Eye Tracking System to Detect Driver Drowsiness using Deep Learning

Final Year Project Report

4th Year Student Name

Nguyen Viet Tung
Hoang Manh

Supervisor: MSc. Luong Trung Kien

Bachelor of Computer Science
Hoa Lac Campus – FPT University
Spring 2021

Contents

Contents

Introduction: Problem

- According to statistics in Vietnam, more than 6 400 people have traffic accidents due to drowsiness every year.

Introduction: Problem

- According to statistics in Vietnam, more than 6 400 people have traffic accidents due to drowsiness every year.

- According to the National Highway Traffic Safety Administration estimated that in 2017 has 91,000 police-reported crashes involved drowsy drivers, and it not decreasing.

Introduction: Problem

- According to statistics in Vietnam, more than 6 400 people have traffic accidents due to drowsiness every year.

- According to the National Highway Traffic Safety Administration estimated that in 2017 has 91,000 police-reported crashes involved drowsy drivers, and it not decreasing.

➡ Driving drowsiness is still a problem that needs to be minimized

Introduction: Scope of Thesis

- In this project, we use deep learning to build a system to track the driver through the eyes to be able to alert in case the driver is showing signs of sleep.

Contents

Background: Approaches for Drowsiness Detection

- Measuring the rotation of the steering wheel or the distance from the lane or lateral direction.

Background: Approaches for Drowsiness Detection

- Measuring the rotation of the steering wheel or the distance from the lane or lateral direction.

- Techniques incorporates data from physiological sensors such as EEG, ECG and EOG data.

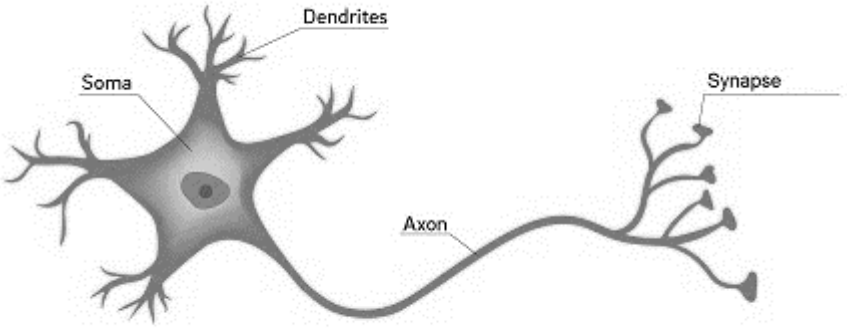
Background: Approaches for Drowsiness Detection

- Measuring the rotation of the steering wheel or the distance from the lane or lateral direction.

- Techniques incorporates data from physiological sensors such as EEG, ECG and EOG data.

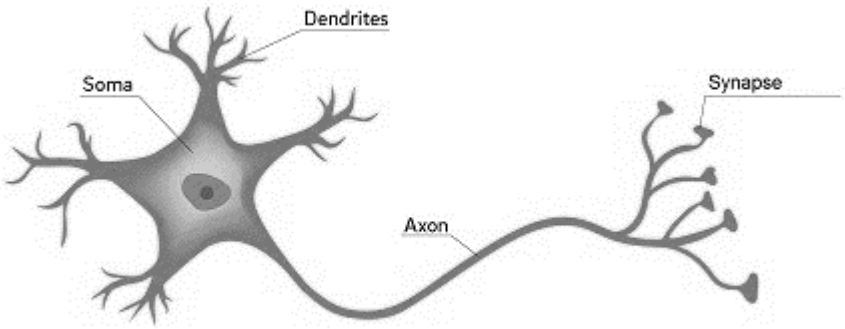
- The retrieval of facial features using Computer Vision, where patterns such as eye closing, shifting of the head, gaze or facial expression.

Background: Neural Network



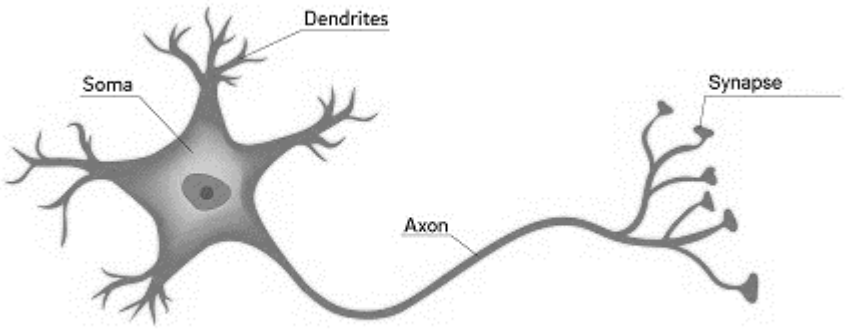
- How regular neural networks work ?
  - Our brain comprises around 10 million neurons, and each neuron connects to 10,000 other neurons.

Background: Neural Network



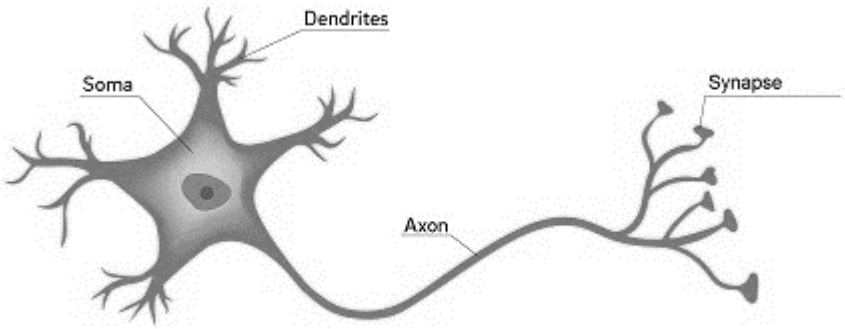
- How regular neural networks work ?
  - Our brain comprises around 10 million neurons, and each neuron connects to 10,000 other neurons.
  - Each neuron has a somatic core body, dendrites input, and axon output signals attached to other neurons.

Background: Neural Network



- How regular neural networks work ?
  - Our brain comprises around 10 million neurons, and each neuron connects to 10,000 other neurons.
  - Each neuron has a somatic core body, dendrites input, and axon output signals attached to other neurons.
  - The dendrites obtain the input data, and the output data are sent to other dendrites.

Background: Neural Network



- How regular neural networks work ?
  - Our brain comprises around 10 million neurons, and each neuron connects to 10,000 other neurons.
  - Each neuron has a somatic core body, dendrites input, and axon output signals attached to other neurons.
  - The dendrites obtain the input data, and the output data are sent to other dendrites.
  - The signal passes through the axon through the dendrites of other neurons if the electrical pulses are adequate to transform the nucleus into a neuron.

Background: Neural Network

- Neural Network (NN) is a model of programming that simulates the functioning of human neural networks.

Background: Neural Network

- Neural Network (NN) is a model of programming that simulates the functioning of human neural networks.

- In combination with deep learning (DL), neural networks are a versatile method that can better achieve numerous complex problems

Background: Neural Network

- In several fields of science and technology, artificial intelligence is today becoming a common subject :

Background: Neural Network

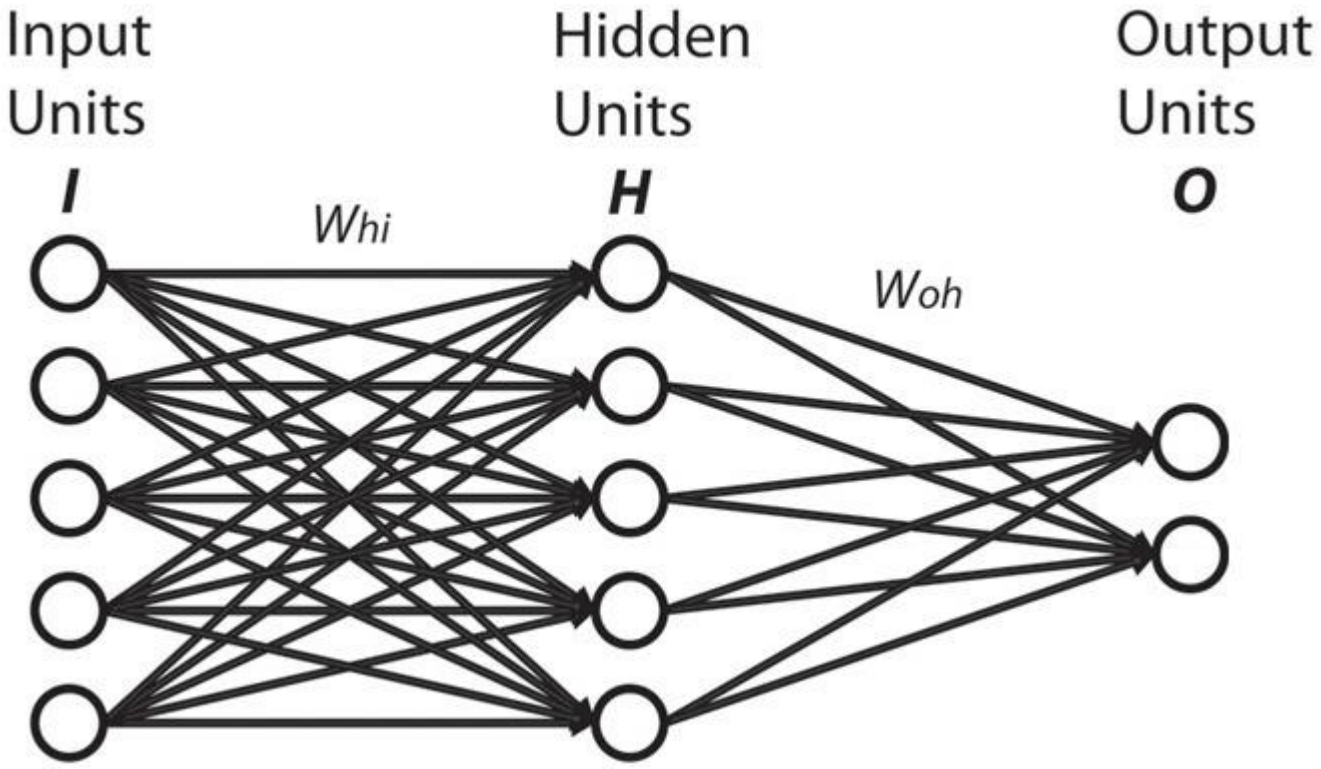
- In several fields of science and technology, artificial intelligence is today becoming a common subject :

