Capstone Project

Live Class Monitoring System (Face Emotion Recognition)

Arman Alam

Content

Introduction

Facial Emotion recognition is a way of identifying the current emotional state of an individual to observer
Facial expressions can display personal emotions and indicate an individual's intentions within a social situation.

Facial expressions and other gestures convey nonverbal communication cues that play an important role in interpersonal relations. Example:- A simple smile can indicate our approval of a message, while a scowl might signal displeasure or disagreement. These cues complement speech by helping the listener to interpret the intended meaning of spoken words. Therefore, facial expression recognition extracts and analyzes information from an image or video feed, it is able to deliver unfiltered, unbiased emotional responses as data.

Product Development : Observing users interaction while interacting with a brand or a product helps the company to assess the effectiveness of any business product.

Video game testing phase. In this phase, usually a focus group of users is asked to play a game for a given amount of time and their behavior and emotions are monitored. By using facial expression recognition, game developers can gain insights and draw conclusions about the emotions experienced during game play and incorporate that feedback in the making of the final product.

Problem Statement

The Indian education landscape has been undergoing rapid changes for the past 10 years owing to the advancement of web-based learning services, specifically, eLearning platforms.

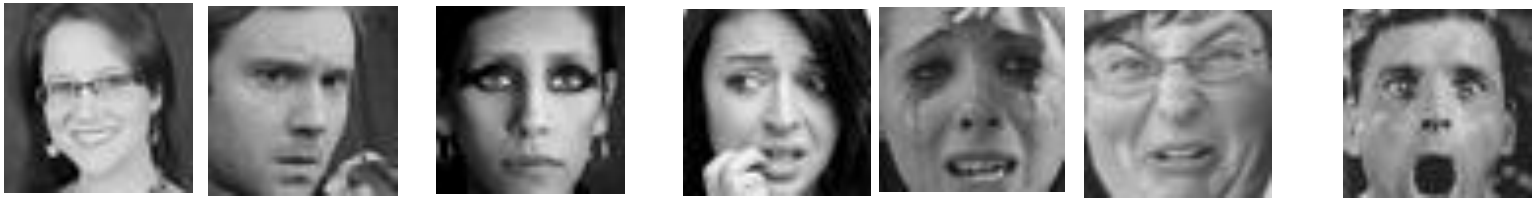
In a physical classroom during a lecturing teacher can see the faces and assess the emotion of the class and tune their lecture accordingly, whether he is going fast or slow. He can identify students who need special attention. Digital classrooms are conducted via a video telephony software program (ex- Zoom) where it's not possible for medium-scale class (25-50) teacher to see all students and access the mood. Because of this drawback, students are not focusing on content due to a lack of surveillance. While digital platforms have limitations in terms of physical surveillance but it comes with the power of data and machines which can work for you. It data can be analyzed using deep learning algorithms which not only solves the surveillance issue, but it also removes the human bias from the system.

Data Summary

Data Set link

https://www.kaggle.com/c/challenges-in-representation-learning-facial-expression-recognition-challenge

This dataset contains 35887 grayscale 48x48 pixel face images with seven emotions.

