Tuijin Jishu/Journal of Propulsion Technology
ISSN: 1001-4055
Vol. 45 No. 2 (2024)

A Deep Learning Approach for Driver Monitoring System

Lakshmi Prasanna,Srinivas Malladi

Koneru Lakshmaiah Educational Foundation (KL University) CSE Department, Vaddeswaram, AP, India

Abstract:- In recent years, the alarming rise in accidents attributed to distracted and drowsy driving has prompted a concertedeffort by automotive researchers and manufacturers to implement innovative technological solutions. The proposedproject uses state of the art Convolutional Brain Organizations (CNNs) to recognize basic states, working throughtwo interconnected modules. The primary module examines ongoing pictures or video feeds to remove facial elements, critical for grasping the driver's state. These highlights act as contribution for the subsequent module, which utilizes a lightweight profound learning design enhanced for edge gadgets. This approach improves discovery precision, giving an exhaustive arrangement contrasted with past techniques utilizing multi-facet perceptrons

Keywords: *Deep learning, Caffinet Model,Blazepose .*

1. Introduction

Since technology is developing so quickly in the modern world, maintaining the safety of our roadways is a major problem. One of the main issues is that driving while intoxicated or distracted might result in collisions. Experts have been working to integrate new technologies into cars to make them safer, but the systems are too large and require more processing power.

Recent times have seen significant advancements in the field of deep learning . Convolutional Neural Networks (CNNs) are a unique type of neural network that we are using to develop a new driver monitoring system. Our goal is to optimise this technology for use on tiny computers that can be installed in automobiles. Our goal is to optimise this technology for use on tiny computers that can be installed in automobiles. We hope to usher in a new era of safety technology that is user-friendly for all and effectively prevents accidents caused by inattentive or fatigued drivers.

Our principal objective is to ensure that fewer accidents occur as a result of fatigued or distracted driving. We aim to save lives and improve driving safety for all by utilising smart technology that can identify when a driver is in danger and alert them immediately. This initiative is about ensuring that everyone gets where they're going safely, not simply about sophisticated technology.

2. Objectives

A lightweight deep learning model that can operate effectively on embedded boards and reliably identifydriver fatigue, distraction, pupil tracking, eye blinking, yawning, and head posture is desperately needed. An ideal model would be smaller than 10 MB, have an inference time of less than 1 second, andhave a frame rate of at least 5 on embedded systems.In order to close the gap between edge devices and high-end devices with GPUs, the system should not only recognise these situations but also notify the driver in accordance with those findings.

3. Methods

In this section we discuss various methods engaged with driver monitoring system. Further these methods will be ported to an implanted gadget (raspberry pi) to reallytake a look at it's working. A DMS framework will take a constant video as an information.

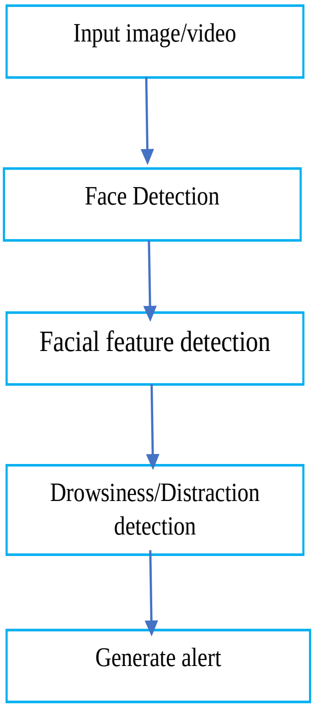

Figure-1 Basic Model of DMS

As seen in Figure 1 above, the model will first receivean image or video as input. The driver's face will thenbe identified by this model from the video, and it willthen continue. Nevertheless, the algorithm will keep searching for a face and won't start extracting features until a face is found. The driver's condition is ascertained by extracting the mouth, eyes, and lip landmarks once the face has been identified. Eye blinking and yawning are signs of sleepiness, whereasturning one's head away suggests distraction. An alertis set off if the driver's head is not looking ahead for more than ten frames. This is considered a sign of distraction. Next, facial features are examined: the irisarea suggests eye blinking, and the mouth aspect ratioindicates yawning. With a 15-frame closurethreshold, alerts are generated whenever mouth opening or eye closure thresholds are exceeded. The whole detection process is depicted in the flowchart in figure-2.

