No camera to know(ing) distraction: An universally compatible device that utilizes a novel CNN to combat distracted driving

Yash Dagade and Ronia Mukherjee, Eden Prairie High School

ABSTRACT

It is well established that drowsy and distracted driving are significant contributing factors to motor vehicle accidents; with the National Highway Traffic Safety Administration estimating that drowsy driving was responsible for over 938,000 crashes, contributing to almost 15% of the total crashes in 2018 alone. Despite the clear need for a device that can detect drowsiness and distraction in real-time to help prevent such accidents, there are no readily-available commercial solutions on the market that can be implemented on the cars you already own. Hence we created EyeDA, a compact device that can be utilized in any vehicle to alert drivers when it detects potential distractions or fatigue. Our approach is to create a novel Convolutional Neural Network (CNN) that will be trained on data from an IMU (Inertial Measurement Unit - comprising an accelerometer and a gyroscope) and a microphone. IMU allows for the measurement of orientation, acceleration, angular velocity, and speed while a microphone enables enhanced accuracy by capturing audio cues. By including this combination of sensory inputs in our model, we can improve the efficiency of our predictions and create a more comprehensive system for detecting driver distraction and fatigue. Additionally, using non-visual features allows for the development of a computationally efficient CNN solution that can be implemented on low-power devices such as the Jetson Nano.

EyeDA aims to develop a universally compatible, cost-effective device that can accurately detect distracted driving and make our roads safer for all users.

Motivation

While algorithms that combat distracted and fatigued driving exist, they usually require you to upgrade to a new car and are hidden behind a massive paywall. The need for a small, inexpensive, A.I.-based device that can be installed on all vehicles to detect and respond to sleepy or distracted drivers is essential. The need for such a device amongst teens is paramount because they have less driving experience and are more likely to drive an older vehicle, which often lacks up-to-date safety features.

Approach and Project Logistics

Currently, EyeDA uses an image processing algorithm that takes real-time footage of the driver's face and uses the nose position to alert the driver when distracted. The hardware components of EyeDA are as follows: Raspberry Pi Camera, a Raspberry Pi processor, and a beeper. However, the use of traditional computer vision approaches results in EyeDA being less accurate and unfit for everyday use. Images 1 and 2 show the first two iterations of EyeDA.

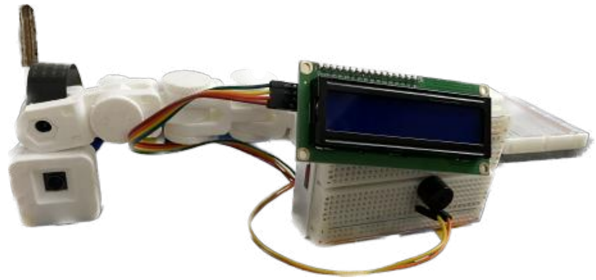

Image 1: The proof of concept, V1

Image 2: Refined V1

The current version of EyeDA is insufficient in several key areas. Firstly, the use of OpenCV for image processing results in slower overall performance, limiting the usefulness of the device in real-time situations. Additionally, the physical size of the device is relatively large, which may be a hindrance to its adoption by users. Finally, the accuracy of the device is limited by its reliance on image processing, which can be affected by factors such as lighting conditions and the presence of obstructions. In order to overcome these limitations, it will be necessary to adopt a more advanced approach, such as the use of a neural network.

Neural networks have been shown to have high accuracy rates in tasks such as image classification, and have the ability to learn and adapt to new data which is beneficial in detecting distractions and fatigue in real time. Yan et al, 2016, used a CNN to achieve an accuracy of 97.76% in detecting driver distraction based on visual features, while Lee et al, 2018, used a CNN to achieve an accuracy of 99.95% in detecting aggressive driving based on images of the driver's face captured by near-infrared and thermal camera sensors.

Techniques for detecting driver behavior that rely on visual features, such as cameras, can be sensitive to changes in light intensity, which can decrease their accuracy. Additionally, these techniques may be perceived as intrusive by drivers, who may be uncomfortable with the use of sensors like cameras. Furthermore, image processing techniques typically require high computational power, which may not be suitable for real-time applications or for use in embedded systems in vehicles.