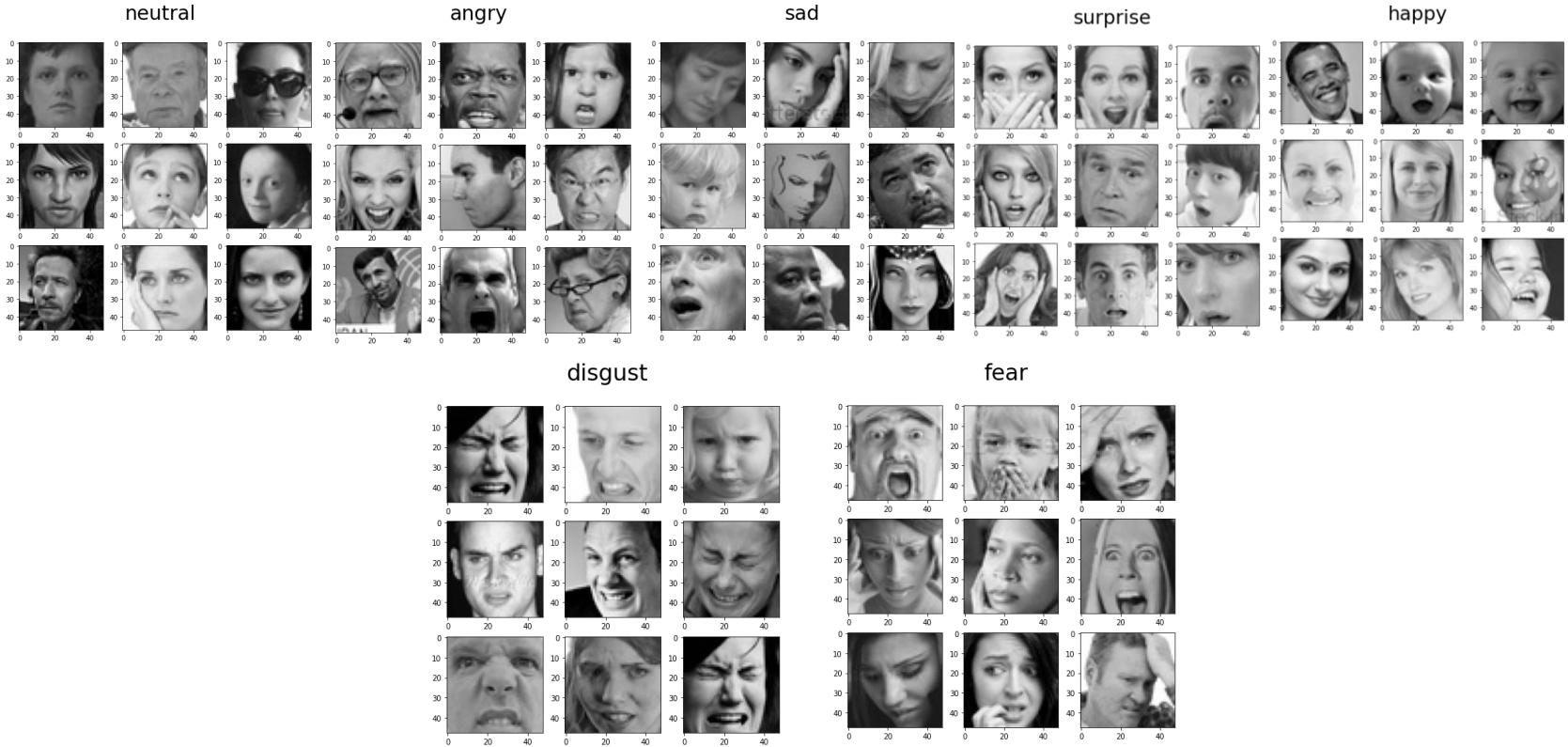


HAPPY ANGRY NATURAL FEAR SAD DISGUST SURPRISE

Data Summary

Emotion	No. of images for Training	No. of images for Testing
Angry	3995	958
Disgust	436	111
Fear	4097	1024
Happy	7215	1774
Sad	4830	1247
Surprised	3171	831
Neutral	4965	1233