Three lightweight modelsfrom Mediapipe are used by the driver monitoring system- Mediapipe Blazeface ,Iris,PoseBlazeface for face landmark detection, Iris for irisIn order to match object scale ranges, the anchor scheme employed in the model adopts a method of defining anchors at several resolution levels.Aggressive down sampling is used to maximise computing resources. This model ends at the 8x8feature map dimensions without performing any more downsampling, in contrast to normal SSD models that employ predictions from several feature map sizes. Anchors at 4x4 and 2x2 resolutions are replaced by six anchors per pixel at 8x8 resolution. Anchors are restricted to the 1:1 aspect ratio, which was determined to be adequate for accurate face detection due to the minimal variety in human face aspect ratios.

Post processing: In post-processing, we use a blending method rather than the suppressionalgorithm to address the increased number of anchorsoverlapping a particular object as a result of the feature extractor's resolution not being reduced below8x8. This technique calculates a bounding box's regression parameters as a weighted average of overlapping predictions, hence minimising the phenomenon. This change leads in a 10% gain in accuracy for our face detection task detection, and stance for stance determination. These models guarantee effective performance when implemented on embedded boards.

Mediapipe Blazeface A portable variant called Mediapipe Blazeface is intended to identify one or more faces photos takenwith smartphones, with a focus on back-facing camera shots. It has been adapted from the Single Shot Multibox Detector (SSD) architecture and is optimised for inference on mobile GPUs. In order to reduce the total number of bottlenecks needed toobtain a certain receptive field size, the model uses 5x5 kernels in its architecture bottlenecks to increase the receptive field sizes. In terms of computing power, this method is economical.

MediaPipe-Iris MediaPipe Iris is a machine learning approach for precise iris estimation that can followat almost noadditional expense when compared to the previous NMS method.

Mediapipe Pose BlazePose is a lightweight convolutional neuralnetwork intended for constant human posture assessment on cell phones. It produces 33 body central issues for a solitary individual, making it ideal for applications like wellness following and communication via gestures acknowledgment. The model uses an encoder-decoder network engineering to foresee heat maps for all joints, trailed by another encoder that straightforwardly relapses to the directions, everything being equal. During induction, the heatmap branch can be disposed of, making it sufficiently lightweight to run on a cell phone..An indicator tracker arrangement is utilized which shows great constant execution on different errands like hand milestone expectation and thick face milestone forecast.

