International Journal of Computer Science and Engineering (IJCSE);
ISSN Print: 2278-9960;ISSN Online:2278-9979;
Vol. 8, Issue 3, Apr - May 2019; 7-14

International Academy of Science, Engineering and Technology
Connecting Researchers; Nurturing Innovations

FACIAL EMOTION DETECTION USING CONVOLUTIONAL NEURAL NETWORKS

Salakapuri Rakesh[1], Avinassh Bharadhwaj[2] & E Sree Harsha[3]

Research Scholar, Department of Information Technology, Chaithanya Bharathi Institute of Technology, Hyderabad, India

ABSTRACT

Human emotions are different mental states of feelings that arise naturally rather than through conscious attempt and are followed by physiological alters in facial muscles which imply different expressions on the face. Some of the emotions are a surprise, sad, fear, anger, happy, etc. Emotion gives us a clue about the state of a person and enables to make conversation with the other person based on their mood. Facial expression plays an important role in non-verbal communication between people. A lot of research work has been accomplished to detect human emotions. But still, it is far behind from the human vision system. In this paper, we are proposing an algorithm which trains the FER 2013 dataset and builds a model. This model is used to predict human emotions using deep CNN (Convolution Neural Networks).

KEYWORDS: *Convolution Neural Network (CNN), Face Detection, Facial Emotions, Facial Expressions*

Article History

Received: 16 Apr 2019 | Revised: 23 Apr 2019 | Accepted: 06 May 2019

INTRODUCTION

Emotions and related changes in the facial muscles are jointly known as facial expressions [1]. Facial expression gives us a clue about the state or mood of a person and enables to make conversation with the other person based on their mood. Furthermore, facial expressions also give support to judge the existing emotion and mood of a person [2]. Facial expression plays a key role in non-verbal communication between people. Various classification of facial expressions might be used in many applications like Medical Rehabilitation [3], Human Behavior Predictor [4] and Surveillance System[5]. Human emotions are majorly classified into 7 categories named as happiness, sadness, anger, fear, surprise, disgust and neutral [6]. Numerous research scholars have used diverse methods [7][8][9] for classifying facial expressions but among all convolution neural networks method given better results. So in our proposed algorithm, we are using deep CNN (Convolution Neural Networks) to build the model and to predict the human emotion from the given input image or video.

The main objective of our proposed approach is to get the percentages of various emotional states (happiness, sadness, anger, fear, surprise, disgust and neutral) in a face. The emotion state having the highest parentage is treated as its resulting emotion state on a specified face. To procure such composite classification of images, a robust and enormous training is essential. Hence in our proposed approach, we have applied deep convolution neural networks to perform training and testing.

Furthermore, the paper organized asComplete system architecture and data set description has been shown in section II. The proposed algorithm and new CNN model are presented in section III. The discussions of the results shown in section IV. Finally, the conclusion and future scope presented in section V.