Background: Neural Network

- In several fields of science and technology, artificial intelligence is today becoming a common subject :

Background: Neural Network

- Neural network components:
  - Units

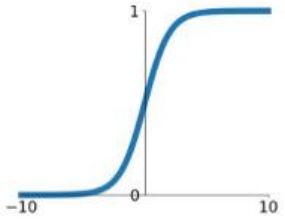


Background: Neural Network

- Neural network components:
  - Activation Function

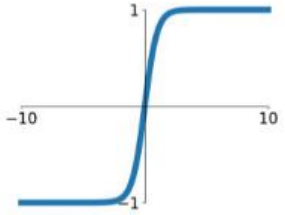
Sigmoid

$$\sigma(x) = \frac{1}{1+e^{-x}}$$



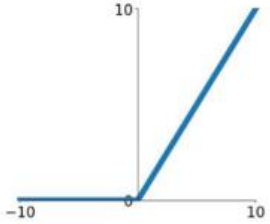
tanh

$$\tanh(x)$$



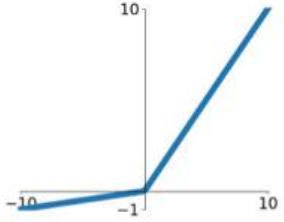
ReLU

$$\max(0, x)$$



Leaky ReLU

$$\max(0.1x, x)$$

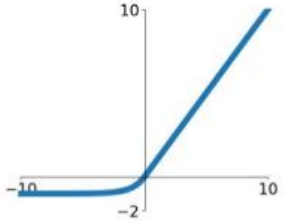


Maxout

$$\max(w_1^T x + b_1, w_2^T x + b_2)$$

ELU

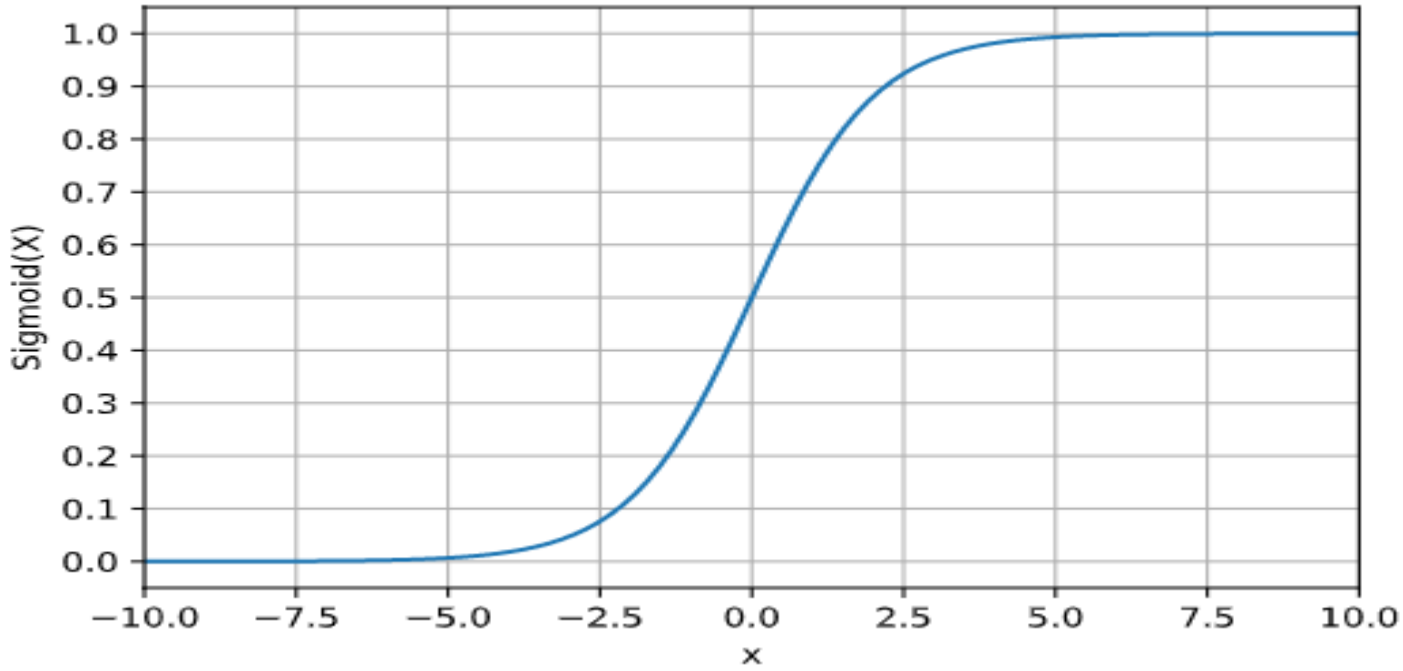
$$\begin{cases} x & x \geq 0 \\ \alpha(e^x - 1) & x < 0 \end{cases}$$



