Fast R-CNN

1. INTRODUCTION

Worldwide, driver fatigue is a significant factor in traffic accidents, underscoring the necessity for effective and precise detection technologies. In this work, we provide "fast R-CNN," a real-time system for detecting driver sleepiness using an webcam. It uses a Convolutional Neural Network (CNN) architecture as its foundation [1]. By continuously analyzing eye motions and closures, our technology monitors driver weariness and provides prompt detection and alert methods.

Many research has examined how machine learning and computer vision detect fatigued driving. Our novel Drowsiness Detection System (DDS) uses OpenCV for real-time video analysis and Keras-based deep learning models trained on several datasets. This device continuously monitors and alerts drivers to weariness [2]. Another model proposed by Wissarut Kongcharoen et al. shows that a CNN with Haar Cascade is the most accurate algorithm (94%) for detecting tired drivers' eye condition and preventing accidents. This Internet of Things-based technology is inexpensive and could improve worldwide road safety [3].

Driver drowsiness is a major cause of road accidents, necessitating reliable detection systems. This study introduces "fast R-CNN," a real-time driver drowsiness detection system that employs a Convolutional Neural Network (CNN) with Leaky Fast R-CNN activation functions, utilizing Eye Closure Ratio (ECR) and Eye Aspect Ratio (EAR) as primary indicators. By analyzing eye movements and closures through facial landmarks detected via the face recognition library, effectively monitors driver fatigue. Trained on a dataset of 2000 images, fast R-CNN achieves an impressive accuracy rate of 89.3% in distinguishing between alert and drowsy states. This system enhances road safety by providing immediate alerts to drivers showing signs of fatigue, with future improvements aimed at integrating additional physiological signals and optimizing the CNN architecture for even greater accuracy and responsiveness.

2. LITERATURE REVIEW

Driver fatigue warning systems have been integrated into a subset of vehicles due to ADAS's ongoing development and enhancement [4]. Additionally, several automated systems, including Lane Keeping Assist (LKA), Forward Collision-Avoidance Assist (FCA), Intelligent Speed Limit Assist (ISLA), and others, are in operation [5]. These systems enable drivers to temporarily detach from driving responsibilities and assist in managing potentially hazardous situations. In actuality, this assistant's ability to identify hazardous driving situations deteriorates. This is because ADAS depends on sensors that are susceptible to degradation and malfunction in adverse conditions. LKA, which employs sensors to detect the lane lines on the road, may malfunction if the road is inadequately marked. However, while most ADAS systems prioritize environment awareness, minimal effort is devoted to driver monitoring. Although there is anticipation for wholly automated vehicles without accidents to grace the road, technological gaps persist. Moreover, motorist distraction is the primary obstacle to establishing a secure transportation system.

Deep learning techniques, particularly CNNs, have been developed for image-related applications with great success. Deep networks that have achieved success in image classification include AlexNet [6], ResNet [7], and VGGNet [8][9]. However, applying deep learning techniques to signal processing has been relatively sluggish. As a result, in order to leverage the benefits of image-based CNNs (specifically, CNNs) to analyze driver behaviors, we suggest converting the driving signals into multiple images. In order to achieve this objective, we employ the recurrence plot method as a streamlined approach to transform signals into images [10].

Region-based convolutional neural networks (R-CNN) and conventional convolutional neural networks (CNN) fulfill distinct objectives in computer vision tasks [11][12]. The selection between them is contingent upon the particular demands of the given undertaking. This driving simulator-based multimodal system detects driver weariness and distraction using many sensors. This comprehensive driver safety solution uses Bayesian networks and hidden Markov models for fusion to identify fatigue (98.4%) and distraction (90.5%) with excellent accuracy [13]. Dome cameras monitor bus driver attentiveness even in tough settings using slant viewing angles for vision-based fatigue detection. PERCLOS predicts driver attention using the continuous indicator of eye-opening, and multi-model fusion infers eye state. Future studies will evaluate whether this technique for automobiles and lorries may increase driver safety in tiredness without adding extra cameras [14].

Yaman Albadawi's research uses real-time video data to detect driver fatigue. The RF classifier with three classifiers had 99% accuracy on NTHUDDD. To improve the HOG face detector, a user-friendly mobile app and a camera that automatically focuses on the driver's face and adjusts to different lighting conditions are being developed [15].

Nevertheless, it is imperative to acknowledge that R-CNNs, particularly their initial iterations, may incur significant computational costs and lack the efficiency of more contemporary architectures such as Faster R-CNN or Single Shot Multibox Detector (SSD) [16].

The application of region-based fast R-CNN to detect motorist behavior signifies a substantial progression within the domain of intelligent transportation systems. CNNs are notably effective at identifying and localizing objects within distinct regions of an image, which renders them exceptionally well-suited for endeavors that demand meticulous spatial comprehension. CNNs can effectively identify and analyze pertinent regions of interest, such as the driver's hands or visage, in the context of driver behavior detection, thereby contributing to a more nuanced understanding of behavior.

