SRM
UNIVERSITY AP
Amaravati

Emotion Detection Through Realtime Facial Gestures - CNN

Team Members

1)Purna Praveen.B(AP18110010379)
2)JenvithManduva(AP18110010418)
3)Ramneesh Valisetty(AP1811010446)

)

INTRODUCTION

- Facial emotions or expressions can be recognized by computers and enhancing modern day machines to understand human emotions from their real life time.
- In 1978 Ekman and Friesen are among the first people who are interested in facial expression recognition they developed a Facial Action Coding System(FACS) in which facial expressions are described by Action Units(AU), they will break down the human expression into 46 AUs each AU is coded with one or more facial muscles.
- Facial expressions are classified into three categories
  i)Face Detection ii)Feature Extraction
  iii)Emotion Classification

- Some companies that make use of “Emotion Detection’ which are Disney, Kellogg’s, Unilever .
- In this paper , to recognize real time facial emotions we have been used Convolutional Neural Network(CNN).
- Convolutional Neural Network(CNN) has multiple layers which are convolutional layers and fully connected layer.These layers will help us to deduct the facial emotions.
- Through our research on Convolutional Neural Network(CNN) we have proposed our Convolutional Neural Network(CNN) it is in Figure 1 .
- We have been used several layers to train the model which are Input, Conv 2,
- Batch Normalization, MaxPooling2D, Activation, Dropout, Dense layers in both convolutional and fully connected layers.
- We have been used OpenCV to capture Real Time Face through System or Machine camera.

DATASET

- The dataset was obtained from Kaggel , developed by author Jonathan oheix. In this dataset it has two directories: the first one is the train set and the second one is the validation set.
- Both train and validation sets have seven directories; they are Happy,Sad,Fear,Anger,Disgust,Surprise,and Neutral. The dataset contain all grayscale images with dimensions of 48*48 pixels.

- The training set consists of 25,121 images examples.The validation set consists of 7,066 images examples.The total dataset consists of 35,887 images examples.

METHODOLOGY

- In this paper we have been used Convolutional Neural Network(CNN) to deduct real time facial emotions, it is a deep learning algorithm.
- The CNN architecture is an analogous to the connectivity pattern of neurons in the human brain and inspired by the organization of the visual cortex.It uses a system much like a multilayer perceptron it is designed for reduced processing requirements.
- The major thing is to remember about any deep learning model is that it requires large amount of data to train and requires a lot of computer resources .
- The advantage of CNN's over other classification algorithms (SVM,Random-Forest,etc..) is that the CNN's learns the best features to represent the objects in the mages and has a high generalization capacity, being able to precisely classify example like car number plate recognition,biometry,etc.

- We have been used Seven layer which are four are convolutional layers and three are fully connected layers .
- In both Four Convolutional Layers and three fully connected layers they are multiple layers inside them follows as:
- **INPUT LAYER**:- The input layer is a tensor with shape:(inputs height(H)) X (input width(W)) X (input channels(C)).

  ---> In our case , the input layer shape:(48X48X1) where 48 X 48 are the shape of an image and 1 is for the gray scale images.
- **CONVOLUTIONAL LAYER**:-Convolutional Layers are used to apply a convolution operation to the input, passing the result to the next layer.

  ---> In our case, we have been using 2D Convolutional Layer and usually abbreviated as conv2D.
- **BATCH NORMALIZATION:-** Batch Norm is the normalization technique done between the layers of a Neural Network instead of in the raw data.Following technique is used as normalization formula of batch norm as:

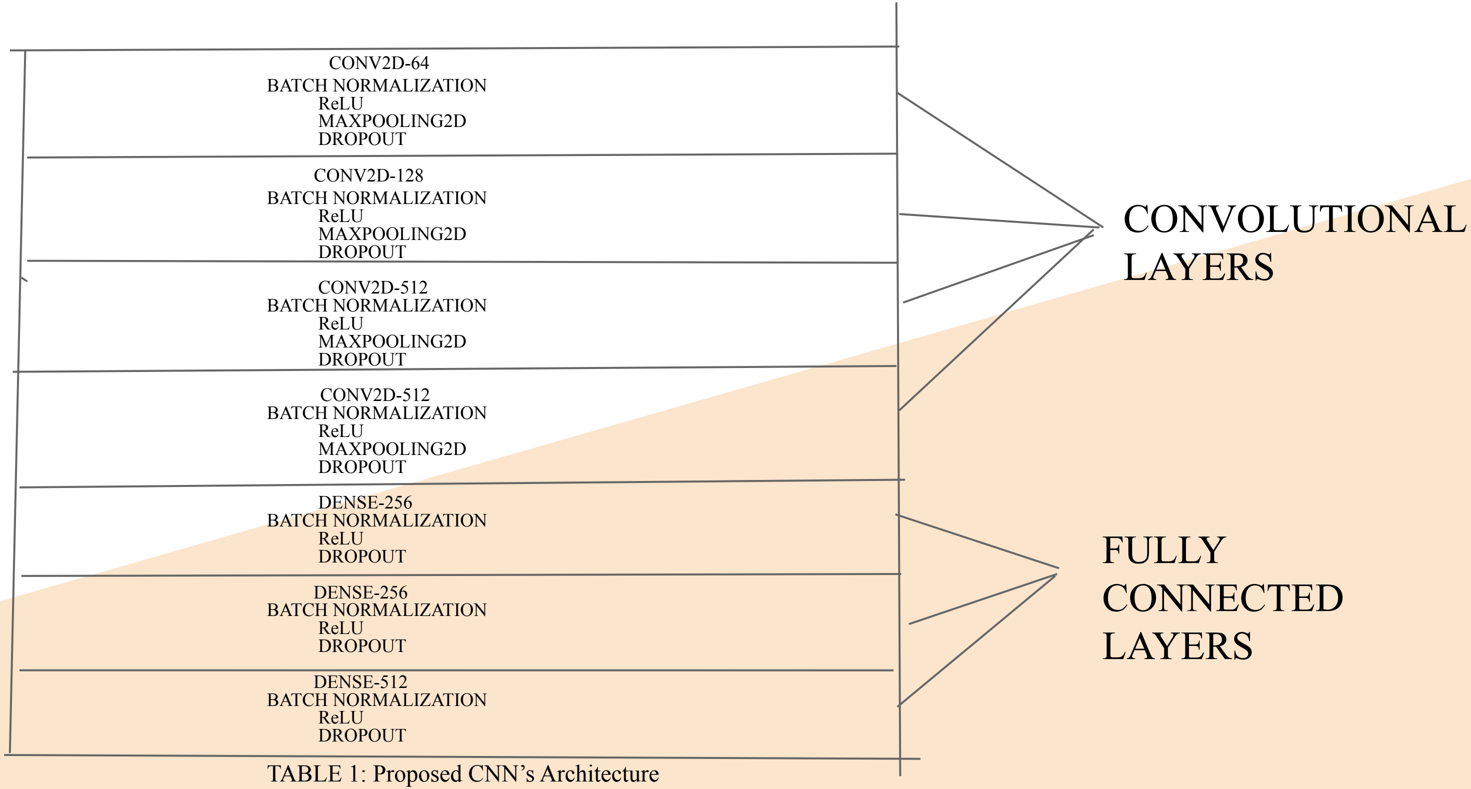
$$Z^N = (Z - m_Z \div s_Z)$$

- **ACTIVATION LAYER:-** The activation function is the non linear transformation that we do over an input signal. This transformed output is then sent to the next layer of neurons as input.It is a node that is put at the end of or in between neural networks.They helps to decide if the neuron would fire or not.
  ----> In our case, we have been using a Rectified Linear Unit (ReLU). ReLU converges six times faster than tanh and sigmoid activation functions.
- **POOLING LAYER:-** Pooling Layer is used for pooling operation, like a filter to be applied for feature maps.
  ---> In our case, we have been using Max pooling with size 2 x 2 pixels.
- **DROPOUT LAYER:-**Dropouts are the regularization technique that is used to prevent overfitting in the model. It is always good to only switch off the neurons maximum to 50%. If we switched off more than 50% then there can be chances when the model learning would be poor and the predictions will not be good..