Background: Neural Network

- Neural network components:
  - Sigmoid

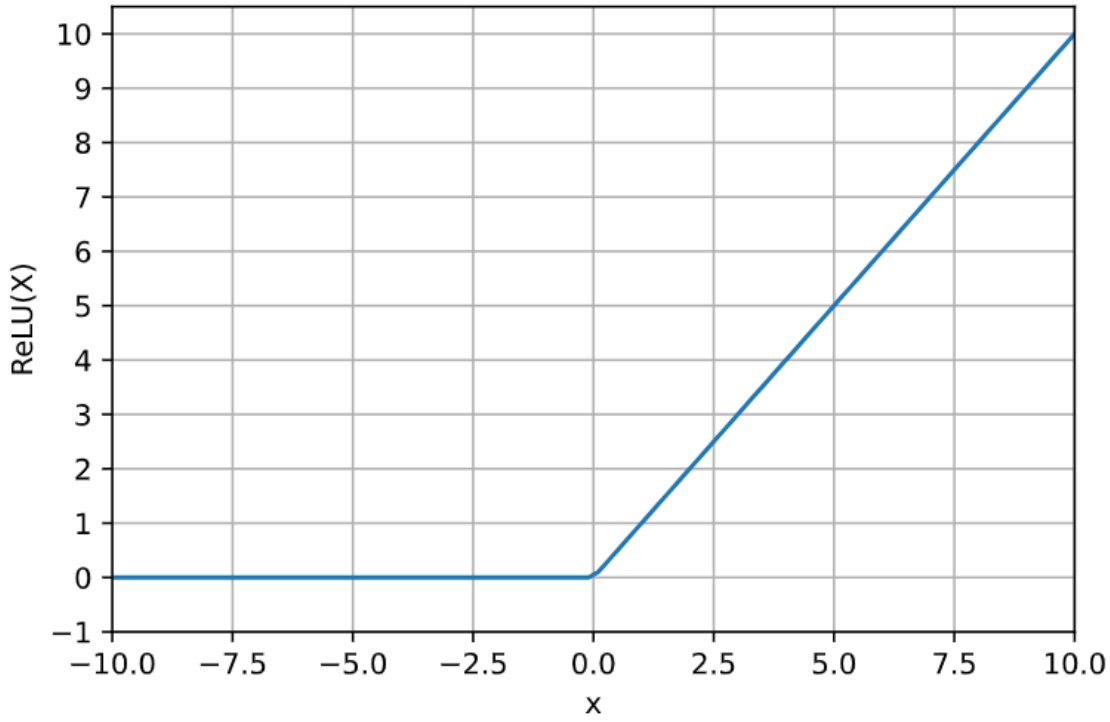
- $\sigma(x) = \frac{1}{1+e^{=x}}$



Background: Neural Network

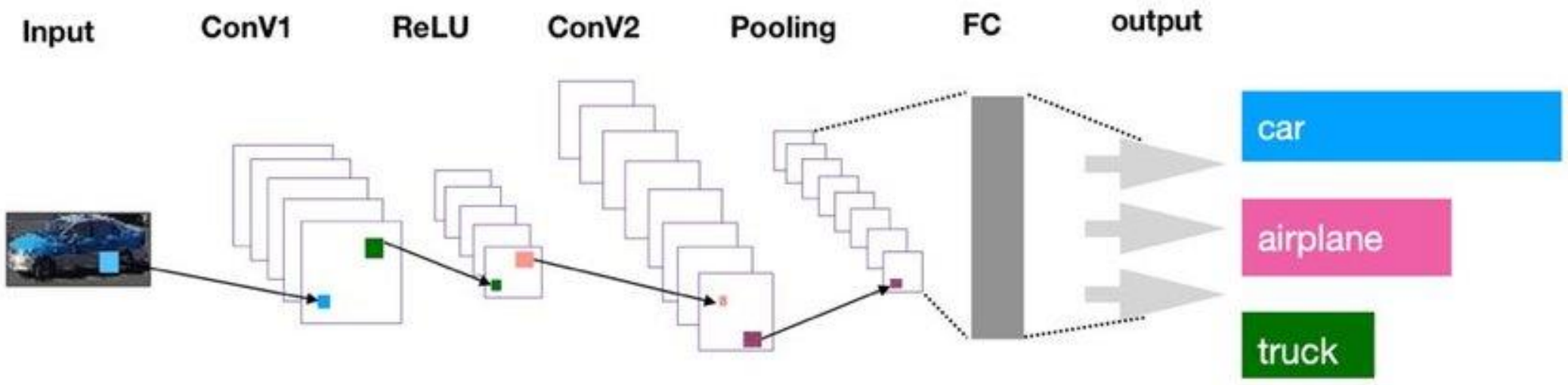
- Neural network components:
  - ReLU

  - F(x) = max(0,x)



Background: Convolutional Neural Network

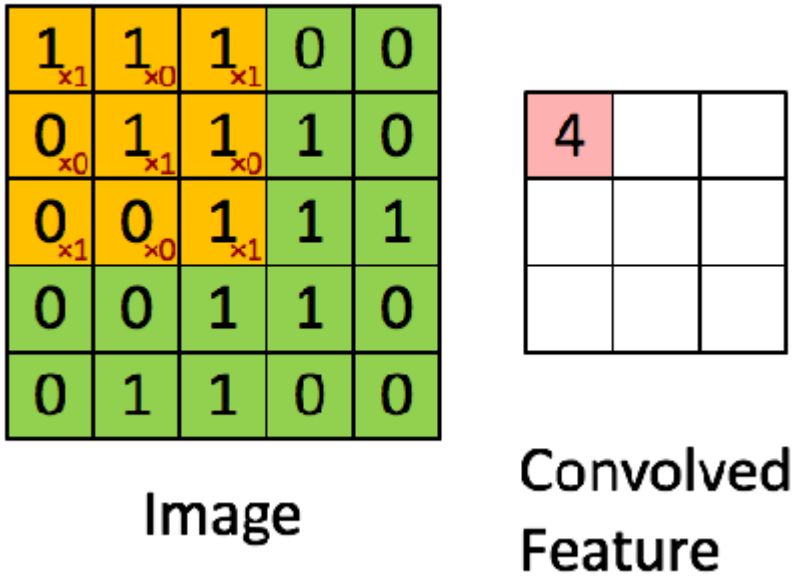
- Architecture



Background: Convolutional Neural Network

- Convolutional layer

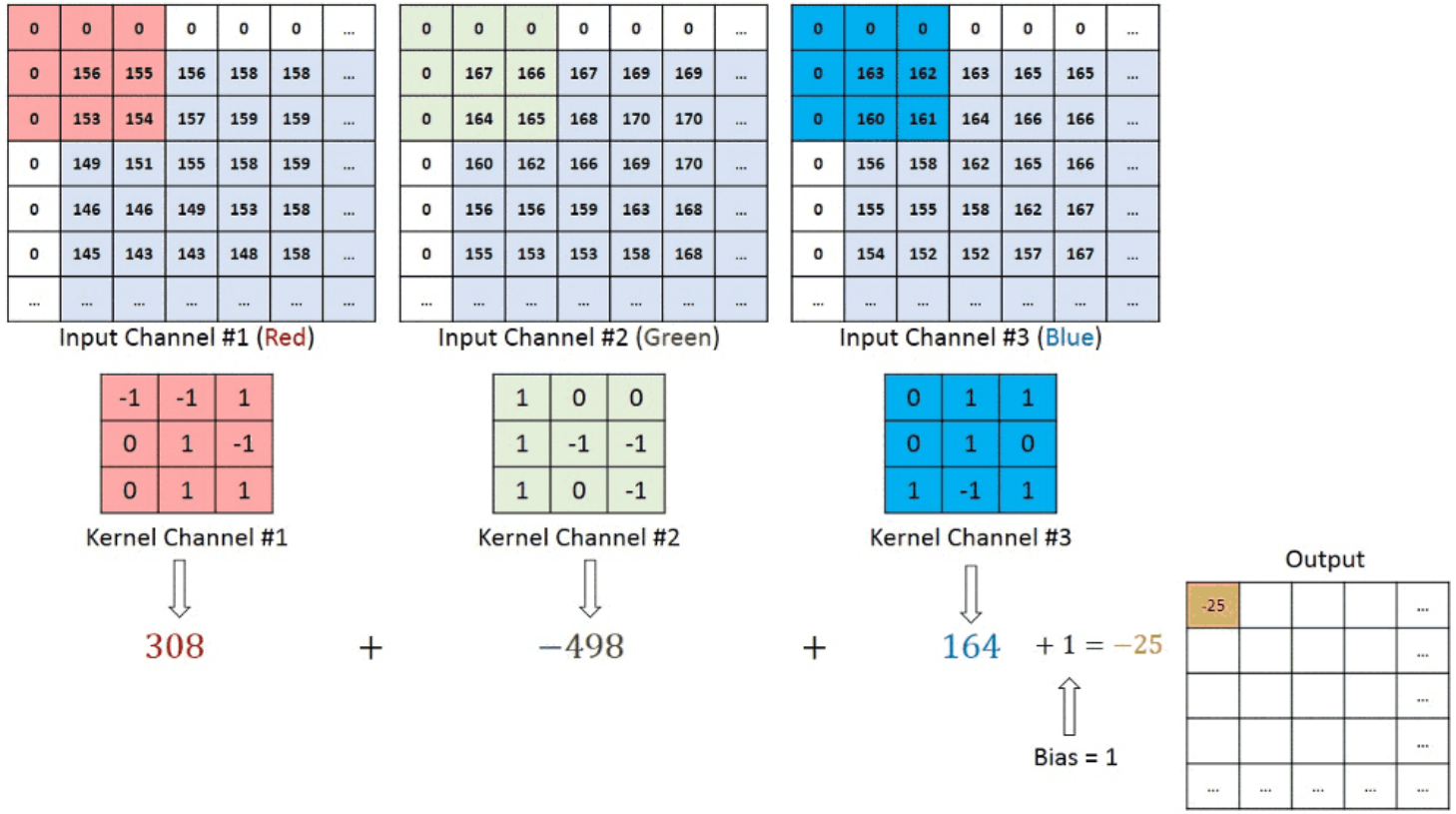
- Gray image:



Background: Convolutional Neural Network

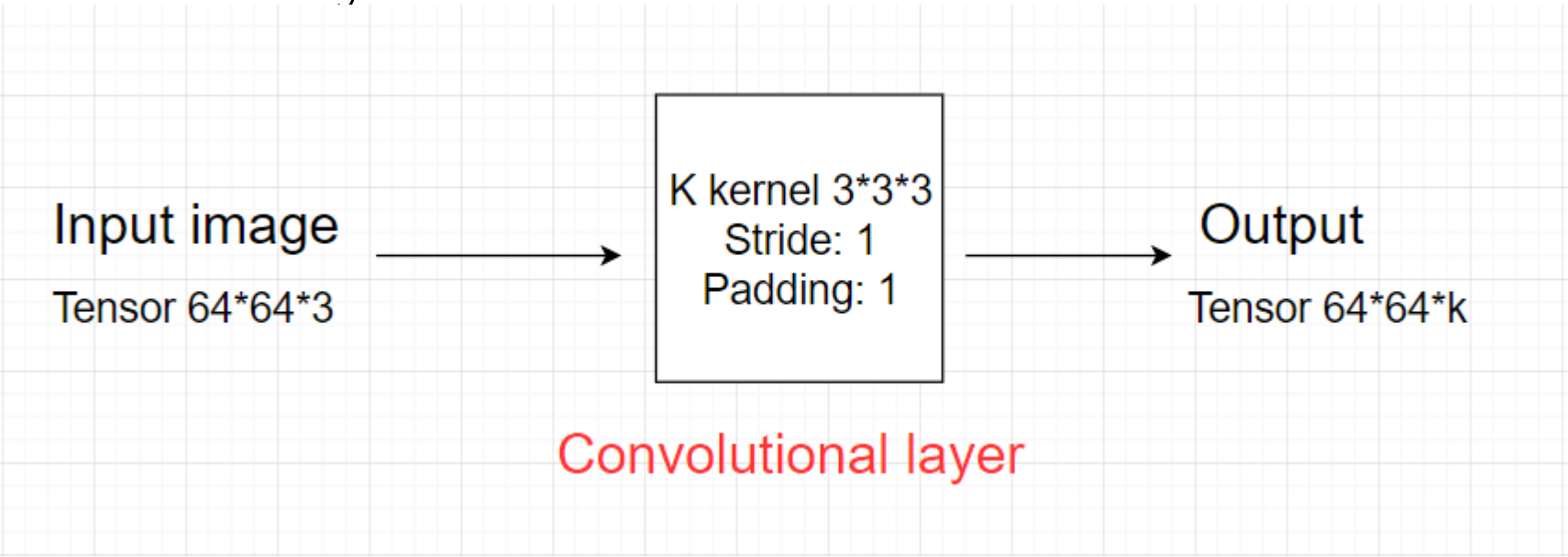
- Convolutional layer

- RGB image:



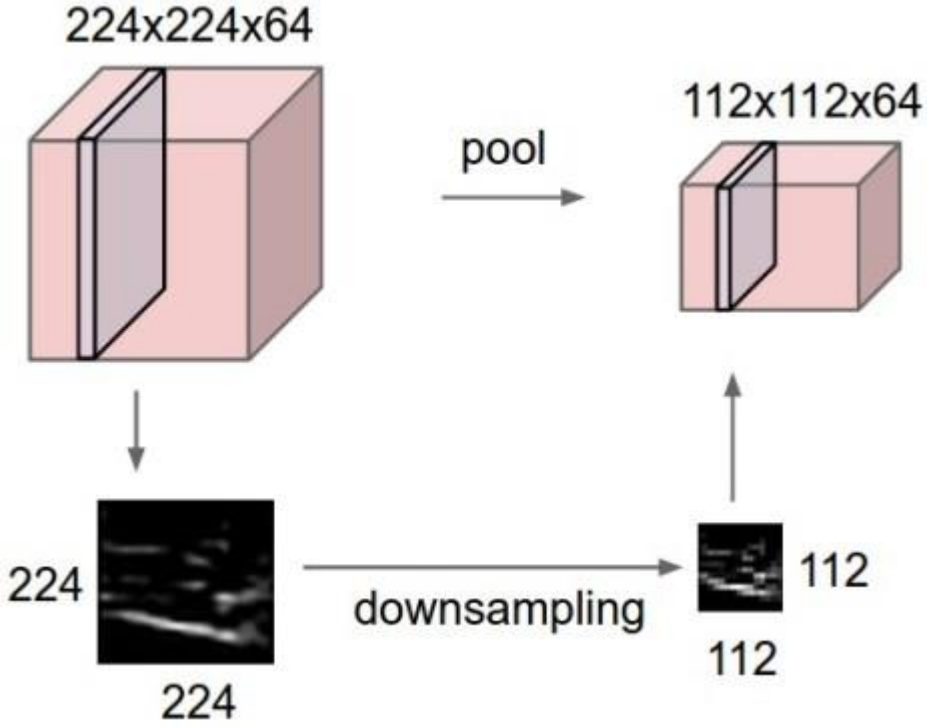
Background: Convolutional Neural Network

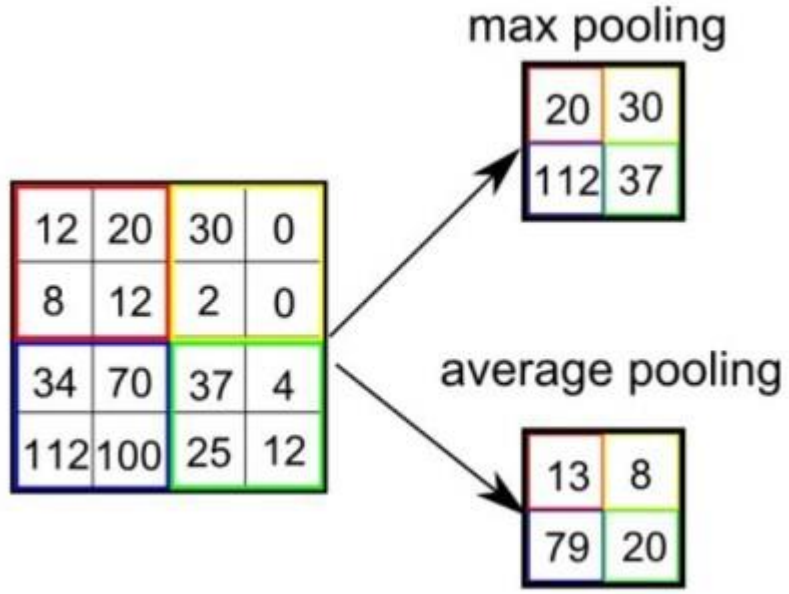
- Convolutional layer



Background: Convolutional Neural Network

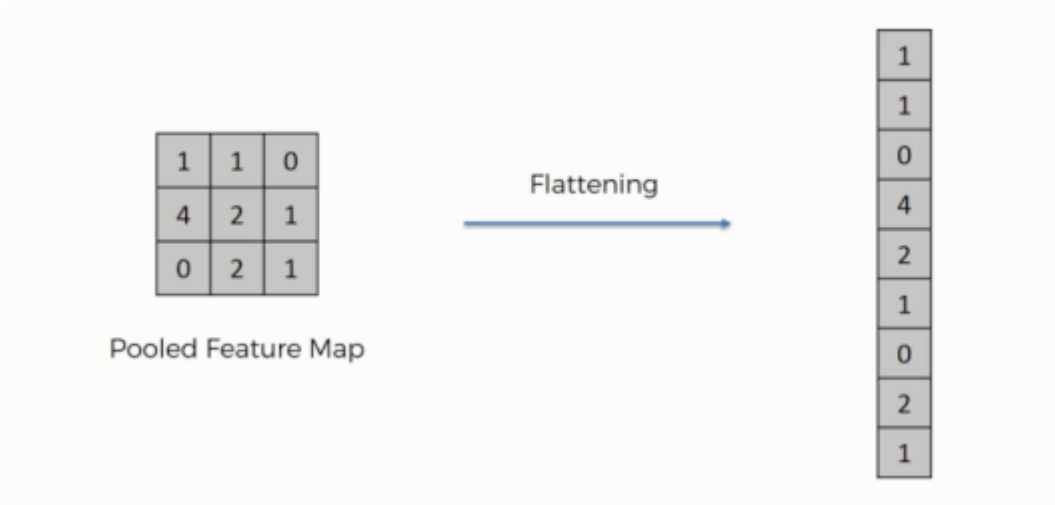
- Pooling layer





Background: Convolutional Neural Network

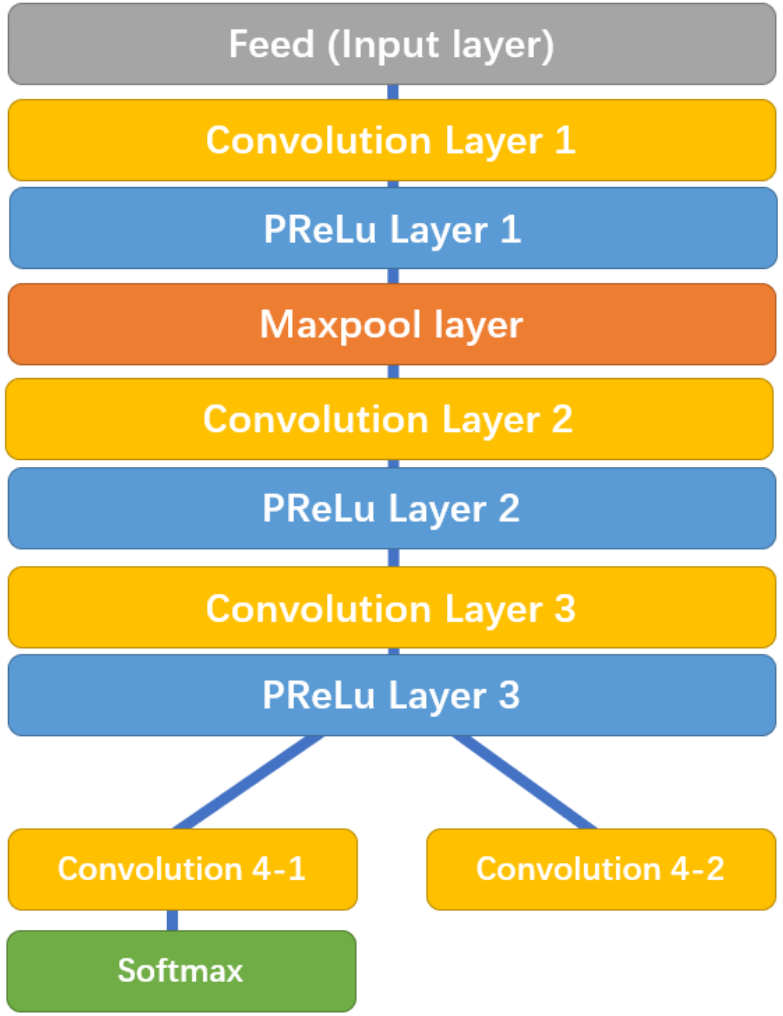
- Fully connected layer



Background: Multi-task Cascaded Convolutional Networks

- P-Net

- The algorithm uses a 12x12 kernel that runs through each image to find faces.

