**Group 6 Project Final Report**

**Project Title**: Driver drowsiness detection using deep learning

**Group Members**: Tianhan Jiang, Peiyun Zhao, David Laditan, David Guo, Tobi Lawal

**Contact Email**: [tianhan.jiang@ucalgary.ca](mailto:tianhan.jiang@ucalgary.ca), [peizhao@ucalgary.ca](mailto:peizhao@ucalgary.ca),

[oluwapelumi.laditan@ucalgary.ca](mailto:oluwapelumi.laditan@ucalgary.ca), [yuhua.guo@ucalgary.ca](mailto:yuhua.guo@ucalgary.ca) and [tobi.lawal1@ucalgary.ca](mailto:tobi.lawal1@ucalgary.ca)

1. **Introduction**

Drowsiness is identified as one of the major causes of fatal traffic accidents. Unfortunately, about 20% of drivers tend to show drowsiness while driving, reported by National Safety Council[1]. This project aims to combine a fine-tuned neural network and python-based face detection and feature extraction module into a real-time drowsiness detection system that will improve road safety.

Current state-of-the-art facial expression recognition models are able to achieve an accuracy of around 75-80%, utilizing the VGG-16 model [11]. Considering that drowsiness detection is arguably easier to differentiate, we would consider an acceptable model performing with 70% validation accuracy and accuracy greater than 75% being widely successful.

The primary motivation for choosing this topic is to apply deep learning concepts and techniques we learned to a real-life problem with practical use.

1. **Methodology**
   1. Dataset and data preprocessing

In the last stage, we have used a [Kaggle dataset](https://www.kaggle.com/serenaraju/yawn-eye-dataset-new) containing 2900 images with four labels: closed, open, no\_yawn, and yawn. Those four labels represent human face images with eyes closed, eyes open, without yawn, and with yawn, respectively. The four classes in the dataset are balanced.

As described in the midterm report, we have developed several models with the best validation accuracy of 0.7098. As initially planned, we worked on finding a new dataset with more data to resolve the overfitting issue presented in various models with deep architectures.

In this stage, we have found another dataset containing 4000 images covering two classes, eyes closed and eyes open. However, we are not able to find the same quantity of images with classes yawn and non-yawn. Due to limited time constraints, we decided not to make or search for 4000 yawn and non-yawn human face images. However, as stated in section 3, we achieve good accuracy in detecting drowsiness based on the binary classification model we built with the new two classes dataset.

We have adopted several data preprocessing techniques, including data augmentations, min-max scaling. The current setup is the best combination that will most positively affect the model accuracy.

* 1. Neural network architecture design

In the project proposal, we mentioned that we had reviewed some similar works[2][3] reporting good accuracy on ResNet, VGG-FaceNet[7], InceptionV3, AlexNet[6], FlowImageNet[8]. AlexNet is fine-tuned to learn features related to drowsiness. The VGG-FaceNet is trained to learn facial features related to drowsiness. It is robust to genders, ethnicity, hairstyle and various accessories adornment. FlowImageNet takes a dense optical flow image extracted from consecutive image sequences and is trained to learn behaviour features related to drowsiness, such as facial and head movements. The plan was to train multiple networks separately and ensemble good performing networks to cover all necessary features essential to detect drowsiness[4].

In the last stage, we have implemented AlexNet and VGG-16 models. The result showed possible overfitting. To have a good performing baseline model, we decided to build a less complex CNN architecture. This model has a total of 28,600 trainable parameters and is in the process of further development. In this model, we use “categorical\_crossentropy” to define the loss function and use “accuracy” as the error metric. After 96 epoch, both accuracy and loss tend to flatten. Metrics at epoch 96: test accuracy is 0.7875, test loss is 0.4176, training accuracy is 0.7931, training loss is 0.3992, validation accuracy is 0.7098, validation loss is 0.4812.

Since our final goal is to have a light yet powerful model that can be run on a mobile device in order to make real-time drowsiness detection on a driver, we shifted our focus to fine-tuning a pre-trained MobileNet. Currently, this improved MobileNet architecture is showing good results on accuracy and acceptable performance on training time, as described in section 3 of this report.