  --->In our case, we have been switched off upto 25% by using a dropout layer.

- **FULLY CONNECTED LAYER:-** The fully connected layer (FC) operates on a flattened input where each input is connected to all neurons.In fully connected layers instead of convolutional layers we have been using the ‘Dense’ layer..
- For each layer of the Artificial Neural Network, the following calculation takes place $g(W_X + b)$.
  → It is repeated for each layer..
- **DENSE LAYER:-** A Dense layer is a fully connected layer, all the neurons in a layer are connected to those in the next layer.
  --->After passing through the fully connected layers, the final layer is added all the fully connected layers to dense and uses the 'Softmax Activation Function’ which is used to get probabilities of the input in a particular class.
- So, finally we have the probabilities of the objects in the image belonging to the different classes.

PROPOSED CNN’S ARCHITECTURE



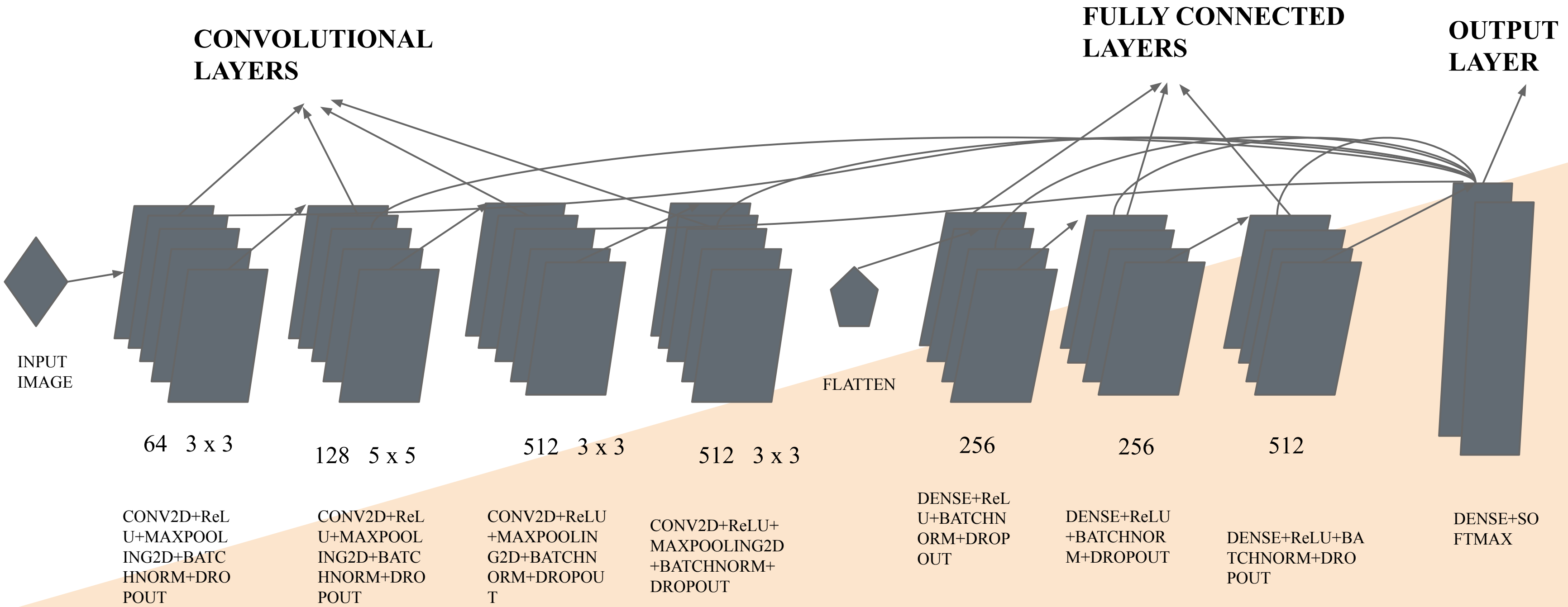
CONV2D-64 BATCH NORMALIZATION ReLU MAXPOOLING2D DROPOUT
CONV2D-128 BATCH NORMALIZATION ReLU MAXPOOLING2D DROPOUT
CONV2D-512 BATCH NORMALIZATION ReLU MAXPOOLING2D DROPOUT
CONV2D-512 BATCH NORMALIZATION ReLU MAXPOOLING2D DROPOUT
DENSE-256 BATCH NORMALIZATION ReLU DROPOUT
DENSE-256 BATCH NORMALIZATION ReLU DROPOUT
DENSE-512 BATCH NORMALIZATION ReLU DROPOUT