4.Implementation

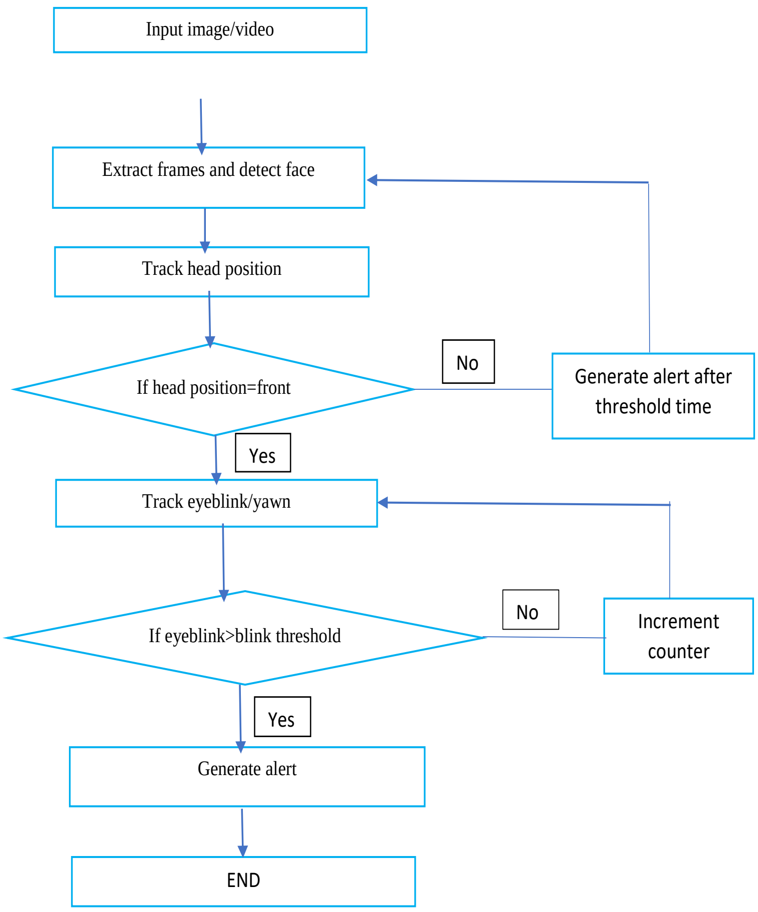

Figure-2 Flowchart of DMS

Data preparation is vital in machine learning, especially for applications such as the Driver Monitoring System (DMS), which require precise predictions. For DMS, we require photos of faces that have been annotated for features such as lips, eyes, and faces. Our custom data set consisted of approximately 800 photos, which is minimal for training. To overcome this, data augmentation techniques were used to boost the dataset's diversity and size. Each of the original 800 photos underwentten augmentation procedures, yielding around 8000images. These were divided into three sets: training,testing, and validation, with roughly 85% (6800 photos) used for training and the remaining 15% (1200 images) for testing and validation. The complete data-set was annotated for 10 classesusing LabelImg application. LabelImg is used to label the images in a square shape. It can store the images in .xml file extension which can be accessed. XML file contains the size, class, image dimensions and other image related informationfor every image in a single file.

4. Results

6800 images were used to train three models, with a batch size of 8 yielding better learning rates compared to a batch size of 4. The complete model achieved an average accuracy of 0.8985 with a learning rate of 0.001. Model validation included evaluation parameters such as precision, recall, and F1 score calculated using the confusion matrix, which summarizes classifier performance.

Total Images	8000
For training(85%)	6800
For testing and validation	1200

Table-1 Dataset Information

Sr. No.	Parameters	Result
1	Train Images	6800
2	Epoch	60000
3	Batch Size	8
4	Step Size	7500
5	Learning Rate	0.001
6	Input Image size	360*240
7	Average Accuracy	0.8985
8	Loss	0.0158

Table-2 Training Result

	Precision	Recall	F1
DMS Model	0.8948	0.8413	0.8668

Table-3 Validation Results

Hardware Implementation

The implementation is done on a Raspberry Pi 4 board after converting from TensorFlow to TFLite.TFLite is recommended over the original TensorFlow format since it has a reduced memory footprint and performs more efficiently.

Raspberry Output

The final DMS implementation on Raspberry Pi yielded desired results, detecting yawning, eye blinks, and head position accurately with minimal lag. These features enable effective detection and alerting for driver drowsiness and distraction. InitialFPS was around 4-5 frames but improved to 7-8 FPS. Inference time for complete facial feature detection was approximately 94 milli seconds.The model's total memory on Raspberry Pi is 8.84 MB, comprising the three models of the DMS. Accuracyand precision are around 89.89% and 89.48%, respectively**.**

Sr. No.	Parameter	Result
1	FPS	7-8
2	Inference time	94ms

3	Accuracy	89.89%
4	Precision	89.48%
5	Memory Usage	8.84 MB

Table-4 Raspberry Pi Results

5.Comparision between Models

The Caffenet and Mobilenet models are Tensorflow models with sizes of 43MB and 28MB, respectively, whilst the DMS model has a total size of8.84MB. Furthermore, the Mobilent model has a higher FPS of approximately 55 when compared tothe Caffenet and DMS models. However, the DMS model works well on the Raspberry Pi, which aids in achieving the desired result.

6.Conclusion

Deep learning algorithms can be used on embedded boards to monitor the driver's status with more precision and accuracy. The successful implementation on Raspberry Pi demonstrates the feasibility of employing lightweight models to detect driver tiredness and attention. However, for efficient detection and alerting, the distance between the DMS camera and the driver should not exceed 75cm, whichcorresponds to the average distance between the driver and the car's steering wheel or windscreen. Themodel's accuracy is influenced by dark or low-light circumstances, which can be improved bysupplementing the training dataset with relevant photos or employing infrared imaging.

