International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 04 Issue: 07 | July -2020 ISSN: 2582-3930

Facial Emotion Detection using Convolution Neural Network

Mahesh U [1], Nikhil K [2], Rajshekar P [3], Ramchandra Murthy C H [4], Varalatchoumy M [5]

[1],[2],[3],[4] Computer Science, Cambridge Institute of Technology, Bengaluru, Karnataka, India.
[5] Assistant Professor, Computer Science dept., Cambridge Institute of Technology, Bengaluru, Karnataka,India.

Abstract– This paper discusses about the application of feature extraction of facial expressions with combination of neural network for recognition of seven different facial emotions happy, sad, neutral, angry, surprised, scared, disgust. Humans are able to produce thousands of facial actions and emotions during communication that vary in complexity, intensity and meaning. Haar cascade is used in this project and it achieved 60% accuracy. Haar cascade method is used to detect an input image and extract features such as mouth, eyes, ears etc. By neural network training 7 different emotion categories areobtained.

Keywords: Neural network, interface, Haar cascade, disgust

1.INTRODUCTION

A Facial emotion expression is the way to represent the action of the mind affective state, intention, cognitiveactivity personality and psychopathology of a person and plays a communicative role in interpersonal relations. It has been studied for a long period of time and obtaining the progress recent decades. Though progress has been made, recognizing facial expression with a high accuracy rate remains to be difficult due to various complexity and facial expressions. Generally human beings can convey intentions and emotions through nonverbal ways such as gestures, facial expressions and involuntary languages. This system can be significantly useful, nonverbal way for people to communicate with each other. The man factor is how efficiently the system detects or extracts the facial expression from image. The system is gainingmore recognition due to widely used in many fields like lie detector, medical assessment and human computer interface.

In this paper, our approach is based on Convolution Neural Network(CNN) for facial expression recognition in which the input image is in pixel format differentiated by the emotion labeled as (i) anger, (ii) happiness, (iii) fear , (iv) sadness, (v) disgust , (vi) surprise and (vii) neutral is classified for the Model of Facial Emotion Detection and thus the Model is trained and that model is Executed and Real Time Video Feed is given with the Image pause of 5 seconds and thus the CNN model is able to Detect the Emotion of the Real Time Multiple faces and provides result and also with the GraphicalChart.

2. PROPOSEDSYSTEM

The process of detecting the emotion involves series of steps. These steps are briefly explained below**.**

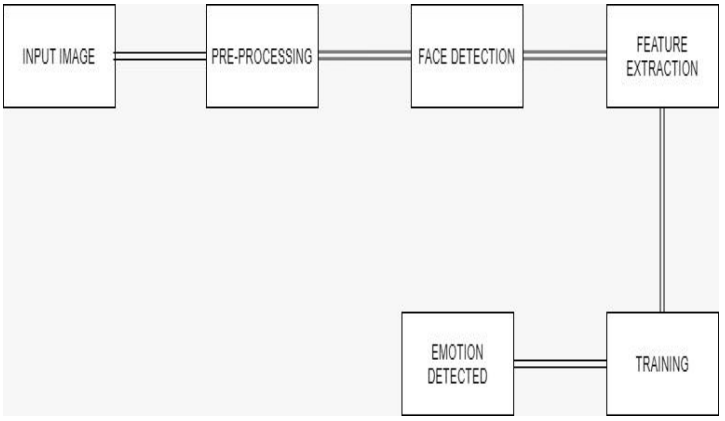

Fig.1 Proposed System

2.1INPUTIMAGE

There are two types of images used in the project real-time images and images from database.

The real time images are used for emotion detection and images in the database are used fortraining.

2.2 PRE-PROCESSING

Pre-processing is a technique that processes its input data to produce output that is used as input to another program.

Advantage of pre-processing is to improve the image data (features) by suppressing unwanted distortions and/or enhancement of image features and also crop the images to the required size so that all output can be of same height and width with the image without being disturbed and change from color image to Grey scale image for better results.

2.3 FACEDETECTION

Local Binary Patterns Histograms (LBPH) algorithm is used to detect face. Once the image has been preprocessed thealgorithm detects the face from the preprocessed output mage image and divides it into blocks and histogram for each block is calculated. Finally blocks are combined into a single histogram and face is detected in thisstage.

