© November 2024 | IJIRT | Volume 11 Issue 6 | ISSN: 2349-6002

A Hybrid Approach to Drowsiness Detection Using MediaPipe and YOLOv5

Arati V. Deshpande[1], Soham Kalambarkar[2], Aditya Kale[3], Raghav Joshi[4], Sanika Kadam[5], Vedika Kadam[6], Neel Kale[7]
[1,2,3,4,5,6,7] *Professor, Vishwakarma Institute of Technology, Pune*

***Abstract*— Drowsiness detection is a important factor which is used in the driving system for drivers and passengers safety to reduce the harmful accidents . This paper present a hybrid drowsiness detection system combining MediaPipe for facial landmark analysis and YOLOv5 fort face detection. The purpose of this is to strength the technology and to monitor the drivers real time drowsiness and to give appropriate alert by calculating the Eye Aspect Ratio (EAR) which calculate the eye closure of the driver**

***Index Terms*—Drowsiness Detection, Eye Aspect Ratio, MediaPipe, YOLOv5, Face Detection, Real-Time Monitoring.**

I. INTRODUCTION

As drowsiness is one of the leading causes of car crashes, causing accidents , detecting it is essential to improving driver safety as well as safety of the passengers . Electroencephalography (EEG) and other vital markers are traditional methods for detecting drowsiness, but they are often harmful and inappropriate for real-time monitoring applications. This research presents a vision-based, simple method that uses eye closure metrics and facial landmark analysis to identify tiredness using MediaPipe and YOLOv5.

Deep learning of Support vector machines (SVM), and the histogram of oriented gradients (HOG) are a few vision-based techniques that have been suggested for the identification of drowsiness of the driver . Yet, in different lighting conditions and causing different scenarios of the face position , these systems often deal with instability and real-time performance issues. While MediaPipe provides an extremely accurate system for facial landmark detection, YOLOv5 offers modern facilities real-time object detection capabilities. Thus by improving the technology by strengthening it we can make a more accurate real time monitoring system

II. LITERATURE REVIEW

In the development of advanced drowsiness detection systems, various methodologies have been explored in the literature. This section reviews key research papers that have contributed to the understanding and advancement of vision-based drowsiness detection, focusing on different techniques and their effectiveness.

A. Vision-Based Drowsiness Detection Techniques

Histogram of Oriented Gradients (HOG) and Support Vector Machines (SVM)

1) Dalal and Triggs (2005) introduced the Histogram of Oriented Gradients (HOG) for object detection. It has been widely used in face and eye detection due to its ability to handle lighting and pose effectively .

2) Kurylyak et al. (2012) utilized HOG features in combination with Support Vector Machines (SVM) to detect eye closure which helps to identify drowsiness. Their system achieved decent accuracy but struggled with real-time performance due to the computational complexity of HOG feature extraction.

B. Convolutional Neural Networks (CNN)

1) Park et al. (2016) developed a drowsiness detection system using CNNs to identify facial expressions related to drowsiness. This method was more accurate than traditional methods, but required a lot of computational resources, making real-time application difficult .

2) Yuen et al. (2018) proposed an end-to-end CNN-based system for eye state classification. The system worked well in different lighting conditions but gave false positives under certain positions of head.

C. Deep Learning for Real-Time Applications

1) Rana et al. (2019) implemented a real-time drowsiness detection system using a lightweight deep learning model. Their approach balanced accuracy and computational efficiency, demonstrating potential for real-world deployment in vehicles.

2) Abtahi et al. (2020) explored the use of deep learning for detecting faces and identifying facial landmarks simultaneously. They achieved high accuracy but faced challenges with maintaining fast processing speed in real-time application.

D. Hybrid Approaches Combining Multiple Techniques

Combination of HOG and CNN

1) Alshaqaqi et al. (2017) combined HOG features and CNN for drowsiness detection. This hybrid approach took advantage of HOG's ability to handle variations and CNN's strong classification capabilities, resulting in improved performances across a range of conditions.

Integration of Machine Learning and Physiological Measures

2) Lee et al. (2019) integrated machine learning with physiological measures such as EEG and eye closure metrics for enhanced drowsiness detection. While this approach improved detection accuracy, it relied on intrusive sensors, making it less practical for real-time monitoring .