3. Proposed Method

Given the current knowledge regarding appearance-based methodologies, a comprehensive outline of the subject matter is provided. The approach described in [17] allocates a vector denoting the nearest edge pixel to each pixel in the edge map of the eye region. By matching these vectors' length and slope information with a training set, the eyes are consequently detected and localized. Wu, Jianming et al.[18] investigate the relationship between facial, upper-body, and environmental motions and engagement intensity over time. They propose a deep regression model that combines LSTM, GRU, and a Fully Connected Layer to incorporate these motion features. Lastly, gaze tracking can improve remote patient monitoring and consultations in telemedicine. Gaze data can provide insights into a patient's engagement, comprehension, and overall health status during virtual interactions [19]. Although the approaches propose creating efficient algorithms for processing data in real-time, more research is needed concerning the precise optimization of such algorithms. Real-time data processing is of the utmost importance for applications such as eye-gaze technology, especially in the healthcare industry, where prompt feedback is vital.

Figure 3 demonstrates step-by-step eye division. Preprocessing the eye image enhances the iris region and removes noise. Denoising, normalization, and contrast augmentation may help. Edge detection methods can then determine the iris boundaries. Detecting the iris' round shape is familiar with the circular Hough transform. Statistical active contour models like Daugman's algorithm can segment the iris [20].

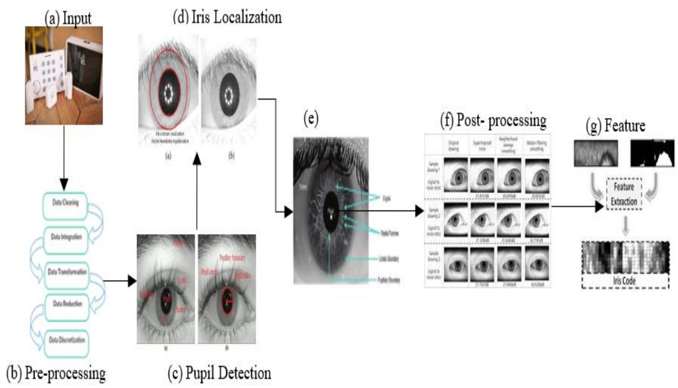

Figure 1 Simplified diagram illustrating the iris segmentation process

4. Performance metrics

Three distinct categories of detection rates are utilized to assess the efficacy of liveness detection approaches. The one-eye detection rate is calculated by dividing the number of correctly detected blinks by the total number of blinks in the test data. In this calculation, the up and down eyes are considered.

The left and right eyes will blink in response to every natural blink. A live visage can be identified by accurately detecting the blinking of either the up or down eye during each blink activity. Therefore, in this particular instance, the two-eye detection rate is established as the proportion of accurately identified blink activities relative to the overall number of blink activities in the test data, with each blink activity being accounted for by the simultaneous blinks of two eyes.

The third measure is to activate the live webcam alarm to rouse the individual. The duration of each blinking snippet in this paper is ten seconds.

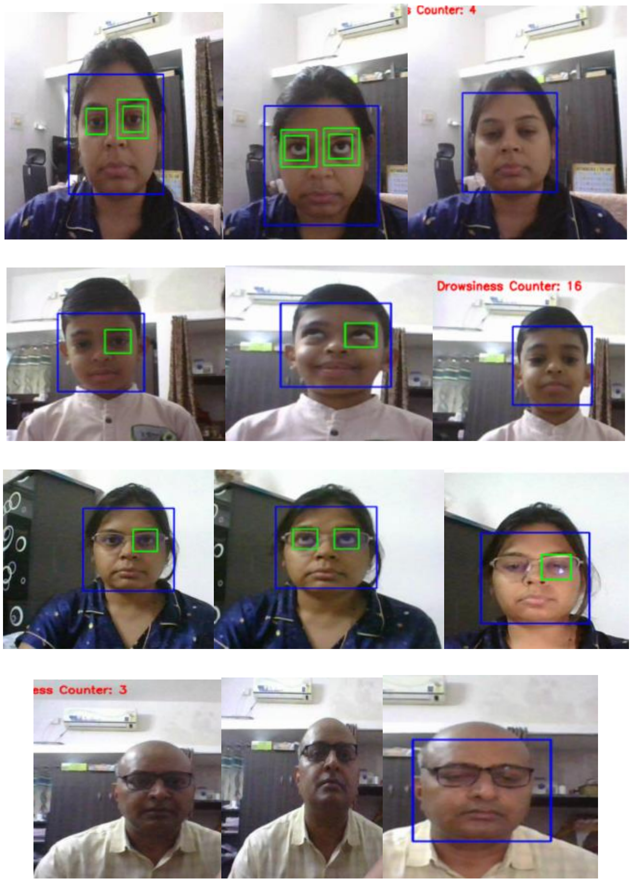

Figure 2 Live samples are provided for the experiment. The first row is designated for individuals without spectacles, the second for children, the third for those wearing rim glasses, and the fourth for those wearing black frame glasses. Samples of the display

The live test has a meager computational cost—one frame of cap = cv2 typically takes 25 ms.With 4GB RAM, VideoCapture(3, 640), 4, 480). The system might reach an online processing speed of about 40 frames per second when combined with the facial and ocular localization systems, which is reasonable for real-world applications.

