Real Time Speech Driven Facial Emotions with Expression

Sumit Rai[1], Shivam Tripathi[2], Sahil Raza[3], Shubham Singh[4], Ms. Neetu Bansla[5]
[1,2,3,4]*Department of Computer Science and Engineering, Raj Kumar Goel Institute of Technology, Ghaziabad, Uttar Pradesh, India*
[5]*Faculty, Department of Computer Science and Engineering, Raj Kumar Goel Institute of Technology Dr. A.P.J. Abdul Kalam Technical University, India*

Abstract **- Because of the heterogeneity present across human faces, recognising a human's facial expression with a computer is a difficult undertaking. This variability encompasses expression, color, position, and orientation. The purpose of this study is to demonstrate how a Convolution Neural Network (CNN) architecture may be utilised to detect facial expressions in real time. The FER 2013 Facial Expression Recognition Challenge dataset was employed in this study, and our neural network was trained to categorise emotion states using it. For the classification of seven different types of emotions using facial expressions, we attained an accuracy of 77.16 percent and a validation accuracy of 57.41 percent in this study.**

Index Terms **- facial expression recognition; human emotion detection; naturalistic expression; recognition of emotional facial expressions; convolutional neural network, image processing, face detection.**

INTRODUCTION

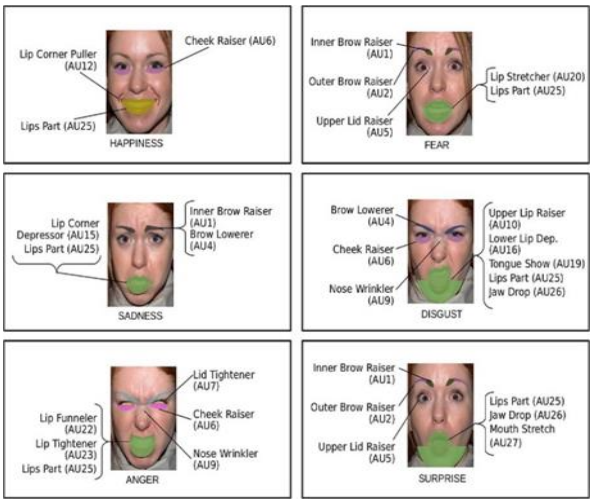
Humans can communicate in two ways verbal communication and non-verbal communication. With the help of their facial expressions humans express their emotions. Surprise, anger, disgust, happiness, fear and sadness are basic expression categories and commonly used as expression labels. If we look at the importance of facial expression in any message that is delivered by the person then we found that message contains 7% effect of verbal part, 38% effect by vocal part and 55% of the message effect is contributed by the facial expressions so we get the point here how important is the facial expressions in any communication. So detecting the emotions of facial expressions of humans while communicating can play a very important role in understanding the true sense of speech. It has various applications like it can be used by any business organization in conducting surveys and understand humans interest in surveys, can be used in camera applications, pain monitoring for patients etc.

Various studies in the field of picture categorization have been conducted using deep learning methods. Convolutional Neural Network (CNN) is a deep learning technique for image categorization, recognition, and segmentation. CNN takes image as an input and then assign weights and biases to various aspects in the image and with the help of this differentiate one image from other. The input layer, convolution layer, dense layer, and output layer are the four layers that make up a CNN. CNN used for "artificial neural network," and it can be used to classify and recognise images. For better accuracy of results we have to feed more and more data to Convolutional Neural Network.

CATEGORIZATION OF FACIAL EXPRESSIONS AND ITS FEATURES

Facial Expressions plays a very important role in detecting emotions of a person. Although humans makes a plenty of emotions based on different physical and mental circumstances but all these emotions can be classified into 7 basic emotions. We are using happiness, fear, sadness, surprise, anger, disgust and neutral as universal emotions to detect the humans emotions in real time. We will understand all these emotions closely that motion of which facial part contribute more in each of these expressions and then try to find the human emotions using our Convolutional Neural Network model.

Fig: Facial Expression Features



PROPOSED METHODOLOGY

FER Dataset:

The Fer2013 dataset is a Kaggle-based face recognition competition that was launched at ICML 2013. Face images at a resolution of 48x48 pixels make up the data. The faces are automatically registered so that they are more or less centred in each shot and take up roughly the same amount of area. Based on the emotion displayed in the facial expression, we'd like to classify each face to one of seven groups(0=Angry, 1=Disgust, 2=Fear, 3=Happy, 4=Sad, 5=Surprise, 6=Neutral).Our training set consists of 28,709 images , test set consists of 3859 images and 3859 images used for validation purpose. Pierre-Luc Carrier and Aaron Courville developed the Fer2013 dataset as part of their ongoing scientific research.

