Creating Sensor System for Safe Motor Navigation: HOMESWEET

Lare Samuel Adeola[1], **Celestine Iwendi**[2], **Aamir Mazhar Abbas**[3]

[1]University of Bolton, Department of Creative Technologies, Bolton, England

[2]University of Bolton, Department of Creative Technologies, Bolton, England

[3]University of Bolton, Department of Creative Technologies, Bolton, England

l.adeola@bolton.ac.uk[1], c.iwendi@bolton.ac.uk[2], a.abbas@bolton.ac.uk[3]

Abstract: 'HOMESWEET' is a deep learning model created with the YOLO object detection algorithm that has been trained to detect certain human physical states that could result in road accidents and deaths. The model achieved an accuracy score of 83% and a Precision rate of over 90% but had a fairly modest Recall rate of just over 70%. The model can be deployed in various other applications as it was able to detect not only facial cues but other micro-expressions and gesticulations that lead to the various states; in particular, in this research, it was created to detect fatigue, drowsiness and lack of total concentration while driving.

Keywords: Computer Vision, Deep Learning, Sensor System, Ubiquitous Computing, YOLOV5

1. Introduction

Unsafe driving is one of the leading causes of death not only in the United Kingdom but also in the world. According to the data from [1], an average of 2400 people die each year from year 2000 to 2020 as a result of road accidents. 34% of these road accidents are caused by motorists who have failed to look properly [2]. Figures 1, 2 and 3, show various plots of the number of lives lost, those who sustained serious injuries and the total injuries as a result of road accidents.

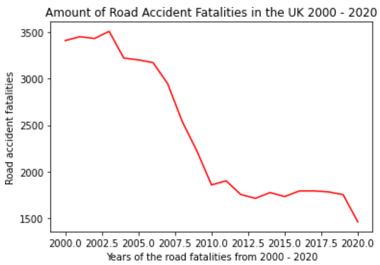

Figure 1: Amount of Road Accident Fatalities in the UK 2000 - 2020

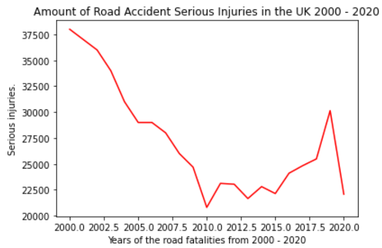

Figure 2: Amount of Road Accident Serious Injuries in the UK 2000 - 2020

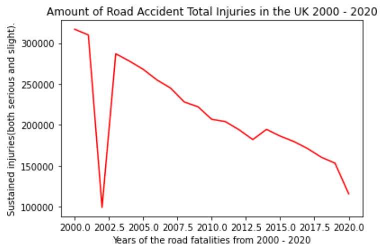

Figure 3: Amount of Road Accident Total Injuries in the UK 2000 - 2020

The charts show a markedly downward trend in the amount of road accidents fatalities from the year 2000 to the year 2020. While this is commendable, it is not enough. For example, in the year 2020, there were a total of 1460 deaths caused by road accidents. This means that an average of 4 people died from road accidents per day in the year 2020. This coupled with the fact that road accidents can be avoided, it is imperative to always strive to reduce these figures further.

This makes the development of a sensor application that can help curtail unsafe driving habits paramount and important. As this application can not only save the lives of drivers and motorists but also save the lives of members of the public like pedestrians and cyclists. In this research, we present a sensor application system that has achieved an accuracy score of 83% in predicting selected unsafe driving habits (like fatigue, drunkenness, distractions) and other causes that may lead to both road accident fatalities and serious injuries.

Aims and Objectives

Build a Computer Vision Model to detect driver fatigue.

Create a Sensor and Feedback system.