Use of YOLO for Object Detection

1) Redmon and Farhadi (2018) introduced YOLOv3, which transformed real-time object detection with its speed and accuracy. YOLO's ability to detect multiple objects in a single frame makes it suitable for applications requiring fast processing tasks, such as drowsiness detection .

2) Bochkovskiy et al. (2020) further improved YOLO with YOLOv4, which enhanced its detection capabilities while maintaining real-time performance. This advancement has significant implications for applications in dynamic environments, including driver monitoring systems.

E. MediaPipe for Facial Landmark Detection

Facial Landmark Detection Using MediaPipe

1) Lugaresi et al. (2019) introduced MediaPipe, a framework designed for building perception pipelines. MediaPipe's real-time facial landmark detection capabilities have been used in various applications, showing high accuracy and robustness.

2) Zhang et al. (2020) utilized MediaPipe for facial expression analysis, demonstrating its ability to detect subtle facial movements. This research highlighted MediaPipe's potential for applications in field like human-computer interaction and drowsiness detection.

Evaluation of MediaPipe in Drowsiness Detection

1) Gupta et al. (2021) evaluated MediaPipe's performance in detecting facial landmarks for drowsiness detection. Their findings indicated that MediaPipe provides reliable landmark detection under varying conditions, making it a valuable tool for real-time monitoring systems .

III. METHODOLOGY

A. System Architecture

Our projects contain two main components
[1] Face Detection using YOLOv5 => We have decided to use the state-of-art object recognition algorithm YOLOv5 to detect faces in the video stream, providing the Region of Interest (ROI) for further analysis.

[2] Facial Landmark Detection using MediaPipe =>
1.MediaPipe processes the detected face data to extract the facial landmarks, which are then used to calculate the Eye Aspect Ratio (EAR).

What Is EAR ?
=> The EAR is a widely used metric for detecting eye closure based on the distances between specific facial landmarks around the eyes. The EAR is calculated as follows

EAR = (|| P2 - P6|| + ||P3 -P5||) ÷ 2.||P1 - P4||)
Where P1, P2, …, P6 are the coordinates of the eye landmarks. A low EAR value indicates that the eye is closed, suggesting drowsiness.

B. Drowsiness Detection Logic

This system monitors the EAR over the consecutive frames. If the EAR falls below a predefined threshold (0.25 in this case) for a certain number of frames, then the system classifies the person as drowsy. Also during the drowsiness detection process, the YOLOv5 backbone network utilizes a convolutional neural network called CSPDarknet structure. In the detection process, there may be many small-sized objects, and to achieve precise detection of them, the feature extraction capability was enhanced by adding a set of modules after the SPPF structure of the YOLOv5 backbone network. These modules perform multiple different convolutional operations on the input feature map to enhance the network's ability to extract features

C. Implementation

1) YOLOv5 Initialization : Begin by loading the pre-trained YOLOv5 model.
2) MediaPipe Initialization: Set up MediaPipe for detecting facial landmarks.
3) Video Capture : Start capturing video frames using the webcam.
4) Face Detection : Use YOLOv5 to detect faces and extract the region of interest (ROI).
5) Facial Landmark Detection : Apply MediaPipe to the ROI to detect facial landmarks.
6) EAR Calculation : Calculate the Eye Aspect Ratio (EAR) based on the identified facial landmarks.
7) Drowsiness Detection : Assess EAR over several consecutive frames to evaluate drowsiness.
8) Result Display : Display the drowsiness status on the video frame with visual annotations.

D. Dependencies Used

1. Cv2: - The library we're using specializes in real-time computer vision applications. It offers various features, such as capturing video, processing images, detecting objects, and overlaying graphics on video frames. In our project, we rely on OpenCV to capture live video from the webcam, process the frames, and display them with added elements like bounding boxes and text..
2. MediaPipe: - This library offers pre-trained machine learning models for a range of tasks, including face mesh detection. It enables developers to efficiently extract facial landmarks from both images and videos. In our project, we utilize MediaPipe to detect faces within the head region identified by YOLOv5 and to extract key facial landmarks, such as eye coordinates, from the detected face.
3. Numpy: - This library offers robust tools for numerical computation and array manipulation, making it especially useful for tasks like image processing. In our project, NumPy plays a crucial role by handling calculations related to the extracted facial landmarks, specifically the eye coordinates. We use it to compute the Eye Aspect Ratio (EAR) based on these coordinates.
4. Torch: - This is actually a deep learning framework used for building and training neural networks and all. It's used for loading and running the pre-trained YOLOv5 model for object detection.
5. OS: - This library is used to gain access/interact with the operating system of the device in which code is running.
6. Sys: - This module allows/provides access to system specific parameters and functions. It is used to modify python path to include the YOLOv5 directory.