2.4 FEATUREEXTRACTION

Feature extraction is used to obtain required shape information data which is shown on the person face in a pattern so that the classifying the pattern or shapes is made easy by a simple procedure. It is achieved by using Haar cascade algorithm. The algorithm needs a lot ofdata of images of faces and images without faces to train and test the classifier. Then features should be extracted. First step is to collect the Haar Features. A Haar feature considers rectangular regions at a specific location in a window which is detected,and adds all the pixels intensities in each region in face and calculates the difference between these sums. AdaBoost which both selects the best features and trains the classifiers that use them. Haar Cascade classifier detects features such as eyebrows, lips, nose and eye.

2.5 TRAINING

Image training refers to the process or training an algorithm or process based on the images in a training database.
Once the image has undergone the extraction process, where the CNN is trained using several thousands of images which is labeled for differentemotions such as sad (0), happy (1), neutral (2), angry (3), disgust (4), surprised (5), sad(6) are trained using Google Collaboratory and the trained program in obtained.

2.6 EMOTIONDETECTED

After the image has undergone comparison, and if the features of the real time image and the image present in the database matches then it can be said that out of 7 facial states listed the person is either happy, sad, surprised, neutral, angry and disgust.

3.SYSTEM DESIGN

3.1 TheDatabase

The dataset that is used for training the model is from a Kaggle Facial Expression Recognition Challenge a few years back (FER2013). It comprises a total of 35000+ cropped, 48-by-48-pixel grayscale images of faces each labeled with one of the
Seven different emotion classes: happy, anger, disgust, fear, sadness, surprise, and neutral.

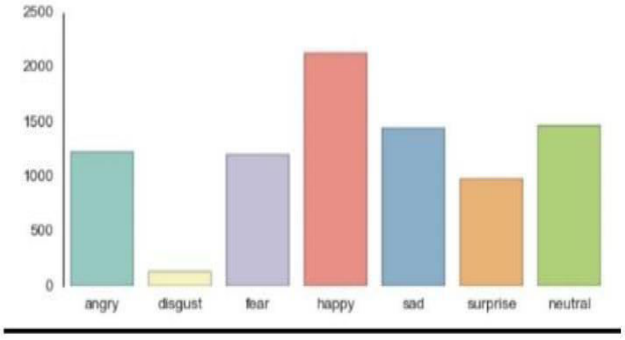

Fig.2 Graph of 7 emotion classes

28709 labeled faces were used as the training set and held out the remaining two test sets (3500/set) for after-training validation. The resulting is a 7-class, balanced dataset, that contains angry, disgust, fear, happy, sad, surprise, and neutral. Now it is ready to train.

3.1 TheModel

Deep learning is a popular technique used in computer vision. Convolutional Neural Network (CNN) had been chosen because layers are building blocks to create our model architecture. CNNs are known to imitate how the human brain works when analyzing visuals. A typical architecture of a convolutional neural network contains an input layer, some convolutional layers, some dense, and an output layer. These are linearly stacked layers ordered in sequence. In Keras, the model is created as Sequential () and morelayers
are added to build architecture.

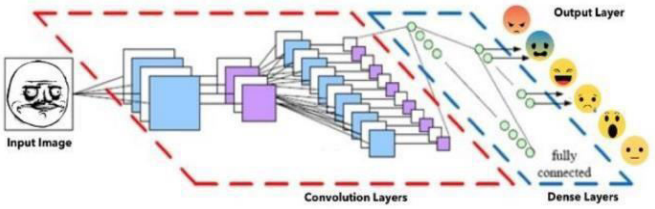

Fig.3 System design

3.2 Input Layer

The input layer is already determined and also has fixed dimensions, so the image must be pre-processed before it can be fed into the layer. OpenCV was used as a computer vision library, and for facedetection parameter in the image. The haarcascade_frontalface_default.xml in OpenCV contains pre-trained filters and finds the cropped face. The cropped face from and image is then converted from color into grayscale using cv2.cvtColor and resized to 48-by-48 pixels using cv2.resize. This step reduces the dimensions as compared to the original RGB format with three color dimensions (3, 48, 48). The pipeliningmakes sure that every image can be fed into the input layer as a (1, 48, 48) NumPy array.