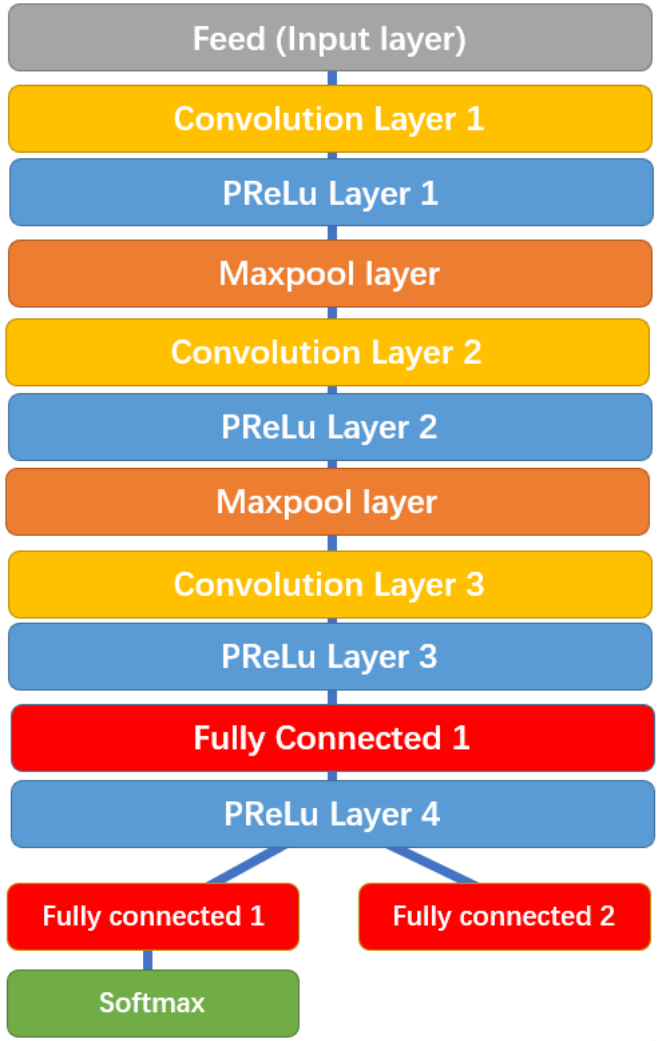


Background: Multi-task Cascaded Convolutional Networks

- R-Net

- R-Net has a similar structure to P-Net.

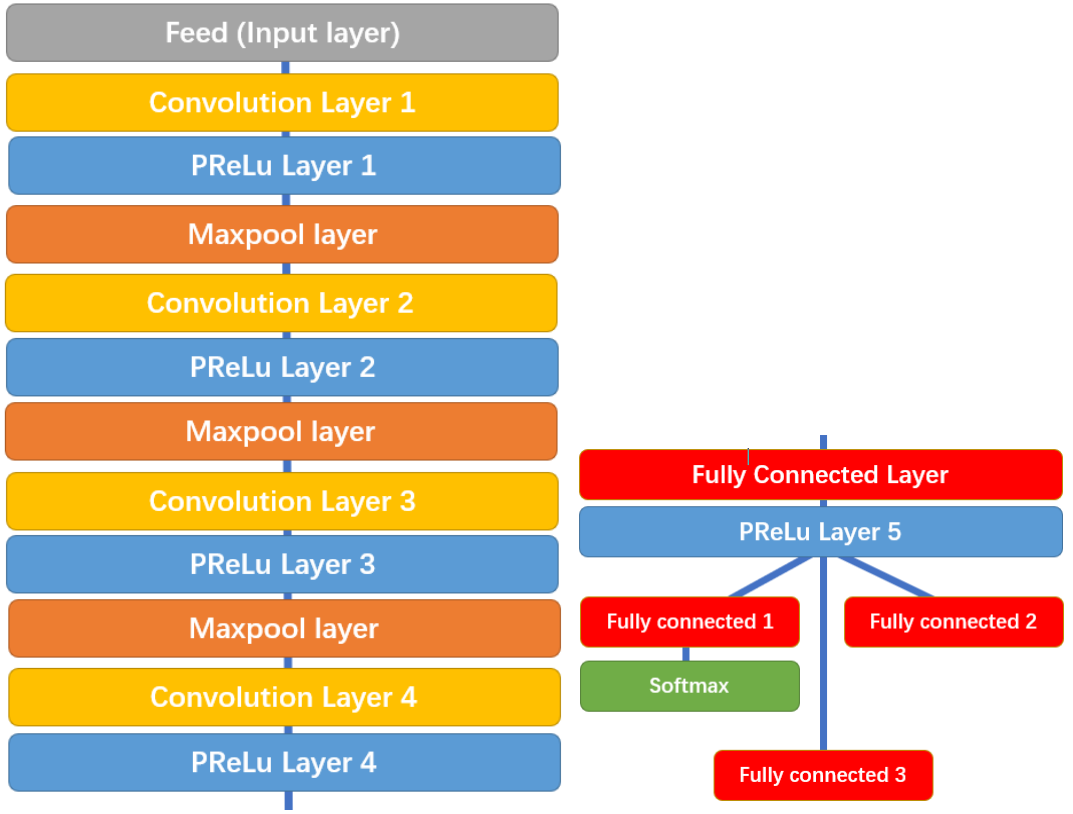
However use more layers. Here, the network will use bounding boxes provided from P-Net and tweaked as coordinates.



Background: Multi-task Cascaded Convolutional Networks

- O-Net

- O-Net takes bounding boxes from R-Net as input and marks the coordinates of the markers on the face



Background: PERCLOS

- **PERCLOS** (percentage of eye closure)

- defined as the 80-100% ratio of the time the eyes are closed over a 1-minute interval, is considered the single most powerful measure of alertness.

Background: PERCLOS

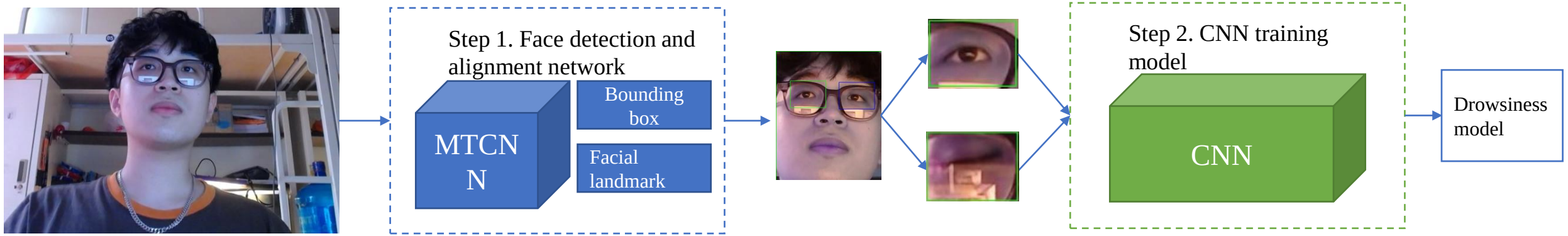
- **PERCLOS** (percentage of eye closure)

- defined as the 80-100% ratio of the time the eyes are closed over a 1-minute interval, is considered the single most powerful measure of alertness.

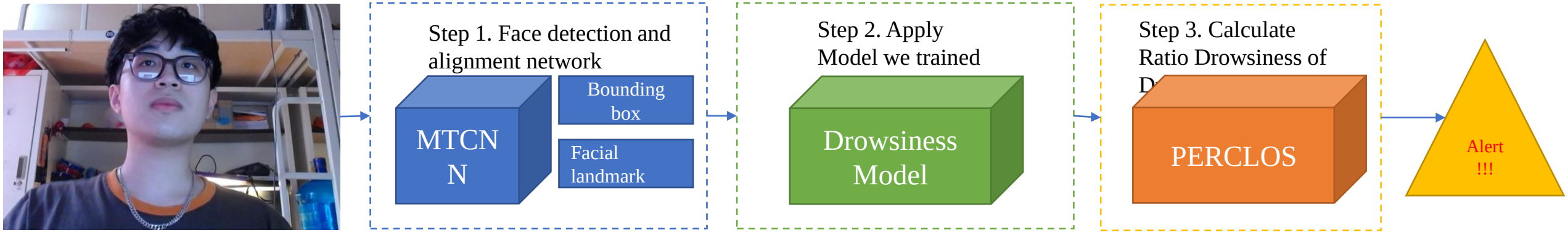
- is also the most effective method to detect drowsiness, analyzing the driver's sleepiness level using the eye state.