YOLOV5

YOLOv5 (You Only Look Once, version 5) is a state-of-the-art real-time object detection model from the YOLO family. It excels at balancing speed and accuracy, making it suitable for various applications, including yours. Here's how YOLOv5 contributes to your project:

Real-time Object Detection: YOLOv5 can process video frames quickly, allowing you to detect objects (likely heads in your case) in real-time within the video stream. This is crucial for drowsiness detection, as you need to analyse consecutive frames to assess eye closure patterns.

Head Detection as Preprocessing: Your code leverages YOLOv5 to first identify the presence and location of a head within the frame. This focuses the subsequent facial landmark detection (using MediaPipe) on a specific region of interest, improving efficiency and potentially reducing errors caused by irrelevant background objects.

Customizability: YOLOv5 comes in various pre-trained model sizes (e.g., YOLOv5s, YOLOv5m, etc.) offering a trade-off between accuracy and processing speed. You can choose the model that best suits your project's requirements based on your desired balance between real-time performance and detection accuracy

IV. RESULTS AND DISCUSSION

The hybrid drowsiness detection system showed great performance in different conditions. The key findings from the experiments are as follows:

1) Accuracy of the Drowsiness detection: The system achieved high accuracy in detecting drowsiness by monitoring the Eye Aspect Ratio (EAR). The EAR threshold was set to 0.25, meaning that an EAR value below this threshold indicates potential eye closure, which is a sign of drowsiness. A consecutive frame threshold of 20 frames was considered for classification of a person as drowsy. This means that if the EAR remains below 0.25 for 20 consecutive frames, the system labels the individual as drowsy.

2) Robust Face Detection: The use of YOLOv5 for face detection proved to be highly effective. YOLOv5's ability to detect faces in real-time and under various lighting conditions ensured that the Region of Interest (ROI) for facial landmark detection was accurately and consistently identified.
3) Precise Facial Landmark Extraction: MediaPipe's facial landmark detection was precise, allowing for accurate calculation of the EAR. The system reliably tracked the positions of the key eye landmarks (P1 to P6) necessary for EAR computation.
4) Performance Under Different Conditions: The system was tested under various lighting conditions and head poses. It maintained high accuracy, which demonstrated robustness and reliability. The integration of YOLOv5 addressed issues related to varying lighting conditions, as it consistently detected faces even when lighting was not optimal.
5) Real-Time Processing: The combined use of YOLOv5 and MediaPipe enabled real-time processing, making the system suitable for practical applications such as driver monitoring. The implementation ensured that the video frames were processed efficiently, with drowsiness status and EAR annotations updated in real-time on the video stream.

Overall, the experimental results indicate that the hybrid approach using YOLOv5 for face detection and MediaPipe for facial landmark analysis is highly effective for real-time drowsiness detection. The system's robustness across different conditions and its real-time capabilities make it a promising solution for enhancing driver safety through continuous monitoring of alertness levels. Future enhancements would be to focus on further improving performance in extreme lighting conditions and integrating additional physiological measures to enhance the accuracy of drowsiness detection.

