Industrial Engineering Journal
ISSN: 0970-2555
Volume : 52, Issue 8, No. 4, August : 2023

A CONVOLUTIONAL NEURAL NETWORK-BASED SYSTEM FOR DETECTING DROWSY DRIVING

Mr. Pranav A. Natekar, PG Student, Department of Computer Science & Engineering
Mr. Ravindra S. Kambale, Assistant Professor, Department of Computer Science & Engineering
D. Y. Patil, Agriculture & Technical University, Talsande, Kolhapur, Maharashtra, India..

Abstract
Deaths caused by automobiles have increased dramatically worldwide in recent years. Therefore, protecting motorists and pedestrians on the world's roads is an issue of paramount importance. Injuries and fatalities from automobiles continue to rise, making this a pressing public health concern. Many current innovations have increased the understanding and technical capabilities of automobiles, allowing them to observe and study road scenarios to prevent accidents and safeguard passengers. Drowsiness in specific, has been recognized as one of the biggest and most important topics of study in the past few decades. This is because drowsiness is an aspect that contributes to the highest accident risk that is the most common cause of fatalities that occur on the world's roadways. In this study, we employ a Convolutional Neural Network (CNN) to a collection of frames of a driver's face to detect and predict drowsiness We utilized a collection of data to guide the development of the proposed approach, and then employed 3D CNN with a Recurrent Neural Network (RNN) architecture to identify driver drowsiness. We were able to create a real-time driving tracking system to decrease collisions on the road since following training, we acquired an appealing accuracy which reaches a 92% acceptability rate.

Keywords Video, CNN, Open-CV, Classifier.

I. Introduction

Everyone in today's society relies on some form of transportation. It is frequently regarded as a comfort, yet in today's world, the daily routine of ordinary individuals cannot function without it. The protection of individuals, as well as the security of their automobiles when it comes to the possibility of a collision or vandalism, is a major worry for people. Automobiles serve several functions, including public transportation, cargo transportation, and personal travel. The driver usually gets fatigued and begins to feel drowsiness after very lengthy driving hours; however, they continue driving towards the purpose of arriving at the destination as quickly as feasible. The driving under unsafe or unfavorable conditions that causes the driver to work harder. Even though driving while asleep is extremely dangerous, the person may nevertheless want to do it. When this happens, the driver loses consciousness and the car accelerates out of control, causing multiple deaths.

Road along with climate conditions, as well as a vehicle's mechanically efficiency, are just a few examples of external factors that can contribute towards an accident's occurrence. Nevertheless, negligent driving accounts for the vast majority of collisions. Reckless driving, exhaustion, and drowsiness are all forms of driver error. Spontaneous reflexes, familiarity, and awareness are just a few of the characteristics that might influence an individual's skill behind the wheel. Thus, sleepy driving is a major contributor to the countless lives lost each year in collisions with vehicles.

Smart automobiles include powerful software programs that manage a variety of driving functions, including motor acceleration, transmission, brakes, steering, etc., to make the experience of driving safer and more pleasurable. The earliest automated navigational systems for cars was developed using Ad hoc networks. These technologies' inability to react instantly to shifting conditions is a clear shortcoming. Whenever time is the primary factor, the decision-making authority rests with the driver. Driver drowsiness can also be assessed by observing their physical state or by using expressions on their faces. For this reason, the need for an accurate "System that recognizes driver drowsiness" is seen as an appealing motivation for its development. The previous method cannot be relied upon in any

way. In order to get accurate readings, extremely sensitive sensors would have needed to be affixed onto the driver, which could be annoying. Keeping tabs on correctness was additionally a time-consuming ordeal. The second method is a non-invasive method of evaluating restlessness by alterations in physical movements or situations such as open/closed eyelids. Another indicative of exhaustion were short sleep durations. In this way, early warning can be accomplished by persistent eye surveillance and detection. In order to reduce the number of accidents caused by drivers falling asleep at the wheel, it has become necessary to advance the state of the art in driving aid systems. This approach is useful for maintaining driver focus.

A. Face Detection

An efficient approach of detecting objects, Haar features-based cascading classifiers are used for identifying faces. This method relies on machine learning and also involves training a cascading function using examples of both negative and positive images. The results of this analysis are then applied to the analysis of other images to identify things. In this work, we'll be focusing on face recognition. To begin, the algorithm must be provided with a large number of examples of positive as well as negative images (i.e., images that do not contain faces). The next step is to take characteristics from that data. The image beneath demonstrates the application of the Haar characteristics in this context. They are identical to the convolutional layer that we have. Every feature is represented by an unique number calculated by dividing the total number of pixels within the white box with the total number of pixels within the black box.