There are several tools for analyzing driver behavior based on non facial features. For example, in Carmona et al. 2015, an algorithm is presented that uses Global Positioning System (GPS), an Inertial Measurement Unit (IMU), and in-vehicle sensors to collect data on driver behavior, and compares the maximum, mean, and standard deviation of this data to standard research on human factors to detect aggressive behavior. Dai et al. 2010 propose a system for identifying and alerting about drunk driving, which uses the accelerometer sensor of a smartphone to gather data and compares it to patterns of drunk driving behavior. Wang et al, 2017 propose a method based on vehicle speed-time series for recognizing risky driving behaviors, using a tuple of speed change consisting of the value and duration for each driver and a Support Vector Machine (SVM) classifier to identify risky drivers. This method achieved a classification accuracy of 95%. However, it only focuses on one risky behavior; it does not consider other unsafe behaviors such as drowsiness or distraction, nor does it consider behaviors such as driving at a constant speed or risky steering. Shahverdy et al. 2020 stand out among other papers due to their relevance to our project, as we are interested in developing a compact device that can use non-visual data to detect driver distraction and fatigue while being computationally efficient and accurate. Table 1 demonstrates the accuracy of such a device.

Various models with different number of convolutional layers, number of filters in the 1st layer, and filter size. The number of filters in the 2nd and 3rd convolutional layers is twice the number of filters in the 1st convolutional layer. Number of parameters, accuracy, and computational complexity are listed for each model.

Model#	#of Conv. Layers	# of Filters in the 1st Layer	FilterSize	#of Params [K]	Computational Complexity [M.FLOP]	Accuracy[%]
1	2	16	2 × 2	3.433	0.007	88.79
2	2	16	3 × 3	6.233	0.012	97.82
3	2	16	5 × 5	15.193	0.030	99.01
4	2	16	7 × 7	28.633	0.057	99.98
5	2	32	2 × 2	10.777	0.021	98.10
6	2	32	3 × 3	21.497	0.043	99.76
7	2	32	5 × 5	55.801	0.111	99.93
8	2	32	7 × 7	107.257	0.214	99.99
9	3	16	2 × 2	7.561	0.015	83.80
10	3	16	3 × 3	15.481	0.031	97.01
11	3	16	5 × 5	40.825	0.081	99.01
12	3	16	7 × 7	78.841	0.157	99.67
13	3	32	2 × 2	27.225	0.054	97.18
14	3	32	3 × 3	58.425	0.116	99.31
15	3	32	5 × 5	158.265	0.316	99.81
16	3	32	7 × 7	308.025	0.616	99.99

Table 1: Adapted from Shahverdy et al. 2020

We propose a novel neural network that uses sensory inputs such as a six-axis IMU (accelerometer, gyroscope), and microphone data to classify driver behavior. We believe our project is technically feasible since other non-visual-based networks have been quite successful in identifying distraction from non-distraction and were able to run quite efficiently.

We believe that incorporating microphone data into a convolutional neural network for detecting distracted driving has several advantages. First, the use of a microphone allows for the collection of more diverse data, as it can pick up sounds and conversations within the vehicle that may not be captured by IMU sensors. This diversity of data can help the neural network more accurately classify driver behavior, as it has access to a wider range of information. Additionally, microphones are more robust in diverse conditions and do not require specific lighting or visibility conditions in order to function effectively. This makes them a reliable and efficient choice for detecting distracted driving in real-world environments while staying computationally efficient.

Our approach is to use similar techniques for developing an image similar to Shahverdy et al. 2020 by using a recurrence plot for the IMU sensors. The IMU data will be converted into an RGB image with the x, y, and z values corresponding to RGB. The optimal size for these pixels is yet to be determined

To improve the accuracy of our convolutional neural network in detecting distracted driving, we propose to preprocess the microphone data by dividing it into three distinct branches based on pitch. The first branch will consist of lower-pitch sounds, such as those corresponding to engine throttle and accelerations. The second branch will include sounds that fall within the range of music and speech. The third branch will consist of higher pitch sounds. By structuring the microphone data in this way, we hope to enable the network to more easily differentiate between distractions and other background noises.

This approach is different from Shahverdy et al. 2020, which relies on a connection to the vehicle at all times. We hypothesize our approach will work on any vehicle, regardless of the year it was manufactured. This means that the compact device would be ready to use right out of the box, without needing additional setup or configuration. Furthermore, the use of these sensory inputs allows for the detection of distracted or fatigued driving in any light condition, making it a reliable and effective solution for ensuring safe driving.