3.3 ConvolutionalLayers

The NumPy array gets passed into the Convolution2D layer where we specify the number of filters as one of the hyperparameters. The set of filters are unique with randomly generated weights. Each filter, (3,3) receptive field of features, slips across the original image with shared weights to create a map of feature that exits. Convolution creates the feature maps that represent how pixel values are enhanced, for example, edge and pattern detection. A feature map is created by adding 1 filter at a time across the completeimage. Whereasother filters are applied one by one after each filter creating a set of featuremaps.
Pooling is a technique used for dimension reduction and is applied after one or several convolutional layers. It is a very important step while building Convolution Neural Network as adding more convolutional layers can greatly affect computational and execution time.A popular pooling method called MaxPooling2D that uses (2,2) windows across the feature map only keeping the maximum pixel value was used. The downscaling of pixelsforms an image with dimensions which is reduced by4.

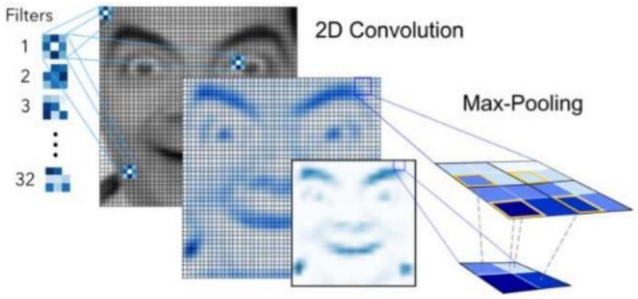

Fig.4 max pooling layers

3.4 Dense Layers

The dense layer is inspired by the way neurons transmit signals through the brain. It takes many input features and transform features through layers connected with trainable weights.

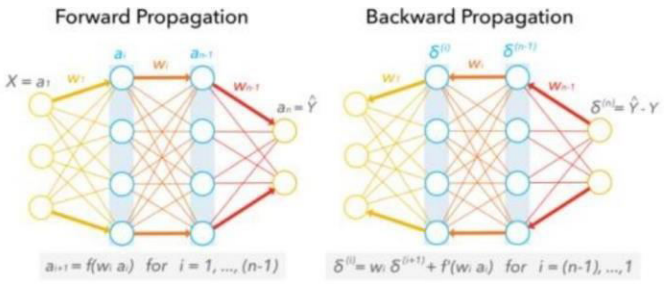

Fig.5 Represent forward and back propagation

Theweights are trained by forward propagation of training data then backward propagation of its errors. Back propagation proceeds fromassessing the difference between prediction and true value, and back calculates the weight adjustment needed for every previouslayer. The speed of the training of the data and the complexity of the model can be controlled byadjusting the parameters, such as learning rate and network density. As we feed in more data, the network can gradually increase accuracy and adjust data until errors are minimized. The more the layers added to the network the better it can pick up signals. As training increases the model also becomes vulnerable for overfitting of the training data. One of themethods to prevent overfitting and generalizing of hidden data is to apply dropout. Dropout can randomly select a portion of nodes which are less than 50% to set their weights to zero during training. This process can successfully control the model's sensitivity to noise during training and maintaining the necessary complexity of the architecture of the model.