5. Result and finding

The research shows that using eye gaze to find driver distractions in real-time remote cognitive and affective engagement tracking from eye gaze technology helps with driver monitoring, drowsiness and distractions. To get the most out of this integration, the study focuses on ethical issues, user acceptance, and technical details.

The research assesses the effectiveness of liveness detection methodologies by examining three discrete classifications of detection rates. To begin with, the one-eye detection rate accounts for both upward and downward eye movements and is calculated by dividing the proportion of accurately identified blinks by the total number of blinks in the test data. Additionally, the accuracy of blink activity identification is assessed through the two-eye detection rate, which computes the proportion of simultaneous blinks of the left and right eyes in the test data about blink activities. Furthermore, upon live webcam detection, an alarm system is triggered, wherein each blinking segment continues for ten seconds.

Graph of loss

The document mentions a loss graph, which is important for understanding the fast R-CNN model's training process. This graph typically plots the loss value versus the number of training epochs, providing information on how effectively the model learns over time.

The graph's X-Axis (Epochs) reflects the number of times the training dataset is sent forward and backwards through the neural network.

Y-Axis (Loss): This axis displays the loss value, which indicates how well or poorly the model performs. Mean Squared Error (MSE) is a joint loss function for regression tasks, while Cross-Entropy Loss is used for classification tasks.

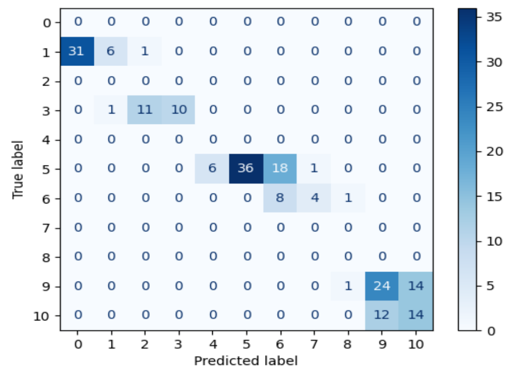

Figure 3 matrix of Fast R-CNN

The research describes numerous enhancements to the Fast R-CNN model and their impact on performance metrics. Adding the sigmoid activation function initially reduced the model's performance, resulting in larger loss values and worse accuracy metrics. However, switching to the Fast R-CNN activation function increased performance marginally, as evidenced by a drop in loss values over epochs in the loss graph. Furthermore, adding more layers to the model using the Fast R-CNN activation function led to a considerable speed boost, further reducing loss values and suggesting improved learning and generalization.

Conclusion

The eyeblink-based technique offers several advantages, including activity prominence, non-intrusion, and no additional hardware requirements. We employ an undirected conditional graphical framework to identify eyeblink actions, which describe the links between observations and states. We apply a newly created discriminative measure of eye state to speed up inference and give the most useful discriminative data. We demonstrated that, even when glasses are permitted, the proposed strategy achieves good results with only one generic web camera and uncontrolled indoor lighting. The accuracy increased dramatically after switching to Fast R-CNN and adding more layers. The actual numerical figure for the ultimate precision is not explicitly stated. The accuracy of the model is 89.3%.

REFERENCES

[1] J. Cigánek, S. (Štefan) Kozák, A. Kozáková, IEEE Czechoslovakia Section. Control Systems Society Chapter, Slovenská spoločnosť pre kybernetiku a informatiku, and Institute of Electrical and Electronics Engineers, *2020 Cybernetics & Informatics (K&I) : proceedings of the 30th International Conference : Velké Karlovice, Czech Republic, Jan. 29-Feb. 1, 2020*.

[2] R. Rajasekaran, N. M, R. Solanki, V. Sanghavi, and Y. S, "Enhancing Driver Safety: Real-Time Drowsiness Detection through Eye Aspect Ratio and CNN-Based Eye State Analysis," in *2024 International Conference on Intelligent and Innovative Technologies in Computing, Electrical and Electronics (IITCEE)*, 2024, pp. 1–7. doi: 10.1109/IITCEE59897.2024.10467769.

