National Conference on Innovations in Computing and Networks (NCICN) - 2019
© 2019 IJSRSET | Volume 5 | Issue 9 | Print ISSN : 2395-1990 | Online ISSN : 2394-4099

Facial Emotion Recognition Analysis using Deep Convolution Neural Network

Kanchi K Sen[*1], ManuDas M[*2], Alma A[*2], Aleesha Navas[*2]

[*1]Assistant Professor, Department of Computer Science and Engineering, Younus College of Engineering and Technology, Kollam, Kerala, India

[*2] B.Tech, Department of Computer Science and Engineering, Younus College of Engineering and Technology, Kollam, Kerala, India

ABSTRACT

Human emotions are the unpredictable fluctuations of the visual model which plays a vital role in non verbal communication. Emotions are the mental states of feelings that occur spontaneously and when combined with facial muscles form an expression. The basic emotions are happy, sad, angry, fear, disgust, and surprise. The paper presents a method for emotion recognition. The proposed methodology involves 2 stages, the first stage is image preprocessing which includes face detection using viola-jones algorithm .In the next stage detection of emotions is done by using convolution neural network(CNN) which shows the intensity changes from low level to high level of emotions on a face. when compared with existing system, our proposed system has better accuracy in recognition, also works as real time system with much less complexity than the existing system, In this system FERC-2013 dataset is used as standard image dataset. The assessment of proposed system gives a quick good result and provides an encouragement for the future researchers in emotion recognition system.

Keywords : Face recognition, facial expressions, emotions, non-verbal communications, face detection convolution neural network (CNN),Deep learning, image preprocessing

I. INTRODUCTION

Human emotions are multidimensional visual model which is very complex to predict Human faces and its emotion is an important application of biometric identification expressions are clue about the mind state of a person which enables us to make communication on the basis of their mood.These fluctuations of expressions play a vital role in non verbal communication.

Developing a emotion recognition model is an exciting field for the researchers due to its wide real world applications like criminal identification, security systems, surveillance systems, human behaviour prediction and medical rehabilitation. The seven primary categories of human emotions which are common among all human across different culture are ;sadness, happiness, angry, fear, surprise, disgust and neutral. The success ingredients of the emotional happenings are the availability of large set of training data. We use [FERC-2013] dataset. There exist some other methods for facial expressions identification which includes identical bilateral amygdala impairment recognition, image pre processing methods and descriptors based local binary patterns, the active shape models arrangement of a multi resolution method, combining multiple kernal method. Convolution Neural Network(CNN) has taken the computer community top with its quick result with great accuracy.

The main focus of our proposed system is to find the standardized expression in face. Human face and hence the expressions all are the blood line. The emotion which has high paternity resulting emotion likewise more experimental outputs, training and supervision of several emotional phases(frame by frame) also stimulates is to enlarge a real time facial emotion recognition system to achieve such compound classification of image , a wide and robust training is necessary.

Hence in this recommend approach abstraction of deep learning using convolution neural network is applied to train and test. the implementation of a neural network mainly stuck on to issues like training data, initial random weight, number of hidden layers, activation function and network structure of system.

In CNN the image is directly inputed as fill-in of handcrafted midway feature, convolutional neural network are used to involuntary learn a stacking of features which can be applied future.

II. RELATED WORK

During past few years,many papers have emerge About facial emotion recognition systems.

Rajesh Kumar G.A proposed an approach for facial emotion recognition using viola jones algorithm trained with FERC-2013 database without Deep convolution neural network.The trained algorithm are compared using accuracy and computing time and finally 84%accuracy is developed.

Yong-Soo Seol proposed facial emotion recognition using a knowledge based ANN during the year 2008 which uses Hybrid approach combining knowledge based ANN gives an accuracy of 81.5% Sebe and Nicu predicts emotion recognition based on joint visual and audio cues using pattern recognition and resulted with accuracy 78% on 2006.

Shan,Caifeng,Shaogang Gong and peter W propoed a method for facial expression recognition based on local binary patterns.This paper gives overview of support vector machine classification and local binary pattern features .The paper provide accuracy of 80.04% and was published in 2009

Table 1:Comparison of related work

AUTHOR	ALGORITHMS	ACCURACY	YEAR
Rajesh Kumar G.A	Viola Jones algorithm	84%	2017
Yong-soo seol	Hybrid approach combining knowledge based ANN	81.5%	2008
Sebe,Nicu	Pattern Recognition	78%	2006
Shan,Caifeng,Shaogang	Support Vector Machine	80.04%	2009

III.METHODS AND MATERIAL

A. System architecture

The complete architecture of proposed system has been represented below.