Contents

Methods: Architecture



Methods: Architecture



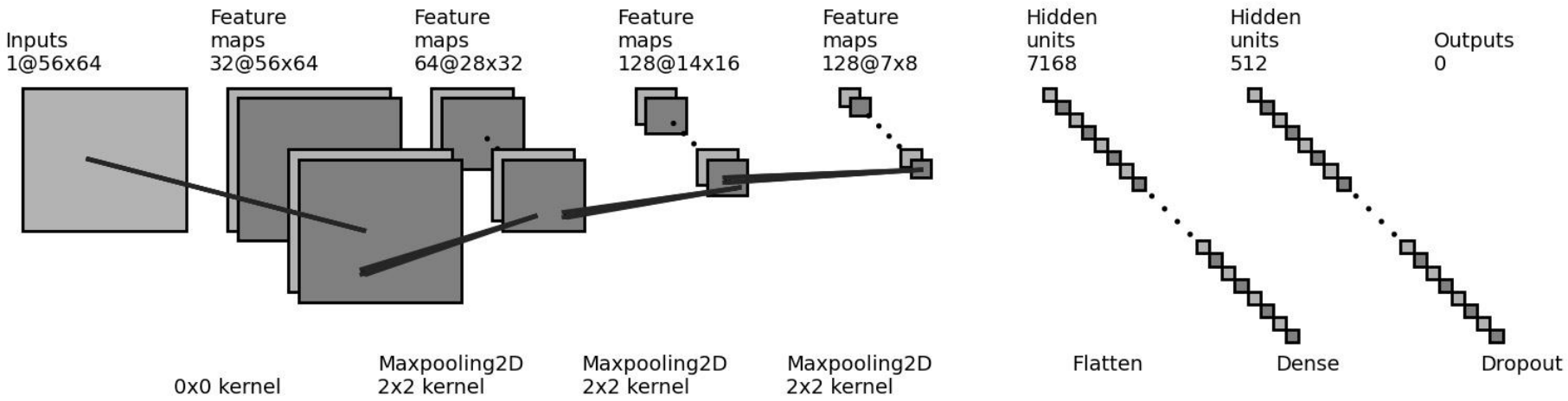
Methods: Datasets

Methods: Datasets

- Kaggle
- References “Pupil Localization Using Geodesic Distance”
- Dataset search

Methods: Preprocessing

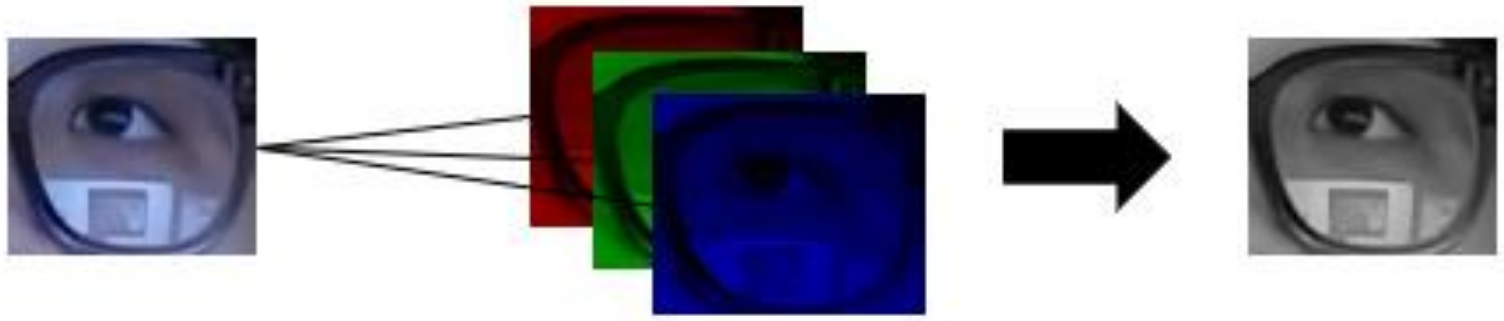
Methods: Convolutional Neural Network for Drowsiness Detection