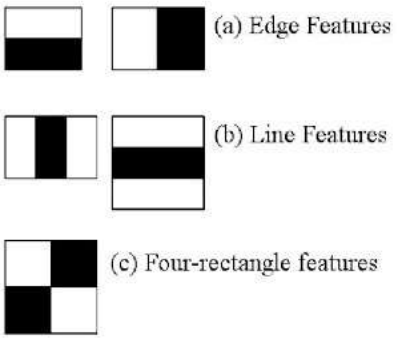

Figure 1. Haar cascading features.

B. Eye Detection

To implement eye identification, we utilized face-landmark prediction. The nose, eyes, jawline, eyebrows, nose, and mouth constitute all examples of face-landmarks that can be utilized to identify and depict these important anatomy characteristics. Aligning faces, estimating body poses, switching faces, detecting blinks, as well as many other tasks have each benefited from the use of face landmarks. By applying shape forecasting techniques, we hope to identify key facial features within the larger context of face-landmarks. Thus, there are two stages involved in face-landmark detection: Find the person's face within this image: Key facial characteristics are identified when a facial image has been localized using cascading classifiers that rely on Haar features. ROI: Multiple face-landmark detectors exist, but they all aim to identify and categorize the same basic face features: Lips, Eyebrows, Eyes, Nose, and finally, the Mouth.

C. Recognition of Eye’s State

Estimating the eye region using optic-flow, sparse monitoring, or frame-to-frame brightness distinction with dynamic thresholding allows for the determination of the extent to which the eyes have been dimmed by the user's eyelids. As an alternative, one can try to determine the eye's openness condition using a single image using techniques like template-based connection identifying for both closed and open eyes, heuristics vertical or horizontal image brightness projections across the area around the eye, parameterized modeling to locate the lids of the eye, or present shape representations.

In terms of the comparative face-camera position (head position), picture image quality, brightness, movement dynamics, etc., the prior methodologies often unintentionally put too stringent constraints on the configuration. Although being able to process images in instantaneously, heuristic approaches that rely on raw image brightness tend to be extremely sensitive. Each video frame's eyes are individually analyzed to determine their aspect ratio. A person's Eye-Aspect-Ratio (EAR) is determined by measuring the angle among their eye's vertical and horizontal axes.

EAR=(|(|p2-p6|)|+||p3-p5||)/(2||p1=p4||)

Where p1, p2,... p6 are coordinates of 2D landmarks.

When one of the eyes is wide open, the EAR is relatively stable, but it approaches zero whenever the other eye is closed. It ignores the position of the user's head and body to some extent. The open-EAR is highly consistent between subjects and is completely invariant to both homogeneous image scaling as well as in-plane facial rotations. Because blinking occurs simultaneously across both eyes, their EARs are averaged.

Finally, the estimated EAR is used to make the overall assessment of the eye condition. The eyes are considered to be closed down if the distance between them is either 0 or very near to 0 and "opened" else. The last stage of the method involves detecting whether or not an individual is sleepy using some established criteria. A human blinking typically lasts between 100 and 400 milliseconds (or 0.1 and 0.4 of a second). Therefore, an individual's eye closing must happen after this time period to confirm that he is indeed tired. A maximum duration of five seconds was decided upon. When the eyes remain motionless for a period of five seconds or even more, the system assumes the user is sleepy and displays a warning message.

D. Open-CV

Shading-based approaches are the simplest way to detect and isolate an object in an image. To successfully divide objects using shading-based techniques, there must be a significant shading difference between an object along with the background. Whenever possible, OpenCV will record both videos and pictures in the uncompressed 8-bit BGR form. What this means is that collected images are capable of being thought of as three matrices, each containing a numerical value between 0 and 255 and labeled RED, BLUE, GREEN, (which is why the term RBG).