3.5 Output Layer

This output represent itself as a probability for each emotion

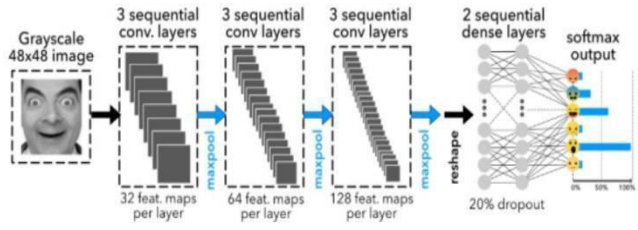

Fig.6 Represent sequential convolution layers

class. Therefore, the model can show the detail probability composition of the emotions in the face. Facial expressions are usually much complex and contain a mix of emotions that could be used to accurately describe anexpression. Building a simple CNN with an input, 3 convolution layers, one dense layer, and an output layer to begin with. As a result, the simple model performed poorly. It showed a very low accuracy where it could only guess the facial emotions. The structure failed to detect the precise details in facial expressions. Which means provided the complexity of the facial emotion detection it was very crucial to build a deep machine learning is needed in order to identify the subtle parts and features present on the face to be more precise and accurate in which the overall rate of accuracy is increased. Having handled combinations of three components to extend the models complexity: Models with various combinations were trained and evaluated using GPU computing g2.2xlarge on Google Collaboratory. This substantially reduced training time and increased efficiency in tuning the model. In the end, our final net architecture was 9 layers deep in convolution with one max-pooling after every three convolution layers as seen previously.

4 IMPLEMENTATION

4.1 Algorithm

Step 1: Collection of a data set of images. (In this case we are using FER2013 database of 35887 pre-cropped, 48-by-48-pixel grayscale images of faces each labeled with one of the 7 emotion classes: anger, disgust, fear, happiness, sadness, surprise, and neutral.
Step 2: Pre-processing of images.
Step 3: Detecting of a face from each pre-processed image.
Step 4: The cropped face is converted into grayscale images.
Step 5: The pipeline ensures that every image can be fed into the input layer as NumPy array of dimension (1,48,48)
Step 6: The NumPy array gets passed into the Convolution2D layer.
Step 7: Convolution generates feature maps.
Step 8: Pooling method called MaxPooling2D that uses (2, 2) windows across the feature map only keeping the maximum pixel value.
Step 9: During training, Neural network Forward propagation and Backward propagation performed on the pixelvalues.
The model can provide the detailed probability composition of the emotion expressions in the face.

5.RELATED WORK

In recent years many researchers have come up with a Emotion detection where it might achieve the goal with pros and cons in it. In Facial expression recognition using fuzzy logic [1] is one useful approach for Fuzzy classification, which can determine the intrinsic division in a set of unlabeled data and representatives for homogeneous groups. Where it uses Bezier's curve, but Accuracy of this system is very low. Whereas in "Facial expression recognition using adaptive robust local complete pattern" [2] ARLCP effectively encodes significant information of emotion related featuresby

using the sign, scale, magnitude and directional information of the response that is sturdier to noise and illumination variation. Which divides into sign magnitude direction histograms which can be and effective method but is difficult to train the data or create the model. "Facial expression recognition with convolutional neural networks"[3] is a process for facial emotion expression recognition systems applying standard machine learning model to identify and extract image features, and these methods generalize poorly hidden data. This gives optimal solution at low resolution. Which can be overcome by training lot of data which is organized based on emotion and it was for a single face whereas we have overcome it by Multiple Face emotion detection and represented via graphical chart.

6. RESULT

Real-time image feeds are provided with the help of the external or an inbuilt camera where the feed is provided as show in the Fig 7. Once the face is detected it identifies the type of emotion present and displays the expression on the face and that emotion graph is plot in the Fig 8 where is the shows the percentage and probability of that particular expression is shown in it and Fig 5.3 shows the probability of the emotion and the image is given pause at every 5 seconds so it doesn't keep changing everysecond.

It has achieved more accuracy compared to other methods. Fig 9 shows the probability of the emotion and the image is given pause at every 5 seconds, so it doesn't keep changing every second.It has achieved more accuracy compared to othermethods.