SYSTEM ARCHITECTURE

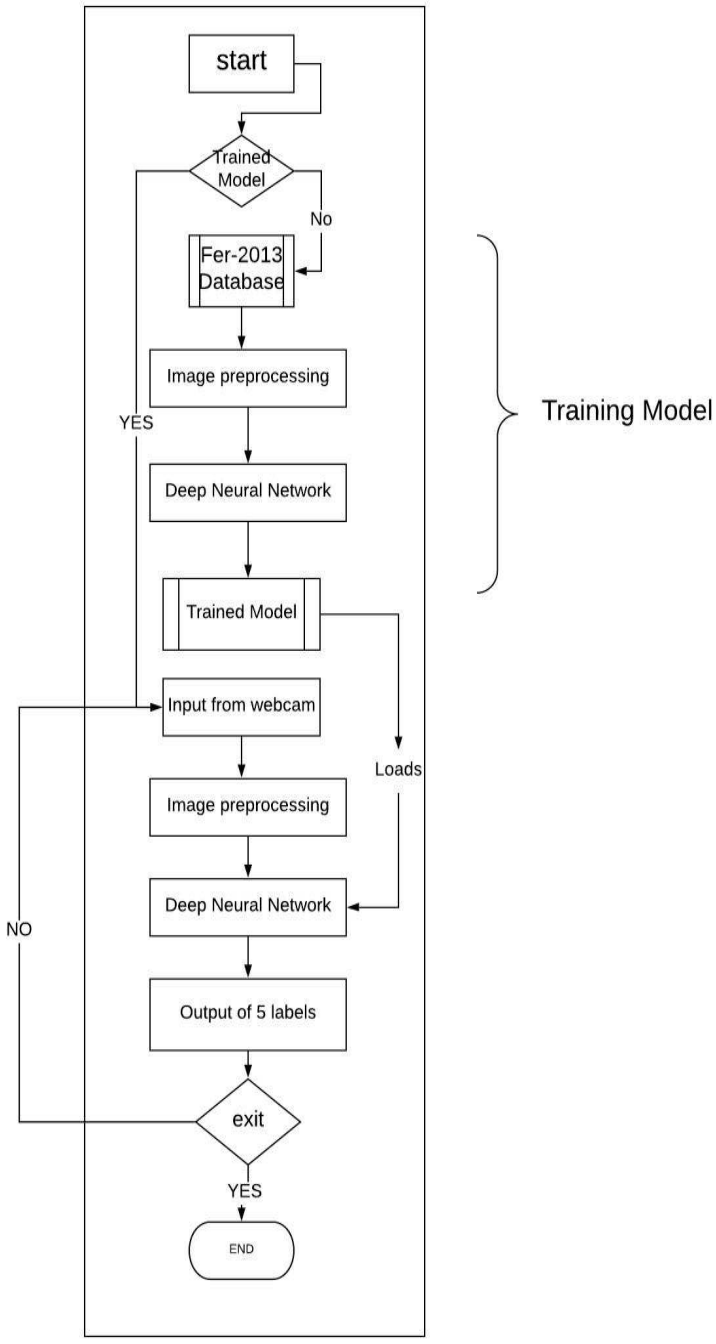

Figure 1: System Architecture

The complete architecture of the system has been represented in above Fig.1. The main algorithm of the proposed system is divided into two parts named as training and testing. First, we need to train the model to classify the face emotions of a given image or video. The first step of our proposed algorithm is to check whether the trained model present

or not. If not, then we have to train the networks to develop the model first and then we have to perform testing for face emotion classification.

The FERC-2013 dataset contains about 32000 low-resolution face pictures of dissimilar age groups and having different degrees of angle is available. The database holds of[48x48] pixels of grayscale pictures of human faces. The faces are automatically processed so that it holds up round a comparably equivalent volume of face space in all images. The prime task is to place each face in view of the emotions of one of seven classes (0: Happy, 1: Sad, 2: Surprise, 3: Angry, 4: Disgust, 5: Fear, 6: Neutral). Thus, the database exists in the form of emotion and its matching pixels array.

PROPOSED ALGORITHM

Step 1: It searches for the model if the model does not exist then it goes to step 2 otherwise goes to step 6.

Step 2: It collects the data from Fer-2013 database and does some pre-processing stuff like converting the pixel string from fer-2013 dataset to images and placing them in related folders.

Step 3: In this step we do some image pre-processing like cropping the image making the dimensions 48*48 etc.

Step 4: In this step, we build a CNN model.

Step 5: We train the model for a different time of epochs.

Step 6: We take the input from webcam or a previously saved video.

Step 7: we take the frame from the video and pass it to Haar Cascade Classifier, then image preprocessing is done as step 3 on the output to produce new frame (image).

Step 8: We give the obtained image to a deep neural network that we build in step 4.

Step 9: We get the outputs from the above model which will be shown to the user.

Step 10: If the user wants to exit then it will exit out of the application otherwise it goes to step 6 for new input.

Layer (type)	Output Shape	Param #
image_array (Conv2D)	(None, 48, 48, 32)	896
activation_7 (Activation)	(None, 48, 48, 32)	0
max_pooling2d_5 (MaxPooling2	(None, 24, 24, 32)	0
batch_normalization_5 (Batch	(None, 24, 24, 32)	128
conv2d_4 (Conv2D)	(None, 24, 24, 32)	9248
activation_8 (Activation)	(None, 24, 24, 32)	0
max_pooling2d_6 (MaxPooling2	(None, 12, 12, 32)	0
batch_normalization_6 (Batch	(None, 12, 12, 32)	128
conv2d_5 (Conv2D)	(None, 12, 12, 64)	18496
activation_9 (Activation)	(None, 12, 12, 64)	0
max_pooling2d_7 (MaxPooling2	(None, 6, 6, 64)	0
batch_normalization_7 (Batch	(None, 6, 6, 64)	256
flatten_2 (Flatten)	(None, 2304)	0
dense_3 (Dense)	(None, 512)	1180160
activation_10 (Activation)	(None, 512)	0
dropout_2 (Dropout)	(None, 512)	0
dense_4 (Dense)	(None, 5)	2565
activation_11 (Activation)	(None, 5)	0