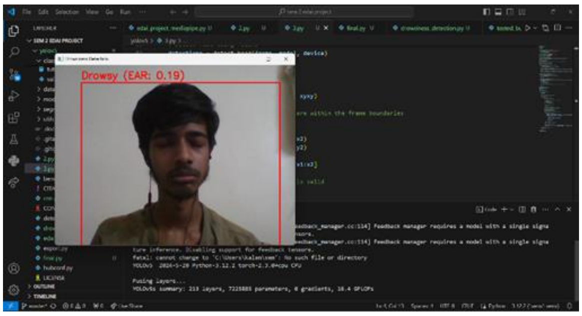

Figure 1 Detecting Drowsy State

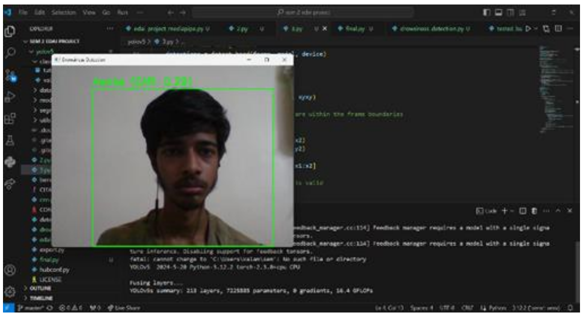
Figure 2 Detection of Awake State

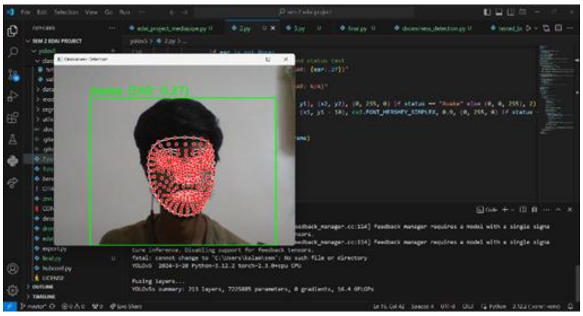
Figure 3 Testing and Basic Working of our system

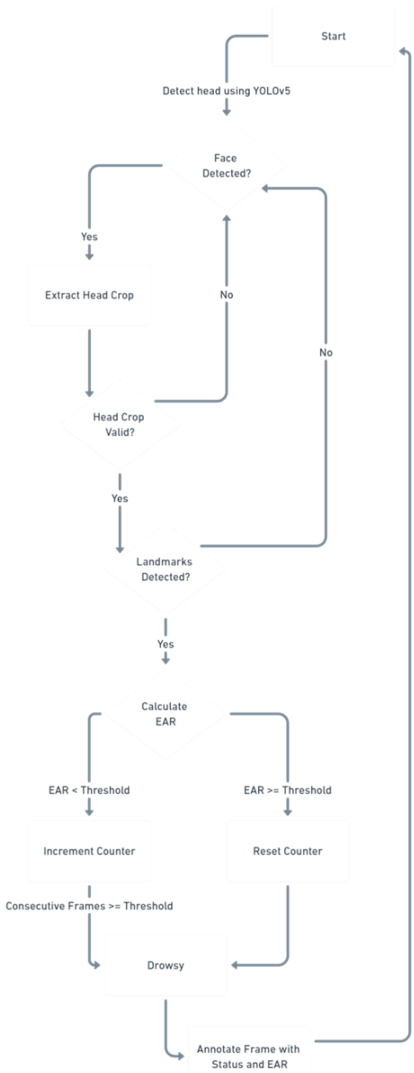

Figure 4 Block Diagram

V. CONCLUSION

The development and installing the drowsiness detection system have the ability to significantly improve the road safety by monitoring the driver and analyze the risks of driver getting fatigue. The integration of advanced technologies such as machine learning algorithms, computer vision, sensor data, etc, have demonstrated that our study has the viable approach to accurately detecting the early signs of drowsiness in drivers, and to give the proper alert Our experiment and results indicate that the proposed system, has a combination of facial feature analysis and real-time monitoring which can achieve high accuracy in identifying drowsy states of the driver making our system more efficient . The system's ability to process and analyze video using YOLOv5 and Media Pipe in real-time ensures timely alerts, allowing drivers to take necessary precautions before causing the accidents crash.

With successful promising results there are still several challenges in some areas to be focus in future research. Enhancing this system's quality across various environmental conditions, such as under different light conditions and road scenarios, is crucial for a broader range of situations which is important to be handled by the system. Also with the , integration of the drowsiness detection system with vehicles we can find the performance of the system performance by which we can understand its functioning of the system and can further help to improve the system's effectiveness and reliability.

In conclusion, in todays world implementation of this system is helpful for safety system purpose of people and to be done efficiently. Continuing the research work going we can move near to make more safer roads and reducing the incidence of fatigue-related accidents.