CONVOLUTIONAL LAYERS

FULLY CONNECTED LAYERS

TABLE 1: Proposed CNN’s Architecture

PROPOSED CNN’S ARCHITECTURE



CONVOLUTIONAL LAYERS

FULLY CONNECTED LAYERS

OUTPUT LAYER

INPUT IMAGE

64 3 x 3

CONV2D+ReLU+MAXPOOLING2D+BATCHNORM+DROPOUT

128 5 x 5

CONV2D+ReLU+MAXPOOLING2D+BATCHNORM+DROPOUT

512 3 x 3

CONV2D+ReLU+MAXPOOLING2D+BATCHNORM+DROPOUT

512 3 x 3

CONV2D+ReLU+MAXPOOLING2D+BATCHNORM+DROPOUT

FLATTEN

256

DENSE+ReLU+BATCHNORM+DROPOUT

256

DENSE+ReLU+BATCHNORM+DROPOUT

512

DENSE+ReLU+BATCHNORM+DROPOUT

DENSE+SOFTMAX

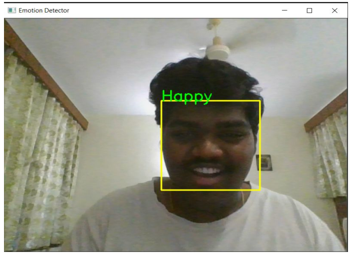
FIGURE 1:PROPOSED CNN’S ARCHITECTURE

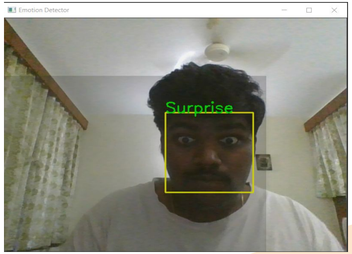
EXPERIMENTAL ANALYSIS

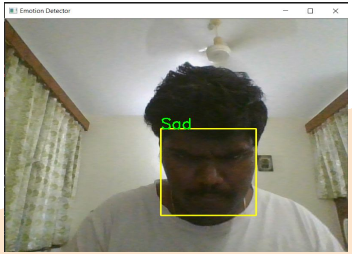
- From proposed CNN’s architecture , we have added all layers to train the model and to fit the model in two directories which are test and validation directories from dataset.
- After all done with training and validation we have measured out trained model accuracy. The output of trained model it is termed as ‘model.h5’.
- By the model.h5 which is output of trained model added to OpenCV to import images and preprocess the images.
- The haarcascade_frontalface_default.xml is added in OpenCV to deduct the face with rectangle box around the face with the help of model.h5 facial emotions get recognized.
- **We have been added 3rd Fully connected layer from existing CNN which contains dense,batch normalization,dropout,Activation layers in it because of this layer accuracy increased.**

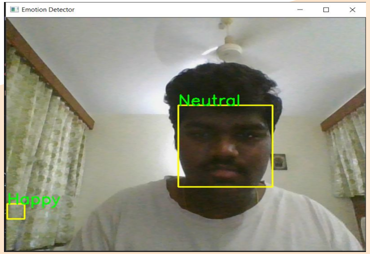
RESULTS

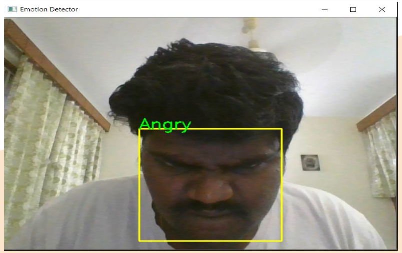
In Figure 2, One of our team member his realtime facial emotions has been recognized with proposed CNN’s architecture.We have obtained accuracy up to 72% from our trained model.