Data Summary



These are some randomly generated images for each emotional expressions

Pipeline

Data Exploration → **Modeling** → **Model evaluation & deployment**

Understanding the data

- Types of emotions
- Images in each category

Modeling structures

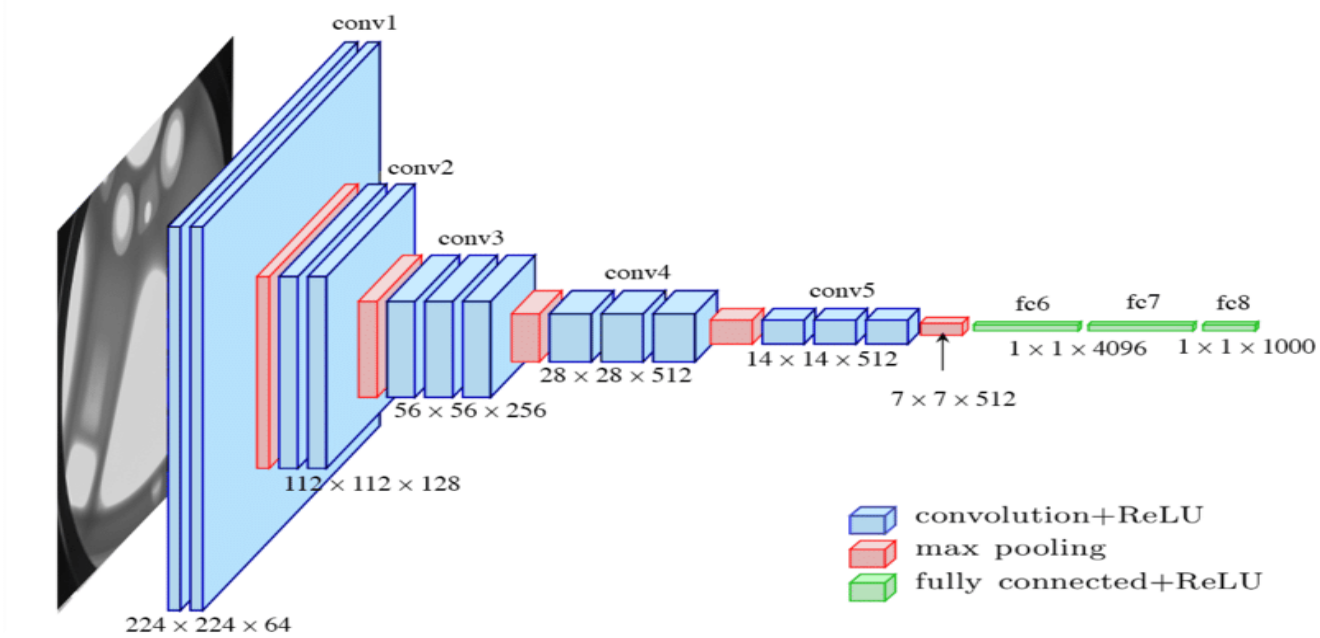
- VGG16
- CNN

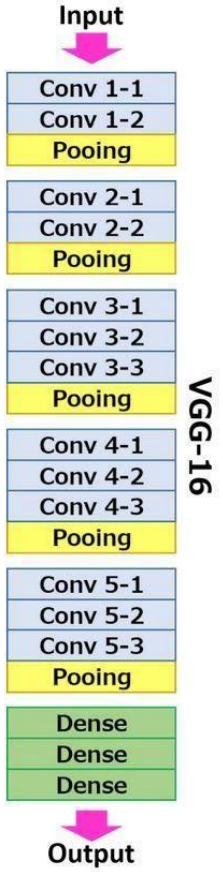
Graphs and applications

- **Loss & accuracy plots**
- **Confusion matrix (Heatmap)**
- **Streamlit**
- **Heroku**

VGG16 Model

1) Transfer Learning (VGG16)





Modeling Steps

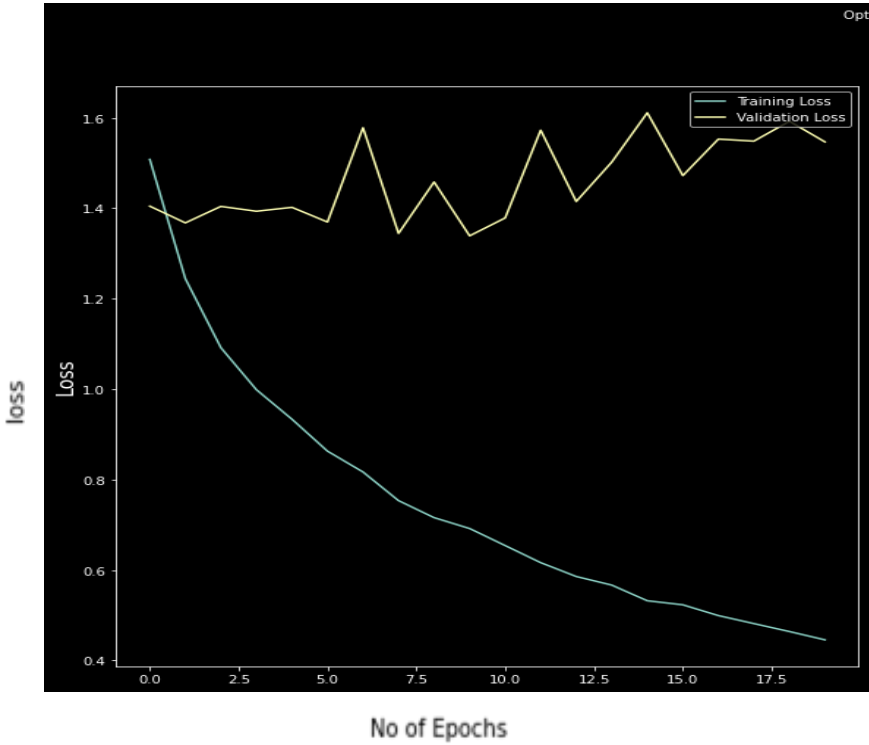
- Pre trained 13 conv layers
- Flatten layer
- Dense Layer