II. Literature Survey

This section has covered the different methods offered by investigators in the past few decades for detecting drowsiness and blinking. An approach for detecting faces utilizing cascading classifier that utilize Haar features was put forth in 2016 by Manu B.N. The approach requires both positive (pictures containing faces) as well as negative (pictures lacking faces) examples for training the classifier which will ultimately recognize an object. In order to identify the facial area, a cascaded Adaboost classification is used in conjunction with several Haar feature-based classifiers. The corrected image is subsequently divided across many different rectangular sections, which can be placed and scaled wherever around the original image. Haar-like features are effective for face recognition in real time because of the diversity of human features. Adaboost's approach will permit all face portions while rejecting non-face portions of images, therefore these differences are able to determined using the total number of each pixel throughout a rectangle's region.

An approach for detecting drowsiness utilizing the state of eyes monitoring using an eye blink technique was suggested by Amna Rahman around 2015. Initially the image is gray-scaled, and then the Harrison corner identification technique is used to locate the outer edges around the upper and lower eyelids. Following the points have been traced, a straight path is going to be drawn connecting the two highest points, the center will be determined from the line, while the lowest point is going to connect to the center. Now, the identical technique is followed for every image, with the eye condition is determined by calculating the distance that exists 'd' between the image's center and its lowermost position. The estimated length 'd' is then used to make the final selection for the eye's condition. The

eyes are considered to be closed down if the range between them is either 0 or very near to 0. and "open" alternatively.

III. Proposed Method

There are a variety of techniques available for detecting drowsiness. Electrodes are employed in this procedure, which is considered to be an invasive approach, in order to acquire data relating to the subject's brain activity, heart rate and pulse rate. The Electrocardiogram (ECG) measures heart rate variability and can be utilized to diagnose drowsiness caused by a variety of factors. Electromyogram (EMG), Electroencephalogram (EEG) and ECG signals are correlated in order to determine if an individual is sleepy, then a result is subsequently given. Behaviorally based methods include keeping a watch on an individual's blink rate, head position, and other indicative signs of sleepiness and alerting them accordingly.

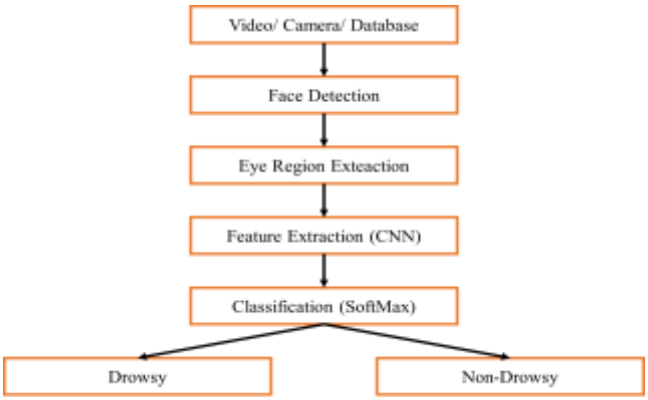

Figure 2. Flow of the presented work.

In Figure 2, the flow of the presented work has been presented. Consider an Image captured by a camera an Input, to determine if an individual is sleepy, one must first locate their facial features in the image then draw a Region-of-Interest (ROI), then use ROI to locate their eyes before feeding that information into a classifier, which will then determine if the eyes were closed or open.

A. Hair Cascade Classifier

Classification is performed using the Haar Cascading Classifier, which takes into account the features of the objects to be categorized. Utilizing the Extensible Markup Language (XML) files generated by the Haar cascading classifier, characteristics of the individual are extracted from the images; these characteristics can then be used throughout the subsequent processing steps. The Haar cascading classifier is used for obtaining characteristics from images. Face detection along with ROI creation within the retrieved image: To identify features within the retrieved image, we don't have to depend on color data. Images have to be converted to grayscale before they are able to identify the facial area. For recognition of faces, we'll be utilizing XML files generated by the Haar cascading classifier. The classifier is fed information about the identified eyes within the ROI. Identical techniques as those employed during eye and face identification. To begin detecting eyes utilizing [3], we must initially configure our cascading classifier for both of the eyes utilizing Right Eye (reye) and Left Eye (leye). Data about the eyes will be successfully retrieved from the images. This is something that can be accomplished by employing the boundaries of the boxes that surrounds the eye, and as a result, we are able to retrieve the shape from an eye using an image. Leye solely stores left-eye information. Following that, feed that information into a CNN classifier to get a prediction of something like whether or not the subject's eyes are open. The right eye's information can be input using reye using the identical manner. CNN will determine if a person's eyes were open [4]: The condition of the eye is predicted with the help of a CNN classifier. The system will activate the alarm whenever the

individuals eyes remain closed for longer than a predetermined threshold. The sound.play() function generates a loud sound to warn the individual.