We plan to utilize various CNN architectures to optimize accuracy and computational efficiency for driver behavior detection. Instead of using a traditional deep fully connected network, we plan to adopt a newer CNN architecture that utilizes global max pooling at the end of the CNN and a small fully connected network. These newer CNN architectures have been shown to be highly efficient, have fewer parameters, and are less prone to overfitting in many applications as described by Warden 2019.

The image generation method we propose has structured patterns and high spatial properties, which means that only a few convolutional layers are needed for training and convergence, compared to traditional natural images. Our approach is heavily inspired by Shahverdy et al. 2020 study as shown in Image 3.

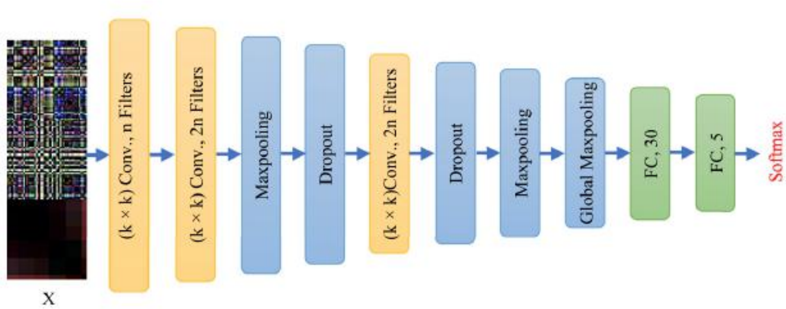

Image 3: Network Architecture adapted from Shahverdy et al. 2020

Going forward, here are the milestones for EyeDA:

1. **Secure a dataset for neural net**: To train a neural network for EyeDA, a large dataset of microphone, and IMU data would be needed. This dataset would need to include a wide range of examples of distracted, fatigued, and normal driving behaviors. Van Ly et al. (2013), Carmona et al. (2015), Carmona et al. (2017). We will scout the relevant literature for dataset that will have the Microhpone, IMU data and labellings.
2. **Create an image using the preprocessing algorithm**: The preprocessing algorithm will convert the data feed from the time series into an image. The pixel size for the images has not yet been determined. The recurrence plot technique will be used to convert the dataset into images for the IMU. We are not certain as to which plotting technique will provide the best differentiation in microphone sounds.
3. **Design neural network architecture**: Once a dataset has been determined, the next step would be to design a neural network architecture capable of effectively classifying the data into appropriate categories. This would likely involve testing a range of different architectures and hyperparameters to find the best-performing model.

4. **Establish network design, training, and evaluation**: After the network architecture has been designed, the next step would be to train the network using the dataset. Once the network has been trained, it would need to be evaluated to determine its performance.
5. **Conduct real-world testing and deployment**: If the network performs well during evaluation, the final step would be to deploy the network on Jetson Nano and test it in the real world.

The following addresses anticipated challenges and their associated solutions:

1. One challenge in developing our convolutional neural network is identifying the optimal method for preprocessing microphone input data to detect distraction. We will experiment with various techniques for dividing the microphone sounds into distinct sections to optimize the model. Our findings will be recorded to identify the most efficient solution.
2. Lack of accuracy in our neural network: we are not confident what accuracy to expect. We are hopeful for the dataset to be at least 95% accurate; however, we have kept 90% as a minimum threshold. By augmenting our training data and carefully designing the network architecture, we hope to achieve the necessary level of accuracy for EyeDA to be a reliable and effective tool for detecting drowsiness and distraction in drivers.
3. One of the challenges in utilizing non-facial features for driver behavior detection is the potential for reduced functionality during periods of cruise control. To address this issue, we propose implementing a feature that can distinguish

between instances of manual and autonomous driving, allowing for the appropriate level of monitoring in each situation. This can be achieved through the integration of an additional algorithm that monitors the drive mode.