FUTURE SCOPE

With evolving in technology we integrated the YPLOv5 and Media Pipe which can accurately give the results i.e. to understand the drowsiness of the driver with the help of EAR(Eye Aspect Ratio) to find closure of the eye By integrating MediaPipe’s hand and body tracking with YOLOV5's neural networks, gesture control systems can be more accurate, offering potential in virtual reality (VR) or augmented reality (AR) for more immersive experiences.

We can integrate it to cars of different models as a safety security which will eventually reduce the number of accidents cause

Further with more detail research and evolving technology we can add more security facility which which will be more accurate and real time monitoring with quick response.

ACKNOWLEDGMENT

Authors thanks Dr. C. M. Mahajan Head DESH Department Vishwakarma Institute of Technology, Pune for his valuable and insightful guidance and providing us with the opportunity to work on this project.

REFERENCES

[1] https://doi.org/10.3390/app13116645

[2] https://doi.org/10.1117/12.2679987

[3] Ghanta Sai Krishna, Kundrapu Supriya, Jai Vardhan and Mallikharjuna Rao K., “Vision Transformers and YoloV5 based Driver Drowsiness Detection Framework”. arXiv:2209.01401v1 [cs.CV] 3 Sep 2022

[4] Redmon, J., Farhadi, A. (2018). YOLOv3: An Incremental Improvement. arXiv preprint arXiv:1804.02767.

[5] Lugaresi, C., Sorokin, A., et al. (2019). MediaPipe: A Framework for Building Perception Pipelines. arXiv preprint arXiv:1906.08172.

[6] Soukupová, T., Čech, J. (2016). Real-Time Eye Blink Detection using Facial Landmarks. 21st Computer Vision Winter Workshop.

[7] Dalal, N., & Triggs, B. (2005). Histograms of Oriented Gradients for Human Detection. CVPR.

[8] Kurylyak, Y., Lamonaca, F., & Mirabelli, G. (2012). Detection of the eye blinks for human's fatigue monitoring. IEEE International Instrumentation and Measurement Technology Conference.

[9] Park, J., & Lee, K. (2016). Driver Drowsiness Detection System Based on Feature Representation Learning Using Various Deep Networks. IEEE International Conference on Big Data and Smart Computing.

[10] Yuen, M. P., & Li, X. (2018). A deep learning framework for driver drowsiness detection. IEEE International Conference on Artificial Intelligence and Big Data.

[11] Rana, R., & Mishra, M. (2019). Real-time Driver Drowsiness Detection Using Convolutional Neural Networks. International

Journal of Innovative Technology and Exploring Engineering.
[12] Abtahi, F., Sefidgar, Y. S., & Whitehill, J. (2020). Deep Learning for Facial Expression Recognition in the Wild. IEEE Conference on Computer Vision and Pattern Recognition.
[13] Alshaqaqi, B., Alshamasin, M., & Ahmad, F. (2017). Hybrid drowsiness detection system using HOG and CNN. IEEE International Conference on Intelligent Transportation Systems.
[14] Lee, B. G., & Chung, W. Y. (2019). Driver alertness monitoring using fusion of facial features and bio-signals. IEEE Sensors Journal.
[15] Redmon, J., & Farhadi, A. (2018). YOLOv3: An Incremental Improvement. arXiv preprint arXiv:1804.02767.
[16] Bochkovskiy, A., Wang, C. Y., & Liao, H. Y. M. (2020). YOLOv4: Optimal Speed and Accuracy of Object Detection. arXiv preprint arXiv:2004.10934.
[17] Lugaresi, C., Sorokin, A., et al. (2019). MediaPipe: A Framework for Building Perception Pipelines. arXiv preprint arXiv:1906.08172.
[18] Zhang, Z., Tao, D., & Lu, B. (2020). Facial Expression Recognition using MediaPipe. International Conference on Pattern Recognition.
[19] Gupta, R., Choudhary, T., & Sharma, R. (2021). Evaluation of MediaPipe for Drowsiness Detection. IEEE International Conference on Computer Vision.
[20] Wang, Y., Luo, M., & Zhou, F. (2018). Comparative Analysis of Drowsiness Detection Techniques. IEEE Transactions on Intelligent Transportation Systems.
[21] Murugan, D., & Sankaranarayanan, S. (2020). A systematic review of driver drowsiness detection systems. Journal of Traffic and Transportation Engineering.