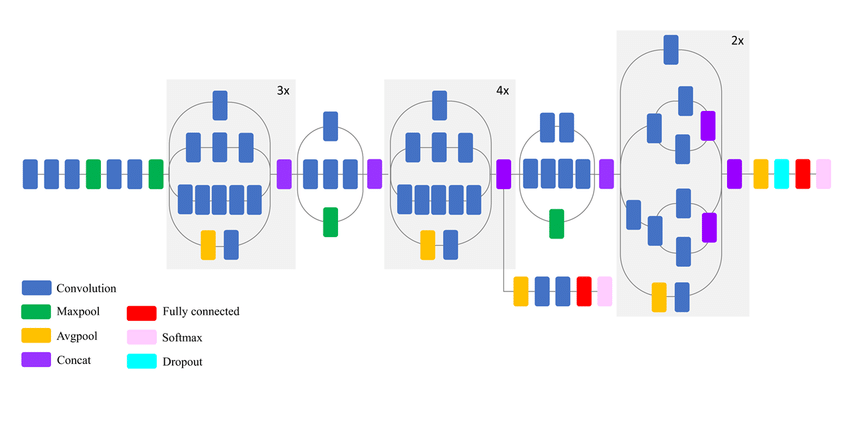
**Proposed Method**

Diagram

Description automatically generated

In this method a simple drowsiness detection for driver is purposed using deep learning. This method consists of transfer learning where inceptionv3 model is used as a based model and combined with purposed neural network to decide either the eyes is close or not to define the state of the driver as discussed before.

1. base model

The inceptionV3 model architecture is depict in ``Fig \ref{}’’. The architecture within the box with dotted border is the part that are implemented in proposed method and it consist of 4 main component, which are convolution, maxpool (max pooling), avgpool(average pooling) and concat (Concatenet).

These components are abstraction of layers of neuron or neural network \cite{practicle}. The main objective of convolution is to extract features from the input image and produce feature maps. The output feature maps in the initial convolutional layer may learn to detect basic features, such as edges and colour composition variation \cite{intelligent}. By considering the input image as 3d matrix or tensor where hight, width and depth of the image is the parameter, the feature extraction is done by multiplying the matrix with another 3d matrix called filter or kernel as depict in ``Fig \ref{}’’. The values in filter tensor are fixed in the way that a certain feature can be extracted, and the filter normally has a smaller size. So, the filter tensor will go through the image tensor by shifting the column and row of the image tensor. The step of shifting is formally called stride. Since the result of multiplication of a matrix will produce smaller matrix, the resulting image also have the smaller tensor. However, some convolution layer can maintain the size of input tensor by expending the input tensor. This process is known as padding. \cite{fundemental}

Pooling is a function where the spatial size of the representation is reduced to reduce the amount of parameter and computation in the network. Pooling layer operates on each feature map independently. There are two type of pooling which are max pooling and average pooling \cite{fundemental}. Max pooling is operated by selecting a region with a fixed size in input tensor and select the highest value to form a new tensor as an output. The region than move by the step of fixed stride. As a result, smaller tensor is produced. This process is depict in ``Fig \ref{}’’. In average pooling layer, the output is the average value of selected region.

Concatenate layer is the layer that combine more than one input tensor to become one output tensor. This out put will be use by the next layer that require tensor that are bigger than one of the input tensors sizes.

1. Purposed network

The purposed neural network is visualized in ``Fig \ref{}’’ where the input from the inceptionV3 model is flatten. Since the final output is only binary decision because only the state of the eyes need to be detected, either they are close or not, the output from flatten layer is reduce to 64 neuron as an input to dropout layer to reduce the number of output into two output.

Activation function

Relu

Softmax

**Experiment**

As discussed before, the neuron has two main parameter, namely weight and bias. These parameters are vital to produce the right output, so that a precise decision can be made. The method that is used to find the right value for the parameter is called training and the data that are used for training is called data set. The model is then tested at the end of training phase or known as inference to evaluate the model.

<<<<work flow>>>>>

Diagram

Description automatically generated

<<<Before discussing the detail about the model use case and implantation, the overall workflow will be explained first. The model will be implemented in two stage, training, and application. In the training phase, the model’s parameter is adjusted, so that the detection accuracy can be improved. In application stage, the model will use the trained parameter to make prediction in a real environment. In the next section, the detail about training stage will be discussed further.>>>

1. Data set

Data sets are used as an input for the model to enable it to be train. The output that produce will be evaluated.

Datasets are usually grouped into batches to handle huge number of data. Some people use the term iteration loosely and refer to putting one batch through the model as an iteration.

1. Training

training a network means nothing more than solving a complex optimization problem \cite{advance}. At first, the value of weight and bias for every neuron is randomly assigned. After that, an image is given to the model as an input. The final output value is compared with an expected value, where the difference value or known as loss value is recorded. The loss values are used by loss function to improve the value of the weight in the way that the loss value can be reduced in the future. The lost function that is used in proposed method is categorical cross entropy.

Since the provided dataset for the proposed method is very large, having more than one epoch is necessary. An epoch indicates the number of passes of the entire training dataset the model has completed. The general relation where dataset size is d, number of epochs is e, number of iterations is i, and batch size is b would be d\*e = i\*b. \cite{epoch}

Determining how many epochs a model should run to train is based on many parameters related to both the data itself and the goal of the model.

1. Inference

During the training phase, not all data from the dataset are used for training, some of them al also used to evaluate the model. This process is known as inference. For instance, 80% ofthe dataset for the proposed method is used for training and 20% are used for inference. As a result of inference, accuracy of the model can be evaluated.

1. Platform

The proposed method is design using python environment and the API used for the model are from TensorFlow\footnote{}. The training phase was executed on computer with 64-bit operating system, x64-based processor, Intel(R) Core(TM) i7-1065G7 CPU @ 1.30GHz 1.50 GHz and 16 GB RAM.

1. Result

Deep learning

neuron

Related work

Proposed method

Base mode  
 purposed network

Work flow

Experiment

Data set  
Training phase  
Inference

Platform

result