Fig: FER Dataset Images

LIBRARIES USED

In recognizing and detecting the facial expressions of humans the libraries used by us are OpenCV, Keras and Tensorflow. Keras and Tensorflow libraries are used in this study to train our proposed CNN architecture and with the help of our deep learning model we will recognize facial expressions.

OpenCV is a open source library and widely used in processing real time images. By using it images can be processed to identify objects and faces.

Keras is one of the libraries that is used to code deep learning models and it uses tensorflow in its backend. Keras acts as an interface for tensorflow library and with the help of this neural network models can be trained and defined in just few lines of code.

CNN ARCHITECTURE

A CNN's architecture is divided into two parts:

1. 1.Feature extraction is a convolutional tool for analysing images that distinguishes and detects the different parts of the image.
2. A connected layer who use the output from the convolution process and predict the behavior of the image based on feature extraction done in CNN process.

There are 3 layers present in CNN architecture:

1. CONVOLUTIONAL LAYER:

This is the initial layer of the CNN architecture, and it is responsible for extracting different features from the input image. Convolution takes place between the user's input image and a M*M-sized filter. Now we perform a dot product between the image and filter with respect to the size of filter which is M*M. The result of dot product or output is termed as FEATURE MAP.

2. POOLING LAYER:

The major purpose of this layer is to shrink the feature map in order to reduce computing expenses.

3. FULLY CONNECTED LAYER:

Weights and biases make up this layer. The main purpose of this layer is used to join the neurons of two different layers. It's frequently inserted in front of the output layer. Activation function is one of the important parameters of CNN model.in our CNN model we use the activation function named RELU. In our proposed CNN architecture we used the

convolutional layer of size 3*3 which are compound together followed by pooling layer whose size was 2*2. Now RELU activation function is applied in the hidden layers of network. RELU introduces nonlinearity in our CNN model. Each hidden layer was followed by a drop out layer and the value has been fix to 0.25.It means to avoid overfitting it randomly deactivate 25% of the nodes from hidden layers. At the end only our output layer will exist consisting different expressions of our model.

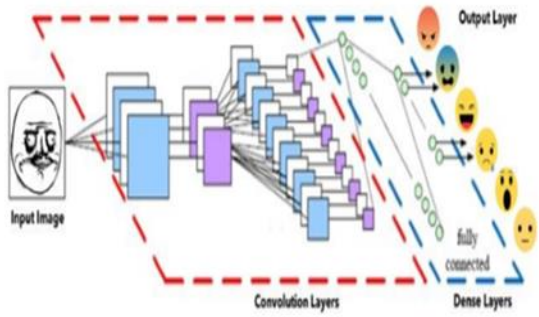
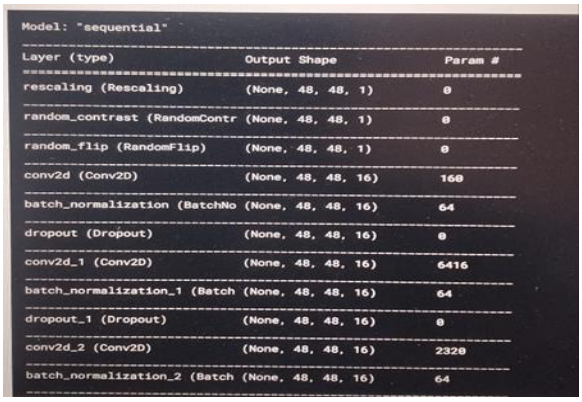
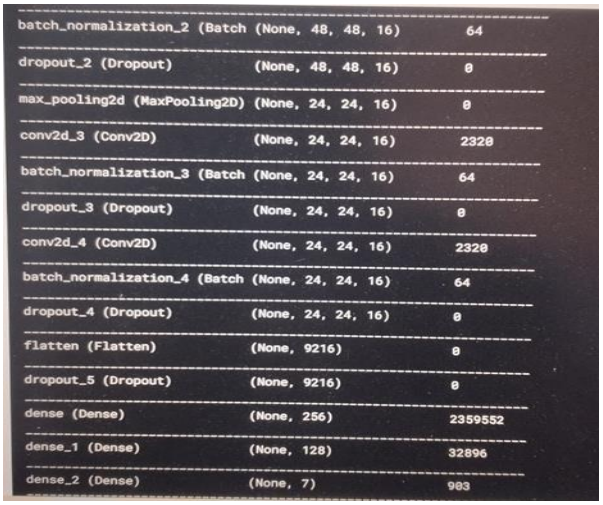