Total params: 1,211,877
Trainable params: 1,211,621
Non-trainable params: 256

Figure 2: Layers used in our CNN Model

We have developed a new Convolution Neural Network (CNN) model and the layers used in our new CNN model are shown in the above Fig.2. The functionality of each layer is described below.

Convolution2D which will detect the edges in the image, that is outlines present in the image. Arguments in Convolution2D are filters, which specifies the number of filters to use, kernel_size is length and breadth of the kernel used, padding specifies whether to use an extra layer of zeros around the image or not.

An activation function is used whether a neuron should be activated or not by calculating a weighted sum and further adding bias with it. Here we have used to activation functions 1. ReLu 2. Softmax.

Max pooling 2D is a sample-based discretization process. The objective is to down-sample an input representation (image, hidden-layer output matrix, etc.), reducing its dimensionality and allowing for assumptions to be made about features contained in the sub-regions binned.

Batch Normalization which is used to normalize the values in the list, we need all the values to be present in between 0 to 1 instead of one value having 100 and other having 1000. This process is repeated four times and finally, we get a 6x6 image.

Then we apply to the Flatten function to convert a 2D array to 1D array. We add additional neurons to perform activation then to eliminate overfitting we drop 50% neurons using Dropout function.

Next, we apply a dense function to add output neurons and use Softmax activation function to generate probabilities of each emotion using those output neurons.

RESULTS

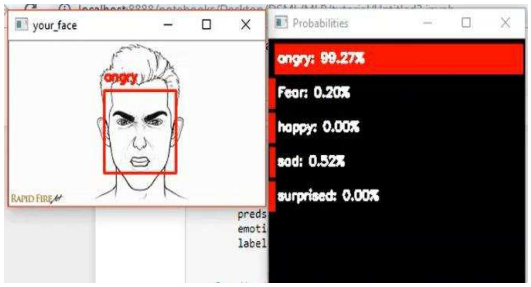

Figure 3: Happy Face

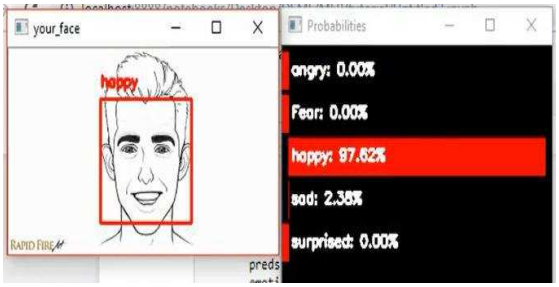

Figure 4: Angry Face

Figure 5: Happy Face (Among Multiple Faces)

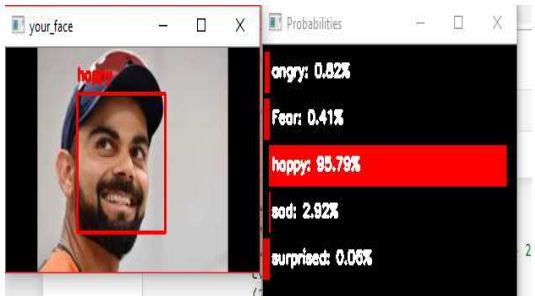

Figure 6: Happy Face

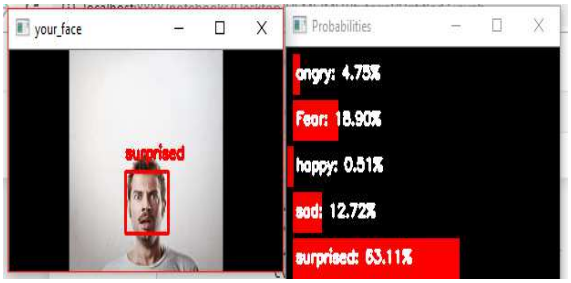

Figure 7: Surprised Face

Figure 8: Sad Face

While testing our algorithm we have given different videos like animated human faces, real human faces, etc. The detected face emotions and their percentage values have been shown in the above figures.