- Activation Function - ReLu, Softmax
- Epoch - 50
- Optimizer - Adam
- Batch size -32
- Callbacks- EarlyStopping, ReduceLROnPlateau

- Loss and accuracy plots

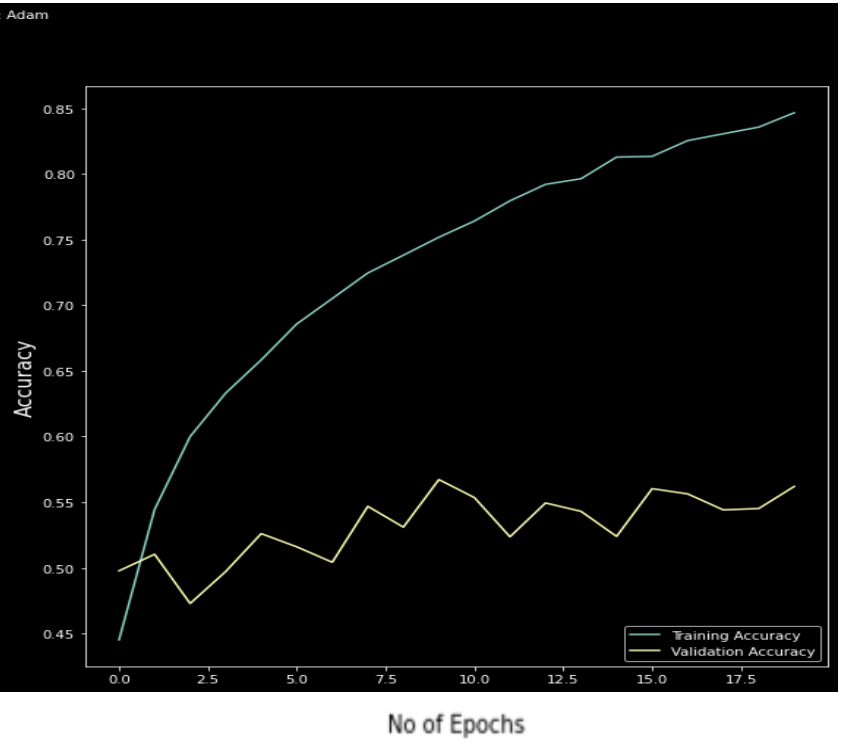
VGG16 Model Evaluation

Categorical Crossentropy



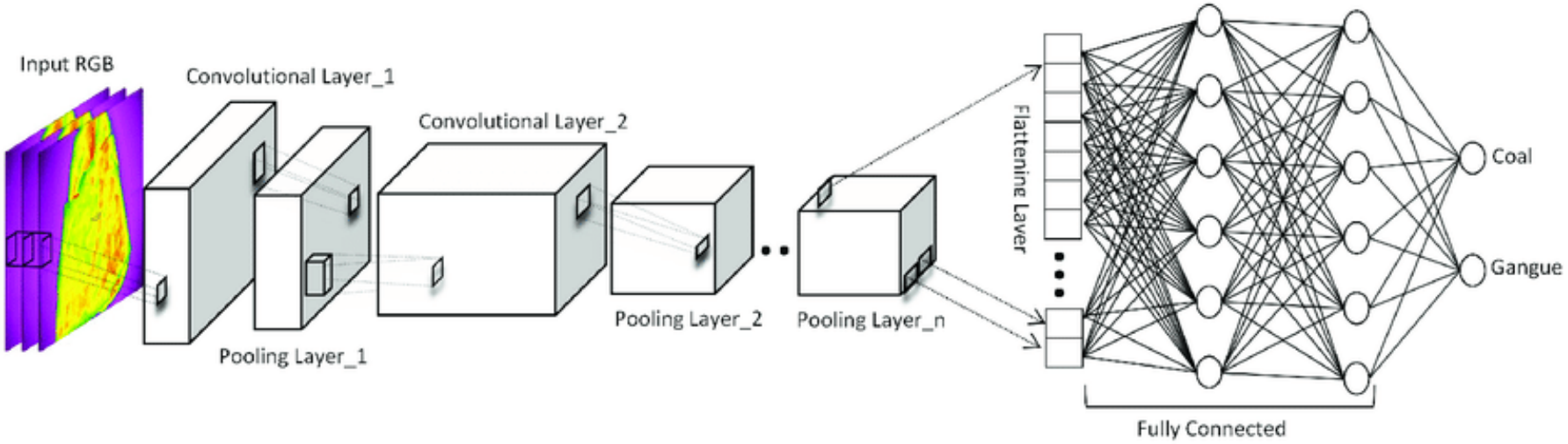
Training Loss is nearby 0.5
Validation Loss is nearby 1.5

Accuracy



Training Accuracy is nearby 85%
Test Accuracy is nearby 55 %

CNN Model


Input RGB
Convolutional Layer_1
Convolutional Layer_2
Pooling Layer_1
Pooling Layer_2
Pooling Layer_n
Flattening Layer
Coal
Gangue
Fully Connected

CNN Model

- We define our CNN with the following global architecture:
- 17 convolutional layers
- 2 fully connected layers
- Basic CNN architecture details:
- Input layer - Input layer in CNN should contain image data Convo layer - Convo layer is sometimes called feature extractor layer because features of the image are get extracted within this layer Pooling layer - Pooling is used to reduce the dimensionality of each features while retaining the most important information. It is used between two convolution layer Fully CL - Fully connected layer involves weights, biases, and neurons. It connects neurons in one layer to neurons in another layer. It is used to classify images between different category by training and placed before the output layer Output Layer - Output layer contains the label which is in the form of one-hot encoded Also we use some common techniques for each layer
- Batch normalization: improves the performance and stability of NNs by providing inputs with zero mean and unit variance. Dropout: reduces overfitting by randomly not updating the weights of some nodes. This helps prevent the NN from relying on one node in the layer too much.

Modeling Steps

- Input Layer - 48,48,1
- Layer 1&2 - 3*3,Conv,64
- Layer 3,4&5 - 3*3,Conv,128
- Layer 6 to 13 - 3*3,Conv,256
- Layer 14 to 17 - 3*3,Conv,512
- Flatten layer
- FC - 256 units
- FC - 512 units
- FC - 7 units

- **Activation Function - ReLu, Softmax**
- **Epoch - 100**
- **Optimizer - Adam**
- **Batch size -32**
- **Callbacks- EarlyStopping, ReduceLROnPlateau**