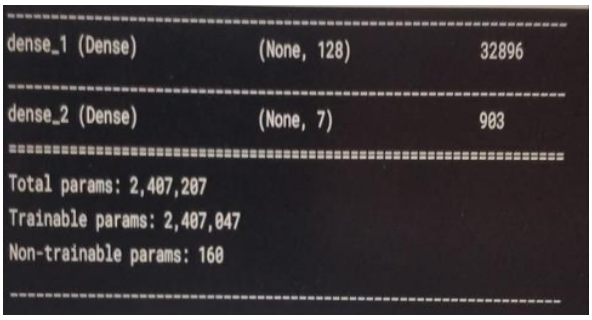
Fig: Convolution Neural Network Architecture



Layer (type)	Output Shape	Param #
rescaling (Rescaling)	(None, 48, 48, 1)	0
random_contrast (RandomContr	(None, 48, 48, 1)	0
random_flip (RandomFlip)	(None, 48, 48, 1)	0
conv2d (Conv2D)	(None, 48, 48, 16)	160
batch_normalization (BatchNo	(None, 48, 48, 16)	64
dropout (Dropout)	(None, 48, 48, 16)	0
conv2d_1 (Conv2D)	(None, 48, 48, 16)	6416
batch_normalization_1 (Batch	(None, 48, 48, 16)	64
dropout_1 (Dropout)	(None, 48, 48, 16)	0
conv2d_2 (Conv2D)	(None, 48, 48, 16)	2320
batch_normalization_2 (Batch	(None, 48, 48, 16)	64



batch_normalization_2 (Batch	(None, 48, 48, 16)	64
dropout_2 (Dropout)	(None, 48, 48, 16)	0
max_pooling2d (MaxPooling2D)	(None, 24, 24, 16)	0
conv2d_3 (Conv2D)	(None, 24, 24, 16)	2320
batch_normalization_3 (Batch	(None, 24, 24, 16)	64
dropout_3 (Dropout)	(None, 24, 24, 16)	0
conv2d_4 (Conv2D)	(None, 24, 24, 16)	2320
batch_normalization_4 (Batch	(None, 24, 24, 16)	64
dropout_4 (Dropout)	(None, 24, 24, 16)	0
flatten (Flatten)	(None, 9216)	0
dropout_5 (Dropout)	(None, 9216)	0
dense (Dense)	(None, 256)	2359552
dense_1 (Dense)	(None, 128)	32896
dense_2 (Dense)	(None, 7)	903



dense_1 (Dense)	(None, 128)	32896
dense_2 (Dense)	(None, 7)	903

Total params: 2,407,207
Trainable params: 2,407,047
Non-trainable params: 160

Fig: CNN Architecture Summary

NEURAL NETWORK TRAINING

In preprocessing of data we have splitted the datasets into 3 parts train, validation and test and then converted it into string of list of integers.

```
train shape: (28709, 3),
validation shape: (3589, 3),
test shape: (3589, 3)
```

We can see the distribution of datasets on train, validation and test from the below bar graph and on the top of their x- axis bar each class of these datasets are written.

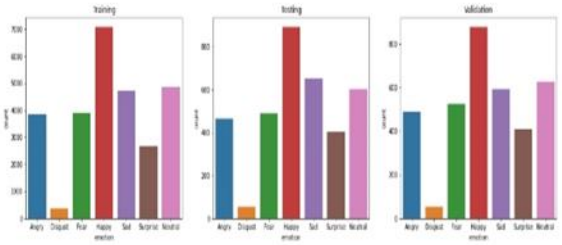


Fig: Bar-plot class distribution of train, value and test

REAL TIME TESTING

The trained model is tested with real-time data once our proposed CNN architecture has been trained. The system prepossesses real-time photos as soon as they are supplied to it. It means that whenever the system receives a picture of any size, it converts it to a 48*48 sized image. The Haar Cascade Classifier is used by the model to recognise the faces in the image. A Haar classifier, often called a Haar cascade classifier, is a machine learning object recognition system that can identify objects in photos and movies. After then, it's mostly the region of interest that's cropped. The

system turns the RGB image, which has three colours red, green, and blue, to a grey image, which contains only one colour, because the model was trained on grayscale images. The machine will next apply image normalisation to the image to make the categorization task easier. The data will next be classified using our proposed Convolutional Neural specification.

RESULT

In this study, we have used keras and tensorflow libraries to train our CNN architecture model. CNN based architecture gives the best result as they can learn and select features automatically and give good results.