In Fig.3 and Fig.4 the animated human face results were shown. Our algorithm works perfectly and given correct results as *happy* and *angry* face emotions.

In Fig. 5 actually the image consists of multiple faces, in that case our algorithm will detect the face which having a large area and processes it produces the result. In the figure, it has shown that the person which has a large area has been processed and produced the correct result as *happy* face emotion.

Fig.6, Fig.7, and Fig.8 are the examples of the correctly detected face emotions as happy, surprised and sad face emotions respectively.

In few cases, our algorithm didn't give accurate results. The following are the test cases in which we get wrong facial emotions.

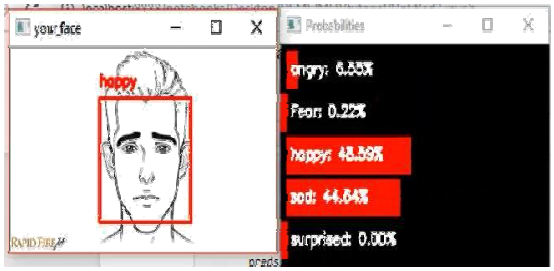

Figure 9: Sad Predicted as Happy

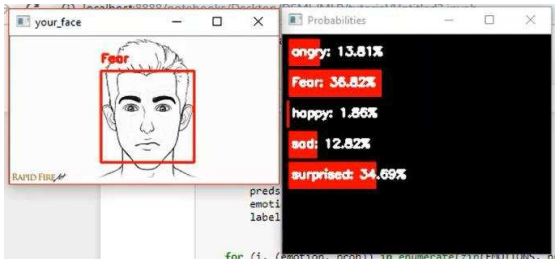

Figure 10:Fear & Sad Predicted as Fear & Surprised

CONCLUSION AND FUTURE SCOPE

Emotion's Identifier application has been computed successfully and was also tested successfully by taking many different test cases. Our model not only takes images as input but also take video streaming as input. In the case of video as input, it processes the video and produces the images as output and these images are processed to detect the face emotions. Our model is user-friendly, and has required options, which can be utilized by the user to perform the desired operation i.e. you can directly give an image or input video for testing or you can train a new model and then give an image or input video for testing. Applications of our algorithm are Human Behavior Predictor, Surveillance System, Medical Rehabilitation. Our model is built to predict up to 5 Emotion's perfectly in future we can develop a new CNN model which can predict the remaining 2 Emotion's also. In future, we can also detect expressions of multiple faces in a single frame or image.

REFERENCES

1. *Ekman, Paul and Harriet Oster. "Facial expressions of emotion." Annual review of psychology 30.1 (1979): 527-554.*
2. *P Forgas, Joseph and H Gordon. Bower. "Mood effects on personperception judgments." Journal of personality and social psychology 53.1 (1987): 53.*
3. *C Borod, Joan. The neuropsychology of emotion. Oxford University Press, 2000.*
4. *Kagan, Jerome, et al. Galen's prophecy: Temperament in human nature. Basic Books, 1994.*
5. *Aggarwal, K Jake., and Quin Cai. "Human motion analysis: A review." Nonrigid and Articulated Motion Workshop, 1997. Proceedings., IEEE. IEEE, 1997.*
6. *Rajesh Kumar G A, Ravi Kant Kumar and Goutam Sanyal. "Facial Emotion Analysis using Deep Convolution Neural Network" International Conference on Signal Processing and Communication (ICSPC'17) – July 2017.*
7. *Matsugu, Masakazu, et al. "Subject independent facial expression recognition with robust face detection using a convolutional neural network." Neural Networks 16.5 (2003): 555-559.*

8. *P. Dinesh Kumar & B. Rosiline Jeetha, A Survey an Facial Expression Recognition, International Journal of Computer Science and Engineering (IJCSE), Volume 6, Issue 1, January – February 2017, Pp.43-50.*
9. *Viola, Paul, and Michael J. Jones. "Robust real-time face detection." International journal of computer vision 57.2 (2004): 137154.*
10. *Viola, Paul, and Michael Jones. "Rapid object detection using a boosted cascade of simple features." Computer Vision and Pattern Recognition, 2001. CVPR 2001. Proceedings of the 2001 IEEE Computer Society Conference on. Vol. 1. IEEE, 2001.*