Futurescope

Later on, the DMS model can be upgraded by consolidating elements like article recognition inside the vehicle, following cell phone use, and distinguishing smoking or liquor utilization by the driver in view of their way of behaving. Carrying out these elements will empower the DMS model to screen the driver as well as the climate around the vehicle, altogether upgrading security measures.

References

[1] Kim, W., Jung, W.S. and Choi, H.K., 2019. Lightweight driver monitoring system based on multi-task mobilenets. Sensors, 19(14), p.3200

[2] Musale, T. and Pansambal, B., 2016. Real TimeDriver Drowsiness Detection system using Imageprocessing Research in Engineering Application & Management(IJREAM), 2(08).

[3] Anupriya R, Binishamol Joby, Liya Joy, Neetha K Nataraj,2020. Driver DrowsinessDetection and alert system based on behavioural measures (IRJET).

[4] Kabir, A.T., Zinnurayen, N., Abrar, T. and Ta-sin, A.J., 2020, August. The Smart VehicleManagement System for Accident Prevention by Using Drowsiness, Alcohol, and Overload Detection. In 2020 10th Electrical Power, Electronics, Communications, Controls andInformatics Seminar (EECCIS) (pp. 173-177).

[5] Bamidele, A., Kamardin, K., Syazarin, N., Mohd, S., Shafi, I., Azizan, A., Aini, N. and Mad, H., 2019. Non-intrusivedriver drowsiness detection based on face and eye tracking.

[6] Luo,X.Q.,Hu,R.and Fan, T.E., 2013, June. The driver fatigue monitoring system based on face recognition technology .In 2013 Fourth International Conference on Intelligent control and Information Processing (pp. 384-388). IEEE.

[7] Pondit, A., Dey, A. and Das, A., 2020, December. Real-time Driver Monitoring System Based on Visual Cues. In 2020 6th International Conference on Interactive Digital Media (ICIDM)(pp. 1-6). IEEE.

[8] Louie, W.Y.G. and Mansour, I., 2019,October.Towards a driver monitoring system forestimatingdriver situational awareness. In 2019 28th IEEE International Conference on Robot and Human Interactive Communication (RO-MAN) (pp. 1-6). IEEE.

[9] Zhu, L., Xiao, Y. and Li, X., 2021, January. Hybrid driver monitoring system based on Internet of Things and machine learning. In 2021 IEEE International Conference on Consumer Electronics and Computer Engineering(ICCECE) (pp. 635-638). IEEE.

[10] Begum, S., 2013, October. Intelligent driver monitoring systems based on physiological sensor signals: A

review. In 16th International IEEE Conferenceon Intelligent Transportation Systems (ITSC 2013) (pp. 282-289). IEEE.

[11] Viola, P. and Jones, M., 2001, December. Rapid object detection using a boosted cascade of simple features. In Proceedings of the 2001 IEEE computersociety conference on computer vision and pattern recognition. CVPR 2001 (Vol. 1, pp. I-I). IEEE.

[12] Howard, A.G., Zhu, M., Chen, B., Kalenichenko, D., Wang, W., Weyand, T., Andreetto,M.and Adam, H., 2017. Mobilenets: Efficient convolutionalneural networks for mobile vision applications. arXiv preprint arXiv:1704.04861.

[13] Bazarevsky, V., Kartynnik, Y., Vakunov, A., Raveendran, K. and Grundmann, M., 2019Blazeface: Sub-millisecond neural face detection on mobile gpus. arXiv preprint arXiv:1907.05047

[14] Ablavatski, A., Vakunov, A., Grishchenko, I., Raveendran, K. and Zhdanovich, M.,2020. Real-time Pupil Tracking from Monocular Video for DigitalPuppetry. arXiv preprint arXiv:2006.11341

[15] Bazarevsky, V., Grishchenko, I., Raveendran, K.,Zhu, T., Zhang, F. and Grundmann, M., 2020. Blazepose: On-device real-time body pose tracking. arXiv preprint arXiv:2006.10204