With the help of the ConvLayer specific features can learn from new dataset so specific features play an important role in improving accuracy. The more data we provide to our deep learning model we will get more accuracy.

When we have tested our model in predicting the human emotions then we found that we have achieved training accuracy of 77.16% and validation accuracy of 57.41 %.

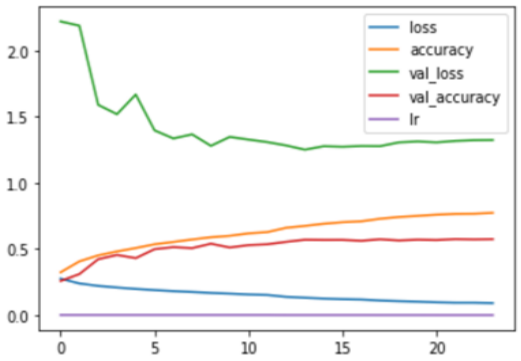


Fig: Performance metrics of proposed architecture

CONCLUSION

This paper gives an effective method for finding the seven basic emotions of human in real time using facial expressions. In this study, we have used CNN to train our model and achieve high accuracy. Now our model is able to predict human emotions but this topic can be further researched and studied and using different algorithms and image processing techniques more accurate models can be built. Although this model gives good result with commendable accuracy but some improvements can be made in some areas like we can add more data in each class to get more accurate results as we know more data given to deep learning models more accuracy is obtained. The area of emotion recognition is gaining popularity in various domains like software engineering, gaming and education. This emotion detection can be used by medical teams for better therapy of patients by understanding their emotions and it can also be used in understanding the consumer reviews in hotels, restaurant and it can also be used in various other fields. This paper presented a detailed approach and technique for detecting human emotions in real time although it can further be improved using different techniques.

REFERENCES

[1] SHAN LI AND WEILONG DENG, “DEEP FACIAL EXPRESSION RECOGNITION: A SURVEY”, ARXIV:1804.08348V2 [CS.CV] 22 OCT 2018.

[2] DEEPESH LEKHAK, “FACIAL EXPRESSION RECOGNITION SYSTEM USING CONVOLUTIONAL NEURAL NETWORK”, TRIBHUWAN UNIVERSITY INSTITUTE OF ENGINEERING.

[3] MAJA PANTIC AND LEON J.M. ROTHKRANTZ“AUTOMATIC ANALYSIS OF FACIAL EXPRESSIONS: THE STATE OF ART”.

[4] TAWSIN UDDIN AHMED, MAHAMMAD NEWAJ JAMIL, MOHAMMAD SHAHADAT HOSSAIN, KARL ANDERSON AND MOHAMMED SAZZAD HOSSAIN “AN INTEGRATED REAL-TIME DEEP LEARNING AND BELIEF RULE BASED INTELLIGENT SYSTEM TO ASSES FACIAL EXPRESSION UNDER UNCERTAINITY”.

[5] DANIEL LLATAS SPIERS, “FACIAL EMOTION DETECTION USING DEEP LEARNING”.

[6] B. HASANI AND M. H. MAHOOR, “FACIAL EXPRESSION RECOGNITION USING ENHANCED DEEP 3D CONVOLUTIONAL NEURAL NETWORKS,” IN PROCEEDINGS OF CVPRW. IEEE, 2017.

[7] TIANYUAN CHANG, GUIHUA WEN, YANG HU, JIAJIONG MA “FACIAL EXPRESSION RECOGNITION BASED ON COMPLEXITY PERCEPTION CLASSIFICATION”.

[8] AYAN BAG "REAL TIME FACIAL EXPRESSION RECOGNITION USING CONVOLUTION NEURAL NETWORK ALGORITHM".
[9] "ROBUST REAL-TIME FACE DETECTION", INTERNATIONAL JOURNAL OF COMPUTER VISION 57(2), 137–154, 2004
[10] D Y LILIANA "EMOTION RECOGNITION FROM FACIAL EXPRESSION USING DEEP CONVOLUTIONAL NEURAL NETWORK".
[11] N. CHRISTOU AND N. KANOJIYA "HUMAN FACIAL EXPRESSION WITH CONVOLUTION NEURAL NETWORK."
[12] Y. LECUN, K. KAVUKCUOGLU AND C. FARABET "CONVOLUTIONAL NETWORK AND APLLICATIONS IN VISION".
[13] Y. MOSES, D. REYNARD AND A. BLAKE "DETERMINING FACIAL EXPRESSIONS IN REAL TIME".
[14] H. YAMADA "VISUAL INFORMATION FOR CATEGORIZING FACIAL EXPRESSIONS OF EMOTIONS".