The architecture is summarized in Table 1.

* 1. Face detection and feature extraction module

As pointed out by previous works[5], eye-based methods and mouth-based methods are the two main categories of drowsiness detection methods. The original intent of this work is to train a network that classifies drowsiness based on both eye region and mouth region. However, we have focused on only eye region classification in the model and eye region recognition in the real-time detection module because of time constraints and dataset limitations.

This python module uses a webcam to capture user face images in stream. After system initialization, four OpenCV algorithms are used to detect faces, eyes, and eyes with glasses. A partial image containing only the eye region is extracted and processed with the same pre-processing steps as the training data. In this work, images are resized to (224, 224), and min-max scaled using (1./225). The processed image is then fed to the trained model. A numerical value of range (0, 1) is to be returned, with 0 representing closed eyes and 1 representing open eyes.

We have designed a set of unit test test cases for this real-time detection module. The test case design covers robustness testing and worst-case testing. Detailed testing results are described in section 3.

1. **Preliminary results**

The improved MobileNet we have built performs well with a generally good training accuracy, validation accuracy, and training time. During training, we have used much fewer epochs (i.e. 7 epochs) than the baseline model, since more epochs tend to overfit the model easily based on our training experiments. At epoch 7, training loss is 6.4966e-04, training accuracy is 1.0000, validation loss is 0.0671, and validation accuracy is 0.9900. A detailed training history is shown in Figure 1.

In regards to real-time testing, we do not have a numerical value for accuracy since we do not have a proper way to analyze each frame of webcam images and label and compare prediction results with ground-truth label value. However, if latency is ignored, the model prediction accuracy is approximating 100%.

In this stage, latency ranges from about 0.5 to 1 second. Since this part of the project is not closely related to the core content of this course, we decided to put devote too much energy into optimizing latency.

In the project files, we have also attached a short video with one team member testing the real-time detection module. This test video has shown a typical reaction of our detection system to closed eye state and open eye state.

1. **Future works**

We intend to keep working on this project after the final presentation and final reports are delivered. We feel the potential of this work to become a better performing tool with some potential to be commercialized.

In order to achieve this goal, the following additional works have to be done:

1. Add more implementations of python/OpenCV modules in order to achieve more facial feature recognition, such as motion (eye movement, face direction), mouth (yawn, talking, smoking), mood.
2. Find or make more human face images to enrich the dataset. The goal is to have a multi-class training dataset of which the class can cover all the above-mentioned facial features in separate classes, such as eyes\_not\_on\_road, not\_facing\_front, yawn, talking, phone\_call, smoking, happy, angry.
3. If the improved MobileNet performs not well, new architectures are to be designed and implemented.