2. Micro-expressions - A Brief Detour

Micro-expressions are those semi-voluntary, often not lasting too long, expressions that our body makes frequently in different behavioural states [3]. For example, whenever we meet someone we genuinely like, there's a web that pulls our eyes and our pupils expand [4]. There are also micro-expressions in other states like fatigue, drunkenness etc. A good example of a micro expression for fatigue can be the yawn. Whenever we yawn, more often than not it is a sign of either current fatigue or future fatigue; with the way we fold our hands (mostly one's dominant hand) to cover our mouths while yawning, this was a common micro-expression noticed in a lot of images sourced from our secondary data sources and the model was fed with this information. Figure 12 shows the various images which had this micro-expression and how the model performed in predicting their label. Other micro-expressions like the closing of the eyes(drowsy/sleepy) or the alertness level (wider opened eyes or a smile or laugh) or common gesticulations (like the stretch when one is tired) were also used to train the model.

Figure 4: Showing the various fatigue/drowsiness micro-expressions. You can see in the image the prevalence of a yawn and the body movements that trail such a fatigue state.

3. Methodology

3.1 Primary Data

The training of 'HOMESWEET' involved both primary and secondary sources of data. The primary source of data was of the author and it was taken through the webcam of the author's laptop. A series of different images and angles were taken to make the model consistent in accurately predicting the various sub-classes. These positions include - i. At different focal lengths. This was achieved albeit mechanically by moving further away. ii. Blurry/grainy images. This was achieved by being in slight motion while the camera was capturing the author's image iii. Side views. iv.Tilts. uncontrolled Environment. All the images from the primary data were taken in the Peter Stocker Laboratory (sci 2.27) in the Department of Computing Sciences, University of East Anglia.

3.2 Secondary Data and Other Image Selection Criteria

The secondary data was sourced from the web. They are - Google Images, Pixabay, and istockphoto. There were three cardinal attributes/features sought after when selecting an image – i. Posture - Does the image narrate honestly the class to which it would belong? ii. Clarity and Crisp - How clear and crisp is the image? What is the pixel density? iii. Image Shot - Is it a full view or a side view? Is it a close-up shot or a mid-range shot? iv. Nationality. Algorithm Bias can hinder the general effectiveness and adaptability of a model in the wild. Hence, various images of different nationalities were selected. Three nationalities in total were represented - African/Black/Black British/Caribbean, Asian/Asian British/Arab, and White/Caucasian. v. Gender - In order to further reduce bias, images of both women and men were selected and sourced. vi. Lighting - Is the image well-lit or under-lit or mid-lit? Efforts were made to select images that tick one of the previous options.

Figure 10 shows the spread of the different criteria used in selecting, sourcing and sorting the images for the secondary data while figure 11 shows some of the images used to train the model.

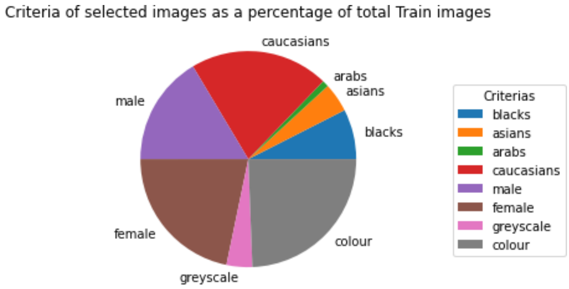

Figure 5: Criteria of selected images as a percentage of total Train images. The total amount of train images was 167.

Figure 6: A selection of different images used in training the model. The criteria from left to right: nationality - black, asian, arabian, caucasian and colour - greyscale

3.3 Data Processing/Interpretation

Most of the Data Processing done for this project involved careful selection and sorting of images that would be able to train the model adequately to be able to classify images both captured and live into two categories - safe and unsafe.

In the process of selecting these images, various criteria were itemized and they have been elaborated in the previous chapter. However, there was one more step that involved processing both the primary and secondary data before feeding them to the model.

3.4 Image Annotation

Image Annotation can be roughly defined as the process of labelling images present in a dataset for the purpose of training a machine learning model. It is an important stage in the Computer Vision Machine Learning process and one important reason for this step is mainly to simplify access to these selected images, using them as the ground-truth for further classification and object detection tasks [5]. Labelling or annotating a picture is paramount in Computer Vision tasks and it brings to the fore the perceived difficulties present in comprehending a visual scene [6]. For this project, the open-source tool LabelImg was used to annotate the train images into the two classes - safe and unsafe.