Figure 3. Eye detection feature

A CNN is a type of Deep-Learning algorithm that may recognize features in an input image, allocate relevance (learnable biases and weights), and then distinguish between distinct characteristics and objects within the image. In contrast to different classification techniques, CNNs need significantly less time for pre-processing. CNNs are able to acquire filters/characteristics independently without necessary training, whereas in early approaches they must be hand-engineered. CNN architectures takes cues from the visual cortex's hierarchical structure, which is similar with the interconnection structure of the brain's human neurons. Only some areas of the visual field, referred to by the term the receptive field, are capable of generating responses from each neuron when they are stimulated. In order to encompass the complete field of view, multiple fields of view must overlap.

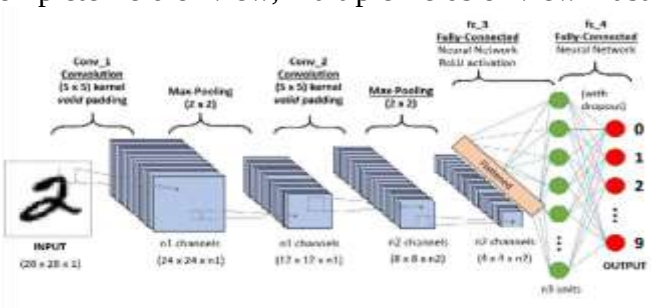

Figure 4. CNN layers.

The approach may achieve an average precise score whenever predicting classes for very simple binary images, however it will likely fail miserably when used to more complicated images with interdependent pixels. By employing the proper filters, the CNN may accurately record the temporal and spatial relationships present in an image. As a result of using fewer variables and reusing weights, the system's design achieves superior matching for the image database. This means that the network may be trained in order to understand more complex images. An RGB image broken down into its individual red, blue and green components. The CNNs task is to simplify the images despite sacrificing information crucial to making an accurate forecast. In order to develop a system that excels at learning characteristics and can scale to large datasets, this information is crucial.

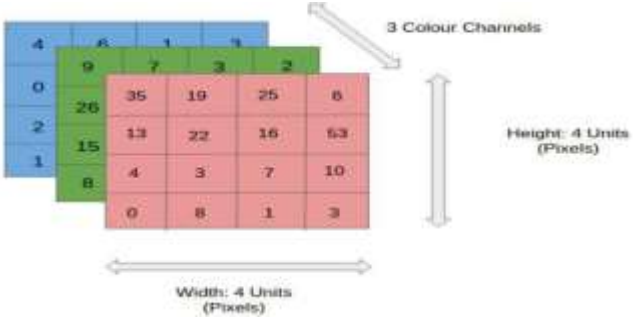

Figure 5. Convolutional layers.

Industrial Engineering Journal
ISSN: 0970-2555
Volume : 52, Issue 8, No. 4, August : 2023

The formula for the size of an image is as follows: Image Dimensions = 5 (Height) × 5 (Breadth) × 1 (Number of channels,eg.RGB). The green part in the preceding Figure is similar to our 5×5×1 source image, which is denoted by I. Kernel/Filter,K, denoted in yellow, represents the component of a CNN Layer's initial stage responsible for performing the convolution process. K is a 3x3x1 matrix that we've decided to use. At Distance Length = 1 (Non-Strided), the center of the kernel moves nine times, every single time executing a matrix multiplication among K alongside the region P for the image across which it hovers. With a fixed Distance Value, the filter's processing will continue to the right until it has processed the entire width. After that, it uses the identical Distance Value to return towards the image's left-hand top corner and start over again. When dealing with multi-channel images (RGB, etc.), the Kernel's depths match those of the original image. After performing matrix multiplication across the KN and in stacks ([K1,I1]; [K2,I2]; [K3,I3]), we add up all of the findings and add in biases to get a compressed one-depth channels Convoluted Feature Result.