**Figures and Tables**

**Table 1. Improved MobileNet network summary**

**Table

Description automatically generated**

**Table

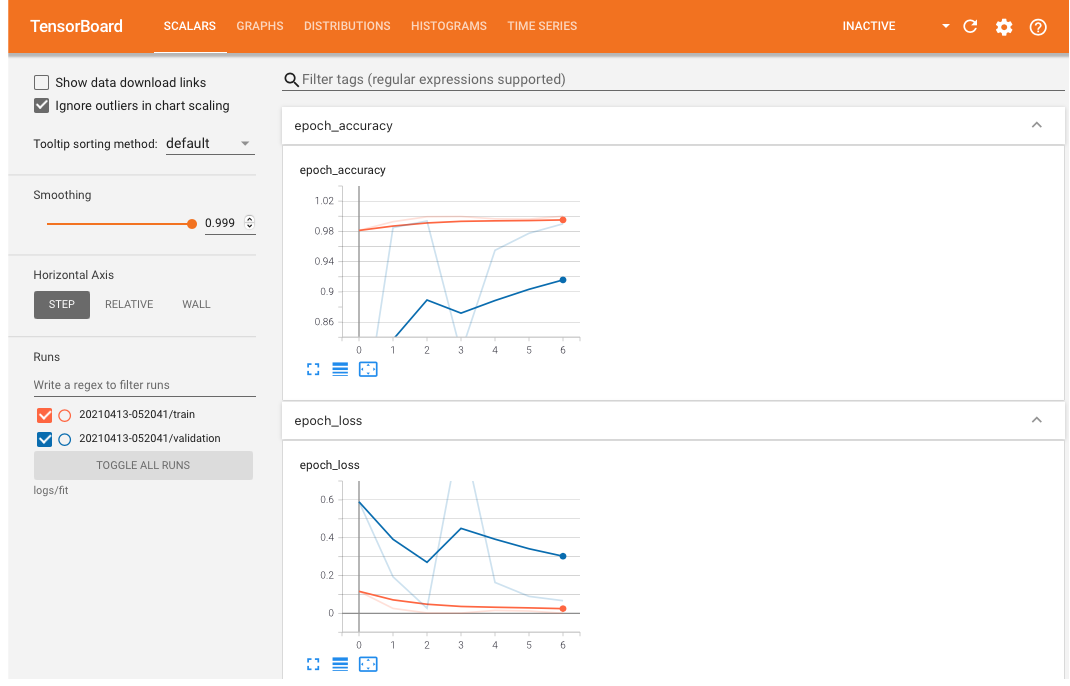
Description automatically generated**

**Table

Description automatically generated**

**Table

Description automatically generated**

****

**Figure 1. Tensor Board training history**

**References**

[1] Drivers are falling asleep behind the wheel, National Safety Council. <https://www.nsc.org/road/safety-topics/fatigued-driver>

[2] Vijayan, Vineetha, and Sherly, Elizabeth. "Real Time Detection System of Driver Drowsiness Based on Representation Learning Using Deep Neural Networks." Journal of Intelligent & Fuzzy Systems 36.3 (2019): 1977-985. Web.

[3] Park, Sanghyuk, Pan, Fei, Kang, Sunghun, and Yoo, Chang D. "Driver Drowsiness Detection System Based on Feature Representation Learning Using Various Deep Networks." Computer Vision – ACCV 2016 Workshops 10118 (2017): 154-64. Web.

[4] Dua, Mohit, Shakshi, Singla, Ritu, Raj, Saumya, and Jangra, Arti. "Deep CNN Models-based Ensemble Approach to Driver Drowsiness Detection." Neural Computing & Applications (2020): Neural Computing & Applications, 2020-07-20. Web.

[5] Zhao, Lei, Wang, Zengcai, Zhang, Guoxin, and Gao, Huanbing. "Driver Drowsiness Recognition via Transferred Deep 3D Convolutional Network and State Probability Vector." Multimedia Tools and Applications 79.35-36 (2020): 26683-6701. Web.

[6] Krizhevsky, A., Sutskever, I., Hinton, G.E.: Imagenet classification with deep convolutional neural networks. In: NIPS, pp. 1097–1105 (2012)

[7] Parkhi, O.M., Vedaldi, A., Zisserman, A.: Deep face recognition. In: BMVC, vol. 1, p. 6 (2015)

[8] Donahue, J., Anne Hendricks, L., Guadarrama, S., Rohrbach, M., Venugopalan, S., Saenko, K., Darrell, T.: Long-term recurrent convolutional networks for visual recognition and description. In: CVPR, pp. 2625–2634 (2015)

[9] Weng, Ching-Hua, Lai, Ying-Hsiu, and Lai, Shang-Hong. "Driver Drowsiness Detection via a Hierarchical Temporal Deep Belief Network." Computer Vision – ACCV 2016 Workshops 10118 (2017): 117-33. Web.

[10] Bhargava Reddy, Ye-Hoon Kim, Sojung Yun, Chanwon Seo, Junik Jang.

“Real-time Driver Drowsiness Detection for Embedded System Using Model Compression of Deep Neural Networks” Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops, 2017, pp. 121-128

[11] T. Vo, G. Lee, H. Yang and S. Kim, "Pyramid With Super Resolution for In-the-Wild Facial Expression Recognition," in IEEE Access, vol. 8, pp. 131988-132001, 2020, doi: 10.1109/ACCESS.2020.3010018.

**Member Contributions**

Each member had a different task and completed various sections of this work, and the workloads are distributed equally.