Methods: Convolutional Neural Network for Drowsiness Detection

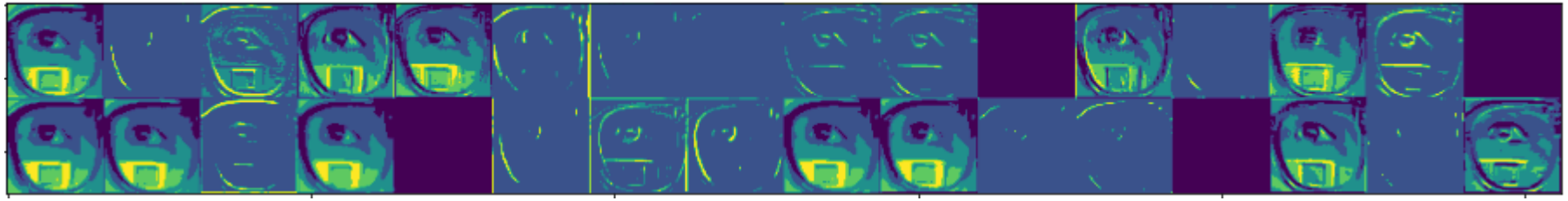



Methods: Convolutional Neural Network for Drowsiness Detection

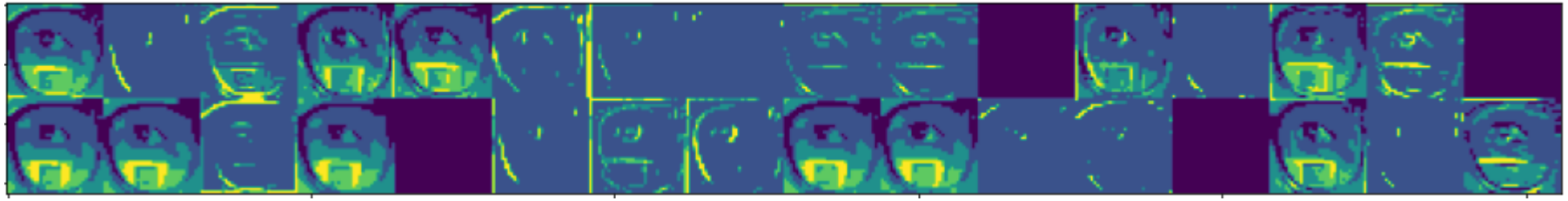


Methods: Convolutional Neural Network for Drowsiness Detection

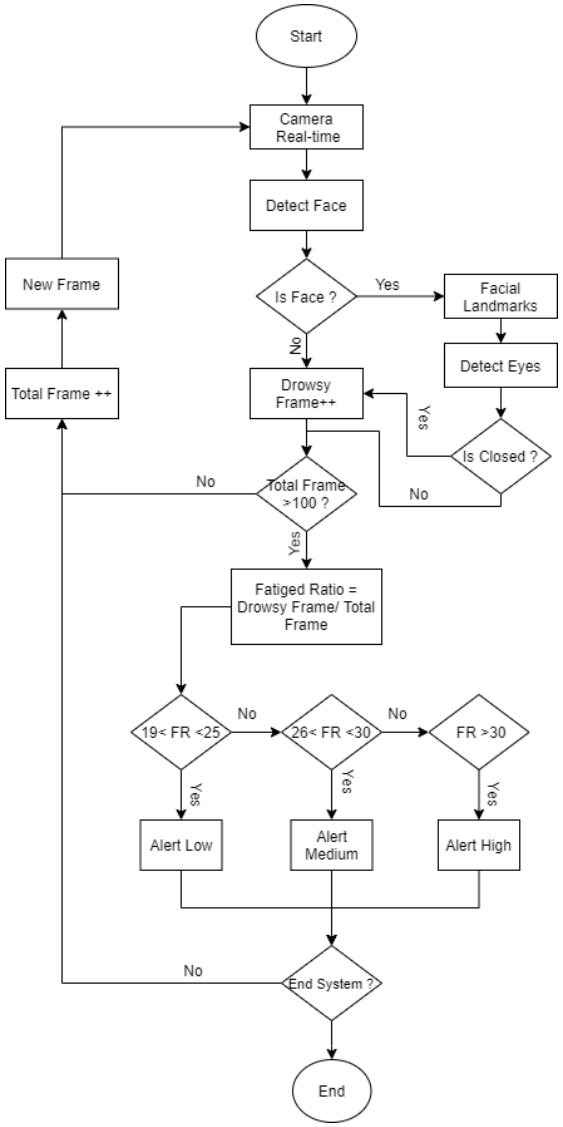
conv2d



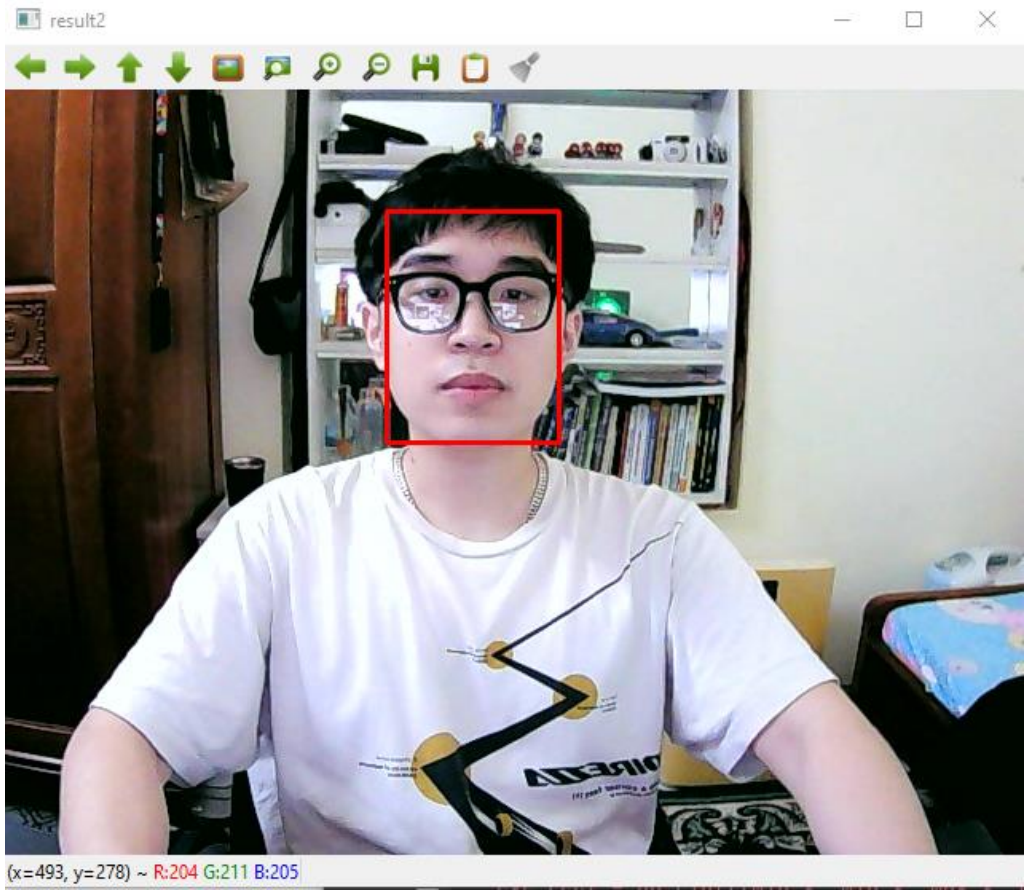
max_pooling2d

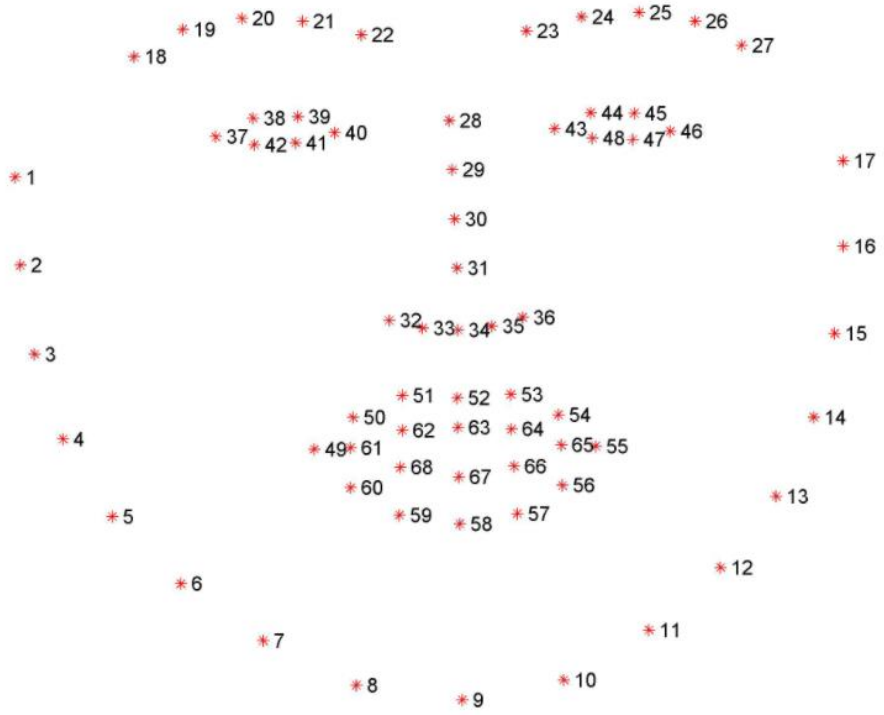


Methods: Drowsiness Detection Tracking Eye System

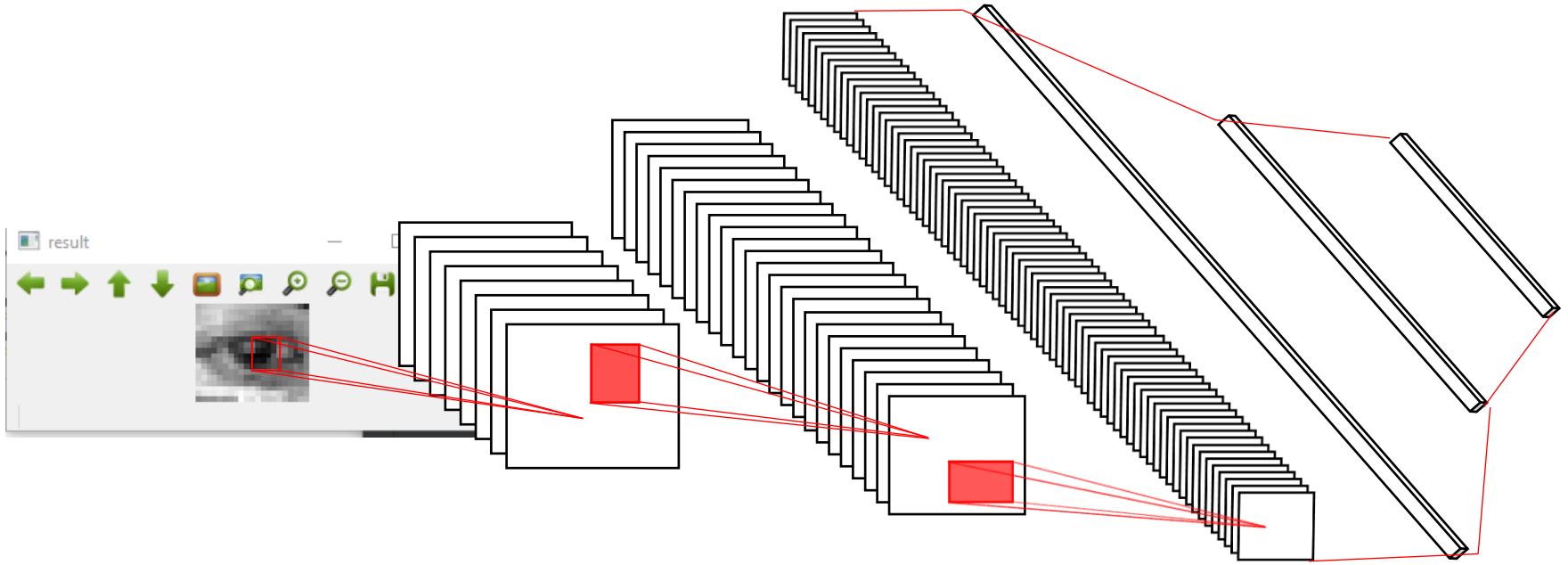


Methods: Drowsiness Detection Tracking Eye System





Methods: Drowsiness Detection Tracking Eye System



Methods: PERCLOS and Alert

- 19-25%: Low Alert
- 26-30%: Medium Alert
- > 30%: High Alert

Contents

Experiment and Results

- A computer with an Intel Core i7 - 10700 processor, an Camera(Webcam) 1080P to help stream and record full HD video at 30 frames per second

Experiment and Results

- A computer with an Intel Core i7 - 10700 processor, an Camera(Webcam) 1080P to help stream and record full HD video at 30 frames per second.

- We want to use pure CPU to embed it on some mobile phone or some car software at an optimal cost easily.

Experiment and Results

- We tried with TensorFlow and PyTorch, and the result shows that PyTorch is more efficient.

Test FPS			
Case	Frame	Time	FPS
Camera normal	100	3.83(s)	26-27
Camera using mtcnn with TensorFlow	100	28.3(s)	3-4
Camera using mtcnn with PyTorch	100	10.6(s)	9-10

Table 4.1: FPS when run with the system testing.