Average (AvgPol) and Maximum Pooling (MaxPol) are the two forms of Pooling. With MaxPol, you get back the highest value that is possible through the region of the image that was processed through the Kernel. AvgPol, on the opposite conjunction, provides an average for each of the image elements inside the region encompassed through the Kernel. MaxPol can also be used to reduce background noise. All noisy stimulation is thrown out, and dimensionality-reduction along with de-noising are also performed. However, AvgPol relies solely on dimensionality reduction to minimize unwanted noise. Therefore, MaxPol is clearly superior to AvgPol. In a CNN, the i-th layer consists of both the Pooling and Convolutional Layer. The total amount of these layers can be raised to capture low-level information more effectively, but doing so requires more computer power and is only possible if the images are sufficiently complex. After completing the steps above, the approach is now able to comprehend the characteristics. The next step is to normalize the output then input it into a standard Neural Network for classifying.

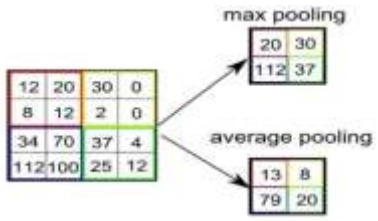

Figure 6. Pooling Methods.

The addition of a Fully-Connected Layer (FCL) is a (typically) low-cost strategy for mastering nonlinear combinations based on the high-level data captured through the CNN layer's outputs. In this context, the FCL begins to acquire a function, which may or may not be linear. After we have finished transforming our input image through a format that is appropriate to use with the Multi-Level Perceptron, we will proceed with flattening the image throughout a column-based vector. A neural network that is feed-forward receives the normalized outputs and is trained using backpropagation with each iteration. The system uses the Softmax Classifier method to categorize images over time, learning how to differentiate among dominant and specific low-level attributes.

.

IV. Result and Discussion

Industrial Engineering Journal
ISSN: 0970-2555
Volume : 52, Issue 8, No. 4, August : 2023

The first step in getting the system up and running is opening a command prompt and navigating to the folder containing the "drowsiness detection.py" program file. It is recommended that you use this command while running the script.

Figure 6. Command Prompt.

Figure 7. Detecting Drowsy Driving.

The time it takes to launch the web camera and initiate detection can vary. When the user's eyes remain closed for too long, the score will rise until it reaches a certain level, at which point the system will sound an alarm. We have used web camera images as input to ensure that the built system is functioning as intended. A web camera takes a number of images at regular intervals and feeds them into the computer. The system retrieves the Eye's ROI to determine its current state. If at least one of the eyes is closed, the score is reset to zero; else it rises by one each time. In Figure, it can be seen that the algorithm has detected that one eye has been opened and assigned it a score of 0. This prevents the onset of drowsiness. Take a look at Figure, the individual is getting sleepy hence their eye state shifts from partially open to partially closed and the score continues rising. However, no alert is given since the result is below the set threshold. As seen in Figure, if a user's eyes remain closed for too long, their scores will rise until they approach a threshold, at which point a warning will be sent out to get them out of their drowsiness

V. Conclusion

The proposed technology can identify drowsiness by monitoring an individual's rate of blinking and eye opening/closing patterns. When drowsiness sets in, the approach will contrast the ratio of open to closed eyes to a predetermined threshold. The device will inform the user and attempt to continue keeping him awake if the measurements go outside of a predetermined range, indicating drowsiness. If the imaging device has higher output, this approach will still function well in dark conditions. A portable approach is provided by the framework that has been described, that can be simply executed on a variety of devices. The suggested approach does not necessitate any expensive equipment or physical equipment unlike those currently on the marketplace.

References

[1] Kirti Dang, Shanu Sharma, “Review and comparison of face detection algorithms”, 7th International Conference on Cloud Computing, Data Science & Engineering - Confluence, 2017
[2] Jongkil HyunCheol-Ho Choi, Byungin Moon, “Hardware Architecture of a Haar Classifier Based Face Detection System Using a Skip Scheme”, IEEE International Symposium on Circuits and Systems (ISCAS),2021
[3] jain, “face detection in color images; r. L. Hsu, m. Abdel-mottaleb, and a. K. Jain.
[4] Wang Yang;Zheng Jiachun, “Real-time face detection based on YOLO” 1st IEEE International Conference on Knowledge Innovation and Invention (ICKII), 2018
[5] Kartika Candra Kirana,Slamet Wibawanto,Heru Wahyu Herwanto, “Redundancy Reduction in Face Detection of Viola-Jones using the Hill Climbing Algorithm”, 4th International Conference on Vocational Education and Training (ICOVET), 2020