- Loss and accuracy plots
- Heatmap of confusion matrix

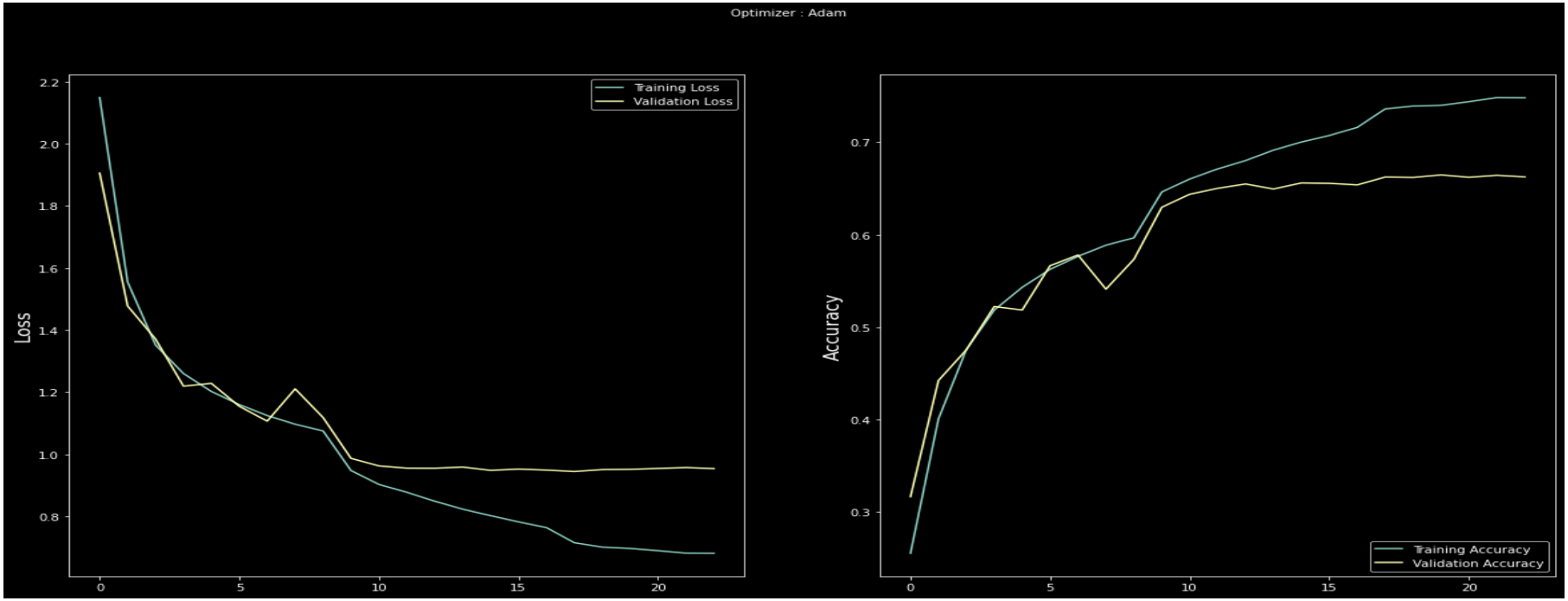
Also we use some common techniques for each layer