Experiment and Results

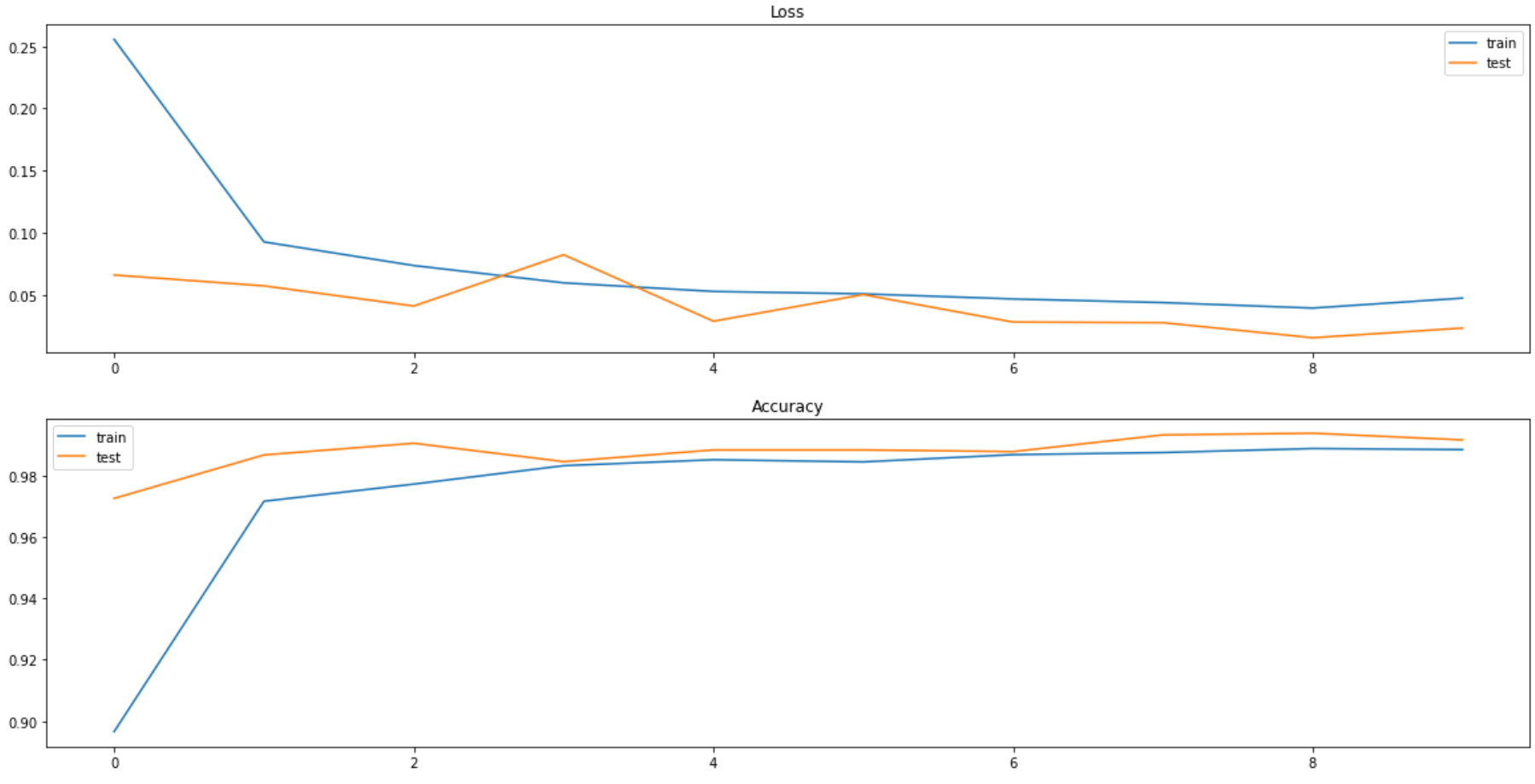
- CPU still takes an extreme amount of time and is not really feasible for fine-tuning in the scope of this project.

Experiment and Results

- CPU still takes an extreme amount of time and is not really feasible for fine-tuning in the scope of this project.

- So we switched to using Google Colab with GPU K80s to optimize the learning model time and possibly help us test faster.

Experiment and Results



Experiment and Results

- We tested on 15 people at FPT University, and the result can be very good with accuracy greater than 90%.

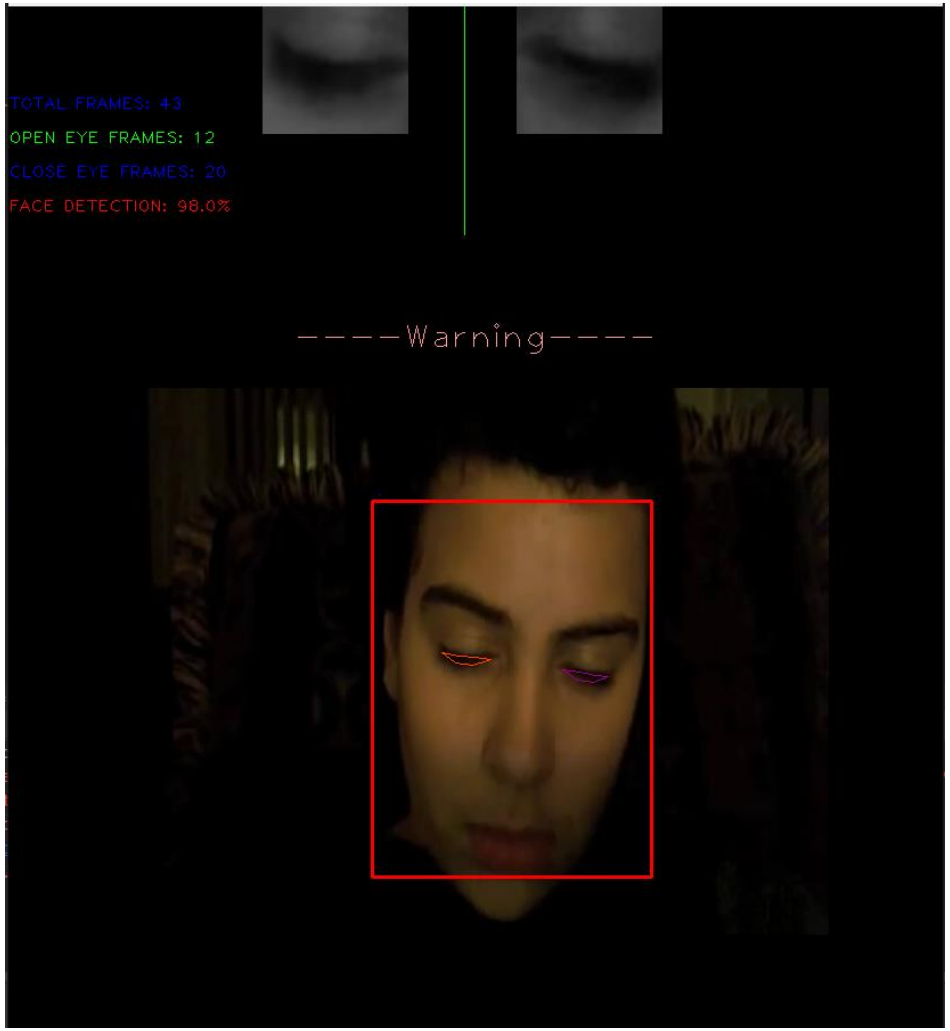
Test Case		
Case	Amount of People	Accuracy
Bare face	15	92.8%
Wearing Glasses	15	90.4%

Table 4.2: Drowsiness Detection Case Testing.

Experiment and Results

- We also tried on some videos on Youtube have content fell asleep and the feasibility of the results.

Experiment and Results



Contents

Conclusion

- Our driver’s drowsiness eye-tracking system is a combination of multiple neural networks.

Conclusion

- Our driver’s drowsiness eye-tracking system is a combination of multiple neural networks.
- Presented NN, CNN, and MTCNN about architecture, its learning process, and how to apply it to a supervised learning problem.

Conclusion

- Our driver’s drowsiness eye-tracking system is a combination of multiple neural networks.
- Presented NN, CNN, and MTCNN about architecture, its learning process, and how to apply it to a supervised learning problem.
- Each model will have certain strengths and jobs, so we have exploited the model’s strengths to make the system work in the best way.

Conclusion

- Our driver’s drowsiness eye-tracking system is a combination of multiple neural networks.
- Presented NN, CNN, and MTCNN about architecture, its learning process, and how to apply it to a supervised learning problem.
- Each model will have certain strengths and jobs, so we have exploited the model’s strengths to make the system work in the best way.
- Our experiments showed that the combination models learn better than the original neural model and achieve higher accuracy.

Contents

Future Works

- The extension ability:
  - Embed the system to the smart devices (phones, car systems, ..)

Future Works

- The extension ability:
  - Embed the system to the smart devices (phones, car systems, ..)
  - The system may recommend the motel driver or where the driver can stop and take naps to stay awake before returning to the road.

Future Works

- The extension ability:
  - Embed the system to the smart devices (phones, car systems, ..)
  - The system may recommend the motel driver or where the driver can stop and take naps to stay awake before returning to the road.

- The accuracy ability:
  - Find some other models that can increase the exact proportions

Future Works

- The extension ability:
  - Embed the system to the smart devices (phones, car systems, ..)
  - The system may recommend the motel driver or where the driver can stop and take naps to stay awake before returning to the road.

- The accuracy ability:
  - Find some other models that can increase the exact proportions
  - Incorporate some more facial details, such as the ratio between the driver's eyebrows- eyes, mount- eyes

Thank You
For Your Attention