4. Tools, Equipments and Libraries

Below is a list of the tools, equipments and libraries in the Data Processing, Analysis and Interpretation stage include: Python - Programming Language, Jupyter Notebooks - Integrated Development Environment, Open CV - This is an open source Computer Vision library, Numpy - A popular library used for manipulating arrays, YOLO - A Computer Vision object detection algorithm framework, LabelImg - an open source tool used for Image Annotation, Pytorch - A python deep learning framework, Matplotlib - A python visualization library. This was used to view images, capture live data and view the model predictions, Webcam - The primary sensor used to capture live and primary data.

5. Feedback System

A Visual feedback system was created in the application as a signal to the user based on the user's label category.

Visual Feedback

Figure 7: the ’sweet blue’ (rgb: 49, 84, 116) used for the safe class to signal safe driving condition to the user

Figure 8: the ’sour red’ (rgb: 135,25,25) used for the ’not safe’ label to signal not safe driving condition to the user - prior or during

6. Results

6.1 HOMESWEET IN THE WILD: Test Data Analysis and Results

A total of 66 images (39% of the total data) were used to test the generalized ability and predictability of the model. Of this number, the model found no detections in 5. Figure 30 below shows a table of the various Classification metrics of the model.

6.2 Interpretation of the Test Data Results

Figure 30 shows the classification performance of the model on the selected test data. We can see that the model has a Recall rate of 0.71875, a Precision of 0.95833 and an Accuracy score of 0.82759. This means that the model has a high specificity rate and a reasonably good sensitivity rate. The model’s accuracy score is also good at 0.82759 and it must be noted that this accuracy can be improved over time as the model is trained with more relevant pictures.

The size of the train images would also have an impact on the strength and generalized nature of the model. This model was trained with only 167 images, a very small number in the Computer Vision domain. A good place to start with training a model would be north of 100,000 - 1,000,000 images and a test data that would be between 15 - 20% of the train data. For this project however, the concern was on taking little appreciable steps and further down the line, the model would be fed with a larger amount of varied and relevant data.

Despite this shortcoming of train data size, the model still performed superbly in the wild which means it is ready for deployment. Section 4.9, shows the predictive ability of the model on randomly selected test data from the two classes - safe and not safe.

7. Conclusion

With an accuracy score of 0.82759, the ’HOMESWEET’ model has proven to be an effective deep learning model able to predict two class labels - safe and not safe - driving habits. The model has a high precision rate but its recall rate is not as high, this means that there would be times in the real-world scenarios where it may incorrectly predict some class labels. While in other innocuous settings and applications, a recall of 0.71875 is good enough, in this application that aims to reduce road accidents and consequentially deaths, this may not be good enough.

This can be solved by feeding the model more data and continually providing it with information under various criteria like lighting conditions, accessories on the face, more micro expressions, more nationalities, backgrounds with artificial lighting etc to further improve ’HOMESWEET’ accuracy score and recall rate and make it a more generalized deep learning model.

7.1 Links

UI/UX and Feedback signals - https://codepen.io/laresamdeola/pen/BaYwmoz
Deep Learning Model Code - https://github.com/laresamdeola/homesweet/blob/main/HOMESWEET.ipynb

Model’s Test Data Predictions

Figure 9: Some selected test data showing the model’s performance at classifying the images into the not safe label

Figure 10: Some selected test data showing the model’s performance at classifying the images into the safe label

References:

[1] ons.gov.uk, “Road casualties,” 2019.

[2] statista.com, “Distribution of contributing factors leading to road accidents in great britain in 2019,” 2019.

[3] G. Zhao, X. Li, Y. Li, and M. Pietikäinen, “Facial micro-expressions: An overview,” Proceedings of the IEEE, 2023.

[4] R. Greene, The laws of human nature. Penguin, 2018.

[5] A. Hanbury, “A survey of methods for image annotation,” Journal of Visual Languages & Computing, vol. 19, no. 5, pp. 617–627, 2008.

[6] A. Barriuso and A. Torralba, “Notes on image annotation,” arXiv preprint arXiv:1210.3448, 2012.