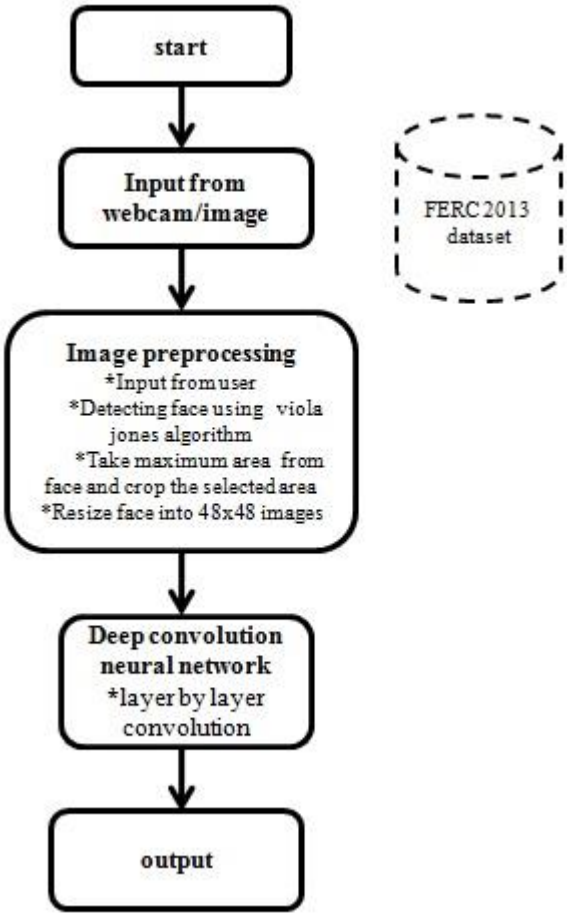

Figure 1: system workflow

The main algorithm involves two parts, testing and training. The first step is to train the network to classify the emotions of given face. The initial step of algorithm is to check whether the trained data are present or not. If not, then firstly we need to train the system and then perform testing for emotion classification.

Algorithm1: Image pre-processing

The feature computation is done by using integral images, an algorithm for generations of sum of pixel intensities is a specified rectangle in an image. it is used for compute the Haar-like features. Calculation of the sum of a rectangular area inside original image is extremely efficient requiring only four additions for any arbitrary rectangle size. Adaboost method used for feature selection of picture, by construction of strong classifiers as linear combination of weak classifiers and also to boost performances .The viola jones is a face

detector algorithm which can able to state whether a picture containing human face or not also viola jones is the most robust face detection algorithm. haar like features which are advanced features and used for object classification .from the haar feature applied to the face, the sum of black pixel and sum of white pixel calculated and they are substracted to get a single value. If the value is more in that region, then its represents a part of the face and is identified as eyes, nose ,neck ,etc.

The integral image concept is used for reducing the limit taken on computational task. The Haar function produces almost 160000 features per image, out of which all are not related for face localizing , so the Adaboost also used to removing irrelavant features. Adaboost determines relevant and irrelavant features. The set of relevant features is named as weak classifier and also constructs a strong classifier as a linear combination of weak classifier

B. Training data
Before training we pre-processed the FERC2013 database images. To pre-process, the Viola-Jones algorithm is used on the dataset. 25304 samples are used for pre-processing and validation, among them we got 11758 valid samples for training. Many samples are failed in face detection task due to the draw back of viola-jones algorithm, some issues are shown below;

Recently, the CNN have confirmed inspiring performance in computer version task for the use of CNN, excessive hardware performance is very important due to the computation difficulty. To achieve better accuracy, utilize CNN architecture according to our classification requirement parameters.
To achieve this, nine main layers are employed while designing CNN architecture.

Algorithm 2 : Deep convolutional neural network

In neural networks, Convolutional neural network (ConvNets or CNNs) is one of the main categories to do images recognition, images classifications. Objects detections, recognition faces etc., are some of the areas where CNNs are widely used. CNN image classifications takes an input image, process it and classify it under certain categories (Eg., Dog, Cat, Tiger, Lion). Computers sees an input image as array of pixels and it depends on the image resolution. Based on the image resolution, it will see h x w x d(h = Height, w = Width, d = Dimension. Technically, deep learning CNN models to train and test, each input image will pass it through a series of convolution layers with filters (Kernals), Pooling, fully connected layers (FC) and apply Softmax function to classify an object with probabilistic values between 0 and 1. The below figure is a complete flow of CNN to process an input image and classifies the objects based on values.