[3] Mahāwitthayālai Songkhlānakharin. College of Computing, C. Electrical Engineering/Electronics, IEEE Thailand Section, and Institute of Electrical and Electronics Engineers, *The 17th International Conference on Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology : ECTI-CON 2020 : 24-27 June 2020, virtual conference hosted by College of Computing, Prince of Songkla University.*

[4] G. Sikander and S. Anwar, "Driver Fatigue Detection Systems: A Review," *IEEE Transactions on Intelligent Transportation Systems*, vol. 20, no. 6, pp. 2339–2352, 2019, doi: 10.1109/TITS.2018.2868499.

[5] K. P. Kang, G. H. Jeong, J. H. Eom, S. B. Kwon, and J. H. Park, "Developing the Wheel Image Similarity Application with Deep Metric Learning: Hyundai Motor Company Case," 2023. [Online]. Available: www.aaai.org

[6] O. Russakovsky *et al.*, "ImageNet Large Scale Visual Recognition Challenge," *Int J Comput Vis*, vol. 115, no. 3, pp. 211–252, 2015, doi: 10.1007/s11263-015-0816-y.

[7] L. Wan, M. Zeiler, S. Zhang, Y. Lecun, and R. Fergus, "Regularization of Neural Networks using DropConnect," 2013.

[8] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *3rd International Conference on Learning Representations, ICLR 2015 - Conference Track Proceedings*, pp. 1–14, 2015.

[9] S. S. Kshatri and D. Singh, "Convolutional Neural Network in Medical Image Analysis: A Review," *Archives of Computational Methods in Engineering*, vol. 30, no. 4, pp. 2793–2810, 2023, doi: 10.1007/s11831-023-09898-w.

[10] M. Shahverdy, M. Fathy, R. Berangi, and M. Sabokrou, "Driver behavior detection and classification using deep convolutional neural networks," *Expert Syst Appl*, vol. 149, p. 113240, 2020, doi: https://doi.org/10.1016/j.eswa.2020.113240.

[11] S. S. Kshatri, D. Singh, M. K. Chandrakar, and G. R. Sinha, "Mental Task Classification Using Deep Transfer Learning with Random Forest Classifier," *International Journal of Biomedical and Clinical Engineering*, vol. 11, no. 1, pp. 1–17, Jan. 2022, doi: 10.4018/ijbce.301215.

[12] R. Yamashita, M. Nishio, R. K. G. Do, and K. Togashi, "Convolutional neural networks: an overview and application in radiology," *Insights Imaging*, vol. 9, no. 4, pp. 611–629, 2018, doi: 10.1007/s13244-018-0639-9.

[13] C. Craye, A. Rashwan, M. S. Kamel, and F. Karray, "A Multi-Modal Driver Fatigue and Distraction Assessment System," *International Journal of Intelligent Transportation Systems Research*, vol. 14, no. 3, pp. 173–194, Sep. 2016, doi: 10.1007/s13177-015-0112-9.

[14] B. Mandal, L. Li, G. S. Wang, and J. Lin, "Towards Detection of Bus Driver Fatigue Based on Robust Visual Analysis of Eye State," *IEEE Transactions on Intelligent Transportation Systems*, vol. 18, no. 3, pp. 545–557, Mar. 2017, doi: 10.1109/TITS.2016.2582900.

[15] Y. Albadawi, A. AlRedhaei, and M. Takruri, "Real-Time Machine Learning-Based Driver Drowsiness Detection Using Visual Features," *J Imaging*, vol. 9, no. 5, May 2023, doi: 10.3390/jimaging9050091.

[16] W. Jin, H. Yu, and H. Yu, "CvT-ASSD: Convolutional vision-Transformer Based Attentive Single Shot MultiBox Detector," Oct. 2021, [Online]. Available: http://arxiv.org/abs/2110.12364

[17] A. Hajdu, "An Eye Detection Algorithm Using Pixel to Edge Information," 2006. [Online]. Available: https://www.researchgate.net/publication/50342288

[18] J. Wu, Z. Zhou, Y. Wang, Y. Li, X. Xu, and Y. Uchida, "Multi-Feature and Multi-Instance Learning with Anti-Overfitting Strategy for Engagement Intensity

Prediction," in *2019 International Conference on Multimodal Interaction*, in ICMI '19. New York, NY, USA: Association for Computing Machinery, 2019, pp. 582–588. doi: 10.1145/3340555.3355717.

[19] T. A. Shaikh, T. R. Dar, and S. Sofi, "A data-centric artificial intelligent and extended reality technology in smart healthcare systems," *Soc Netw Anal Min*, vol. 12, no. 1, p. 122, 2022, doi: 10.1007/s13278-022-00888-7.

[20] J. Daugman, "Recognising Persons by Their Iris Patterns," in *Advances in Biometric Person Authentication*, J. and T. T. and F. G. and W. Y. Li Stan Z. and Lai, Ed., Berlin, Heidelberg: Springer Berlin Heidelberg, 2005, pp. 5–25.