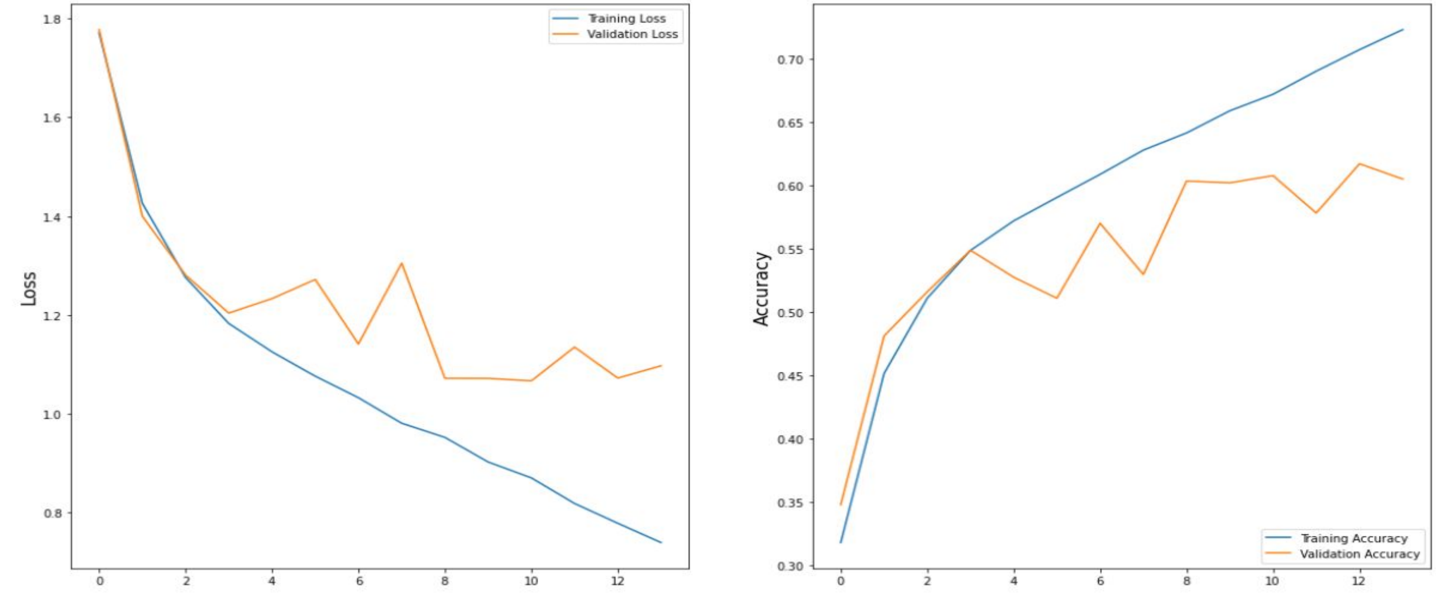








PLOTTING ACCURACY AND LOSS



CONCLUSION

- In this paper the main aim is to recognize emotions through real time facial gestures.we have obtained accuracy of 72%.The achieved results are satisfactory as the average accuracies and therefore, this CNN model is accurate.
- **Before adding 3rd fully connected layer an accuracy in between 68% to 69%.After adding 3rd layer the accuracy improved to 72%.**
- For an improvement in this model and its outcome , it is recommended to to change the parameters wherever useful in CNN model and removing unwanted parameters.Resize the learning rate and adapting with in the location may helpful to improve the accuracy and model.
- The number of epochs can be set to higher number to attaining the accuracy as output.But by increasing the number of epoch may lead to overfitting.
- This similar CNN model can be used to different datasets to be trained and tested and check for its accuracy.
- In this projects , further development can obtain accuracy for each emotion , covariance between each emotion, Pi-chart for each emotion for image or real time face.

THANK YOU