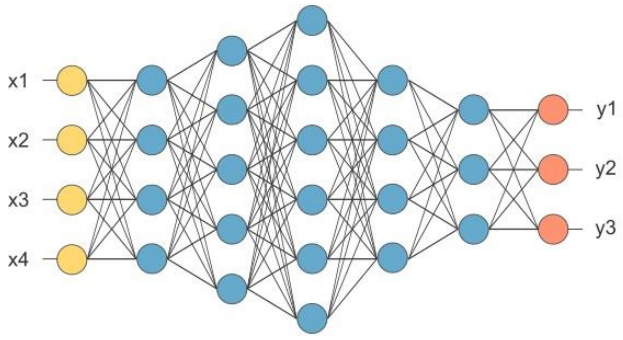

Figure 2:convolution neural network

Pooling layers section would reduce the number of parameters when the images are too large.The fully connected layer flattened our matrix into vector and feed it into a fully connected layer like neural network In the above diagram, feature map matrix will be converted as vector (x1, x2, x3, …). With the fully connected layers, we combined these features together to create a model. Finally, we have an activation function such as softmax or sigmoid to classify the outputs as cat, dog, car, truck etc.

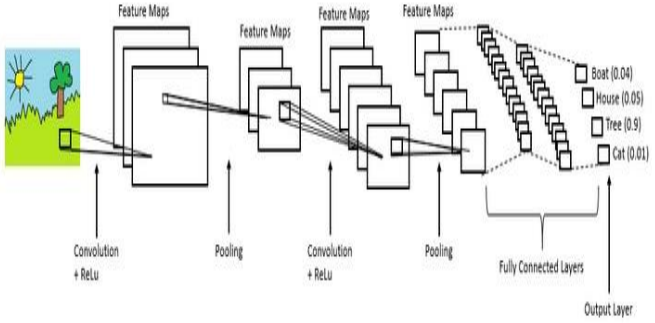

Figure 3 : Complete CNN architecture

Phase1: initialize all the weights and filters with random values
***Phase2*:** Input the training images to the network and goes over between propagation phases and detects output probabilities for all class.
For example: the output probabilities for the first input image are [0,0,0,0.3,0.3,0.2,0.5].
The weights are randomly assign for the first training image ,so output probabilities are also random.

Phase3: Calculate the entire error at output layer by using the equation,
Total error = [(target probability-output probability)^2]

Phase4: Using the back propagation algorithm, calculate the gradients of error for all weights in the network. To

minimize the output error use that gradient decent to update all weights, filter values and parameter values.
The weights are updated is proportion to reducing the total error when the same image is inputed again, the probabilities might be [0,0,0,0.1,0.7,0.1,0.1] which is closer to the target letter[0,0,0,0,1,0,0] this implies that by altering its weights and filters the network has learnt to categorize this particular image correctly. so that error of output is reduced. factors such as network architecture ,number of filters used and filter sizes, etc have been fixed before first step and don't change during training process. only the values of filter matrix and connection weight updated during the training process

Layer by layer explanation of CNN

Layer 0: input layer
[48x48x1] the pixel value of the input image. Width 48,height 48 and with one colour channel

Layer1: convolutional layer
Compute output of all neurons associated to local regions in the input layer, each calculating a dot product among their weights and a small region they are related in .

Layer2: Rectifier linear unit layer
Most commonly deployed activation function for the output of the CNN neurons. This is done usually as max(0,x)zero. This leaves the volume size same ([44x44x64]),and group normalization is done

Layer 3: Max pool layer:
This will perform a down sampling operation along the spatial dimensions (width, height), resulting in volume such as [22x22x64]. Max-Pooling with 3×3 filter and stride 2, gives size [22x22x64], i.e. (44-3)/2+1=22 is output size, depth is same as before, i.e. 64 because pooling is done independently on each layer.

Layer 4:
Convolution with 64 filters, size 5×5, stride 1, now size is [18x18x64],i.e. (22-5)/1+1=18; is size of output 64 depths because of 64 filters.

Layer 5:
Max Poling Layer with 64 filters, size 5×5, stride 1,now size is [18x18x64],i.e. (18+2*1-3)+1=18 original size is restored..