Fig.7 Represent graph of neutral emotion

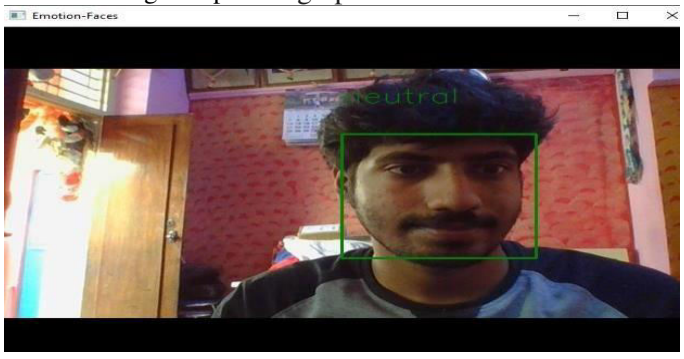

Fig 8. Represent the image of neutral emotion

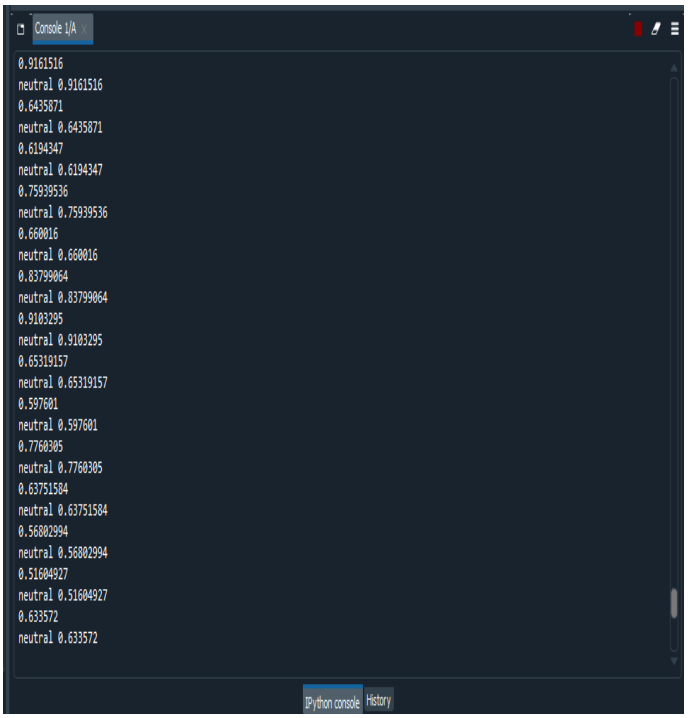

Fig 9. Represent emotion reading based on a facial expression

The accuracy in recognition of emotion depends on feature extraction method. HAAR cascade is used to detect facial features such as nose, eyes, ears etc.

The accuracy score of 63% is achieved in this work which is more than predicted score of previous work [4] which was about 48% accuracy was achieved.

To avoid the continues change in real time video, feed the image and pause for 5 seconds of delay to obtain that certain emotion at that instance

CONCLUSION

This paper provides a proposed model to solve problem of facial emotion detection using neural network. The accuracy in the recognition of emotion depends on the feature extraction method. HAAR cascade classifier is used to detect facial features such as nose, ears, eyebrows and eyes. A accuracy score of 63% is achieved in this project which is more than predicted (48%). The project is further implemented to identify emotions of multiple faces.

REFERENCES

[1] RojaGhasemi and Maryam Ahmady "Facial expression recognition using fuzzy logic"2014.

[2] Al Shahriar Rubel,Adib Ahsan Chowdhury "Facial expression recognition using adaptive robust local complete pattern"2019

[3] Arushi Raghuvanshi, Vivek Choksi, "Facial recognition with Convolutional neuralNetwork." 2019

[4] AI shahriarRubel,Adibahsan Chowdary."Facial recognition using adaptive robust local competepattern".

[5] XueMei Zhao, ChengBingWei. "Real-time facial recognition based on LBPH algorithm.",

[6] Nazil Praveen, kesari Verma and S. Gupta, "Facial recognition using gini Index andcharacteristics",2012

[7] D. C. Ali Mollahosseini and M. H. Mahoor. Going deeper in facial expression recognition using deep neural networks. IEEE Winter Conference on Applications of Computer Vision, 2016.

[8]] S.-Y. D. Bo-Kyeong Kim, JihyeonRoh and S.-Y. Lee. Hierarchical committee of deep convolutional neural networks for robust facial expression recognition. *Journal on Multimodal User Interfaces*, pages 1–17,2015

[9] P. Ekman and W. V. Friesen. Emotional facial action coding system. Unpublished manuscript, University of California at San Francisco,1983.

[10] F. Chollet. Keras. https://github.com/fchollet/Keras,2015.

[11] Y. Jia, E. Shelhamer, J. Donahue, S. Karayev, J. Long, R. Girshick, S. Guadarrama, and T. Darrell. Caffe: "Convolutional architecture for fast feature embedding". 2014.