- **Batch normalization**
- **Dropout**

Model Evaluation

Categorical Crossentropy

Accuracy

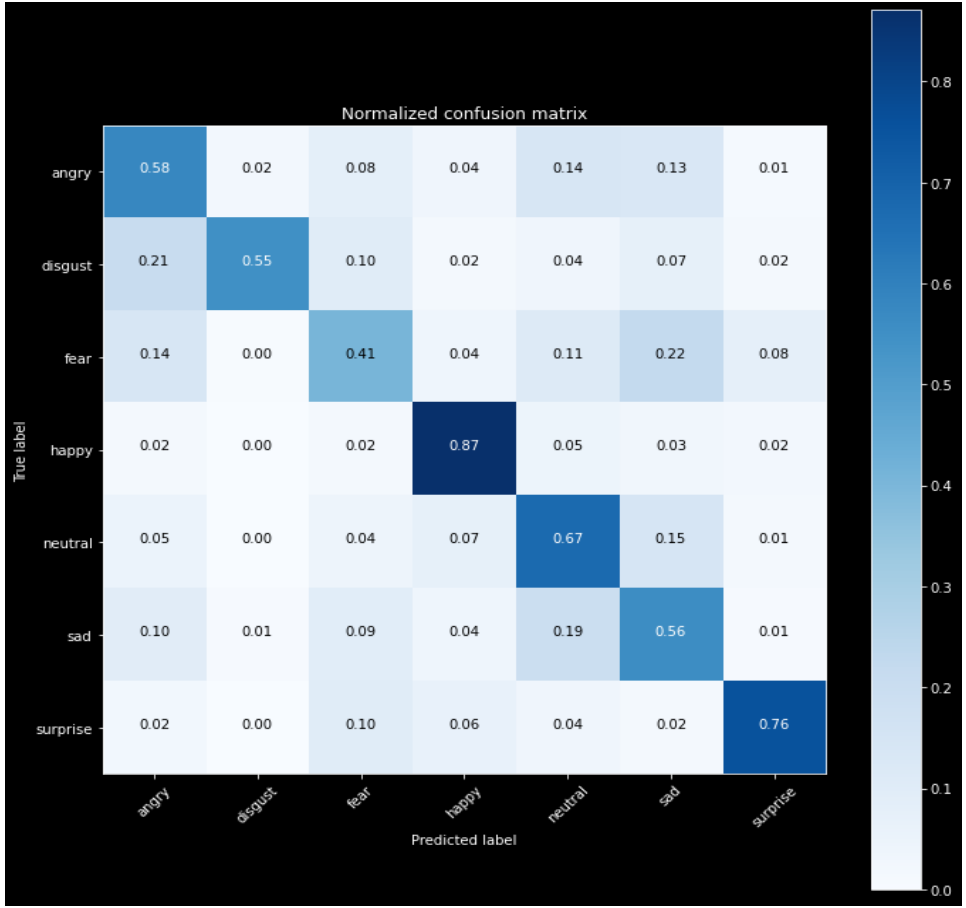


Training Loss is nearby 0.5
Validation Loss is nearby 1

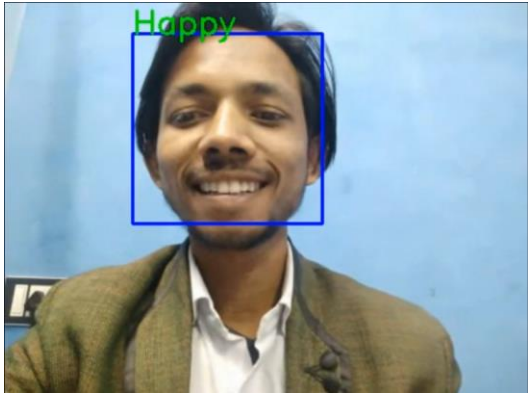
Training Accuracy is nearby 76%
Test Accuracy is nearby 66%

Confusion matrix (Heatmap)

- It is doing a great job in identifying happy, neutral and surprised facial expressions but its getting confused between angry and disgust.it is also performing poorly on recognizing feared emotion as it is getting confused with sad expression



Real Time Face Emotion Detection



Deployment

Creating Web App Using Streamlit

Streamlit is an open-source python framework for building web apps for Machine Learning and Data Science. We can instantly develop web apps and deploy them easily using Streamlit. Streamlit allows you to write an app the same way you write a python code. Streamlit makes it seamless to work on the interactive loop of coding and viewing results in the web app.

https://share.streamlit.io/armanalam6342/face-emotion-recognition/app.py

Deployment

Deployment in cloud platform

Heroku is a container-based cloud Platform as a Service (PaaS) supporting several programming languages as Java, Node.js, Scala, Python, PHP, and Go.

https://face-exp-recognizer.herokuapp.com/

Challenges

- Large image dataset to handle
- Google Collab GPU working Slowly that's why I trained my models on kaggle
- Selecting No. of filters and neurons
- Selecting batch size to avoid crashing of the system
- In Deployment File uploading limit size on heroku is 500 mb and my model size was 800+ mb thus I faced a lot of issues in deploying my model

Conclusion

- The CNN model gave us training accuracy of 76 % and validation accuracy of 66 %.
- It is doing a great job in identifying happy, neutral, sad and surprised facial expressions but its getting confused between angry and disgust.it is also performing poorly on recognizing feared emotion.
- A front-end model was successfully created using Streamlit and run on a local webserver.
- Successfully deployed Streamlit web app on Heroku and streamlit share that runs on a web server.

Thank You