Layer 6:
Convolution with 128 filters of size 4x4 and stride 1and we used padding 0,therefore now size is given as [15x15x128], i.e. (18-4)/1+1=15, is size of output 64 and depths of 128 filters.

Layer 7:
Fully connected with 3072 neurons. In this layer, each of the 15x15x128=28800 pixels is fed into each of the 3072 neurons and weights determined by back-propagation.

Layer 8:
Fully-connected layer calculates the class scores, resultant volume of size [1x1x7], where each of the seven numbers correspond to a class score, such as among the seven classes of emotions. As with normal neural networks and as the name implies, each neuron in this layer will be linked to all the numbers in the previous volume and soft max layer with 3072 neurons.

Layer 9:
Soft max layer with 7 neurons to predict 7 classes output

IV. RESULTS AND DISCUSSION

The detected facial emotions and their percentage have been shown below

Figure 2 :Row wise (1)is Angry face,(2)is disgusted face,(3)is surprised face,(4)is happy face,(5)is sad face,(6)is fearful face

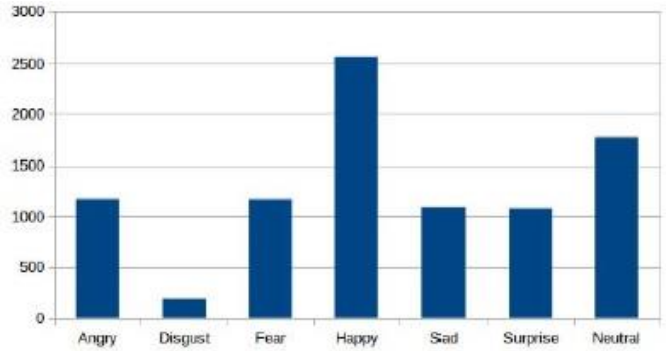

Figure 3.successfully detected percentage of facial emotions

	Algorithm	Accuracy
Proposed System	Viola Jones algorithm	84%
	Hybrid approach combining knowledge based ANN	81.5%
	Pattern Recognition	78%
	Support Vector Machine	80.04%
Existing System	Viola jones,Deep Convolution Neural Network	90%

Table 2. Existing v/s Proposed system

This section describes the results of both proposed and existing systems. when comparing the proposed system with the existing system the rate of accuracy spots much increase to better extend with the less consumption of time.This system also performs as a real time emotion detector and also so in return provides better precision.

V. CONCLUSION

Feature extraction from the emotions are quite tricky phase in the process of recognition .the percentage of emotions differs from stages to stages.In our proposed system we use 9 layers of convolution neural network for classification and training of 7 types of standard emotions.FERC-2013 database provides the data. Viola jones algorithm is implemented for face detection. By using convolution neural network our proposed system projects an accuracy of around 90+% than the many existing systems.

VI. ACKNOWLEDGEMENT

The Authors would like to acknowledge Dr. p. Sreeraj, Principal of younus college of engineering & technology, Dr. Nijilraj. N, Head of CSE Department in ycet, prof.Kanchi K Sen, Guide and Prof.Yasir.A,project coordinator of CSE department and younus college management, supporting for this work

VI. REFERENCES

[1]. P Forgas, Joseph., and H Gordon. Bower. "Mood effects on personperception judgments." Journal of personality and social psychology 53.1 (1987): 53.

[2]. Adolphs, Ralph, et al. "Impaired recognition of emotion in facial expressions following bilateral damage to the human amygdala." Nature 372.6507 (1994): 669.

[3]. Zhang, Zhengyou, et al. "Comparison between geometry-based and gabor-wavelets-based facial expression recognition using multi-layer perceptron." Automatic Face and Gesture Recognition, 1998.

[4]. Proceedings. Third IEEE International Conference on. IEEE, 1998.

[5]. Viola, Paul, and Michael J. Jones. "Robust real-time face detection." International journal of computer vision 57.2 (2004): 137-154

[6]. Pritpal singh ”Study of Face Recognition techniques” ." International journal of computer vision .4 (2017): 23-95

[7]. Shan, Caifeng, Shaogang Gong, and Peter W. McOwan. "Facial expression recognition based on local binary patterns: A comprehensive study." Image and Vision Computing 27.6 (2009): 803-816

[8]. Sebe, Nicu, et al. "Emotion recognition based on joint visual and audio cues." Pattern Recognition, 2006. ICPR 2006. 18th International Conference on. Vol. 1. IEEE, 2006.