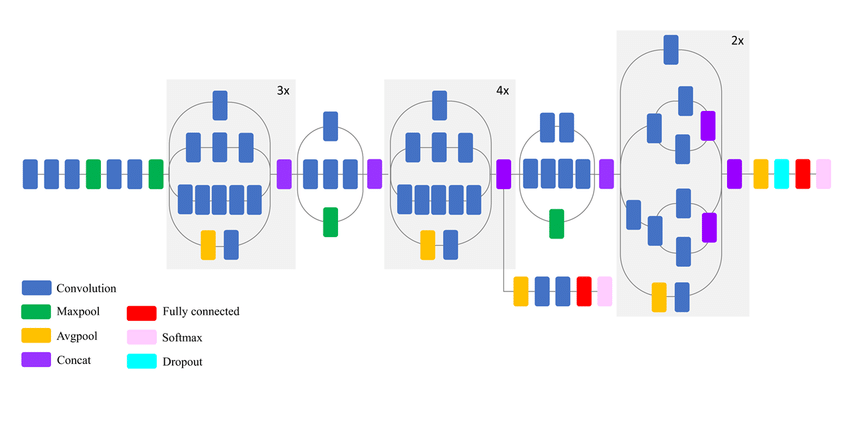
**Proposed Method**

Diagram

Description automatically generated

In this section a proposed method for detecting driver drowsiness using deep learning is discussed. As previously brought up, this method will use module technique, transfer learning and image augmentation. The module technique will allow the proposed neural network to work on fewer cases which is detecting either driver’s eyes is close or open to detect the driver drowsiness base on the paper’s drowsiness definition.

First, we will use cascade classifier provided in OpenCV package to crop the region of interest from the input which is the image from camera. The output image, which is the image of the eyes of the driver will be the input to the proposed model to detect either the eyes are close or not. The input image resolution is converted to 80 x 80 with Grayscale and 3 channel to be compatible with the model. If close eye is detected with longer period than usual, an alarm will be executed as a sign that the driver drowsiness is detected.

The proposed model consists of transfer learning where inceptionV3 model is used as a based model and combined with purposed neural network. With the help of transfer learning, knowledge may be adapted or transferred from one set of activities or domains to another. The structure of inceptionv3 is depict in ``Fig. \ref{}’’ and the part that implemented in the proposed method is within the box with red dotted border and it consist of 4 main component, which are convolution, maxpool (max pooling), avgpool(average pooling) and concat (Concatenate). These components are abstraction of layers of neuron or neural network \cite{practical}.

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{fundamental}.

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{fundamental}. 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 output will be use by the next layer that require tensor that are bigger than one of the input tensors sizes.

The output from inceptionV3 is then become the input to the proposed neural network, depict 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 neurons as an input to dropout layer to reduce the number of output into two output.

Aforesaid, every neuron has its own activation function to solve a certain problem. As shown in ``Fig. \ref{}’’, the activation used in proposed neural network is ReLU and Softmax, where ReLU is use for neurons in the first dense layer and Softmax is used for neurons in the second dense layer.

Deep learning

neuron

Related work

Proposed method

Base mode  
 purposed network

Work flow

Experiment

Data set  
Training phase  
Inference

Platform

result