Proposed timeline:

1. 2/7/23: Scout for potential datasets that have the required measurements and land on a dataset that we will use throughout the process. Document why the dataset was chosen. Guidance from the think team will be beneficial to this project.
2. 3/5/23: Create a preprocessing algorithm for IMU and microphone measurements to create grayscale images. Determine the optimal frequency and pixel size. Document findings.
3. 3/26/23: Develop and train a CNN to classify images into three categories: distracted, drowsy, and normal. Optimize architecture for accuracy and efficiency. Document findings.
4. 4/9/23: Compare the network with other industry standards and compress for smaller processors and faster frame rates. Document findings.
5. 4/30/23: Buffer time or fine-tune and explore next steps, such as adding PCBs, more sensors, making it open-source, etc.

Funding will go towards the purchase of a cheap and compact GPU for testing.

Item	Amount	Cost per unit:	Link:
Jetson Nano Developer Kit	2	$229	https://developer.nvidia.com/buy-jetson?product=jetson_nano&location=US

PERSONAL:

I got involved in developing Eyda by coming across the Shreya Dixit foundation created by Shreya's family, who lost Shreya in a car accident that involved distracted driving. Shreya was close to my age, and as a teen who has just begun to drive, I understand the driver's responsibility on the road. I also understand how quickly a small distraction can lead to catastrophic crashes and deaths. It is scary knowing that distracted driving is so prevalent, so I am inspired by Shreya's story to use AI to create EyeDa, a tool designed to specifically combat distracted driving. I have taken two AP Computer Science classes that have taught me the basics of programming, and I am proficient in JAVA.

SOURCES

National Highway Traffic Safety Administration. “Distracted Driving.” Retrieved from November 2019, www.nhtsa.gov/risky-driving/distracted-driving

Center for Disease Control and Prevention. (n.d.). Retrieved from https://www.cdc.gov/motorvehiclesafety/distracted_driving/

National Safety Council. “Distracted Driving: Research & Data.” December 2022, www.nsc.org/road-safety/safety-topics/distracted-driving

"Fatigue-Related Crashes." National Highway Traffic Safety Administration, December 2022.
https://www.nsc.org/road/safety-topics/fatigued-driver#

Warden, P. (2019, December 16). TinyML: Machine learning with tensorflow lite on Arduino and Ultra-Low-Power Microcontrollers. Google Books. Retrieved December 21, 2022, from https://books.google.com/books/about/TinyML.html?id=tn3EDwAAQBAJ

Cedeno, Lorenzo. "On the Road to Safety: Improving Driver Fatigue Detection using Photoplethysmography." Mechatronics Rig Weblog, 4 December 2022. https://www.researchgate.net/publication/277788122_Drowsiness_Detection_Using_Photoplethysmography_Signal

Galarza, G., Ruiz, F., Casas, J., & Armingol, J. M. (2018). A monitoring system to alert the driver about drowsiness driving. IEEE Acces, 6, 23 964-23 975.

Yan, C., Teng, X., Smith, J., & Zhang, Y. (2016). Detecting driver distraction using convolutional neural networks. In 2016 IEEE International Conference on Multisensor Fusion and Integration for Intelligent Systems (MFI) (pp. 556-561). IEEE.

Lee, J., Yoon, J., Song, W., & Park, J. (2018). Detecting aggressive driving using convolutional neural networks. In 2018 IEEE Intelligent Transportation Systems Conference (ITSC) (pp. 1-6). IEEE.

Abouelnaga, Y., Eraqi, H. M., & Moustafa, M. N. (2017). Real-time distracted driver posture classification. CoRR, abs/1706.09498.

Das, N., Ohn-Bar, E., & Trivedi, M. M. (2015, September). On the performance evaluation of driver hand detection algorithms: Challenges, dataset, and metrics. In 2015 IEEE 18th International Conference on Intelligent Transportation Systems (pp. 2953-2958).

Hari, C. V., & Sankaran, P. (2021). Driver distraction analysis using face pose cues. Expert Systems with Applications, 179, 115036. https://doi.org/10.1016/j.eswa.2021.115036

Shahverdy, M., Mahmoudi, F., Berangi, R., & Sabokrou, M. (2020). Driver behavior detection and classification using deep convolutional neural networks. Expert Systems with Applications, 149, 113240. https://doi.org/10.1016/j.eswa.2020.113240

Carmona, J., García, F., Martín, D., Escalera, A., & Armingol, J. (2015). Data Fusion for Driver Behaviour Analysis. Sensors, 15(10), 25968–25991. https://doi.org/10.3390/s151025968