International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 07 Issue: 04 | Apr 2020 www.irjet.net p-ISSN: 2395-0072

DETECTION OF HUMAN EMOTIONS IN AN IMAGE USING CNN

Kuppa Srivalli[1], Inkollu Chandana[2], Guddanti RamyaSri[3], Kothapalli Swathi[4], Dr. T. Kameswara Rao[5]

[1,2,3,4]Graduation students, Dept. of Comp. Sc. & Engg., Vasireddy Venkatadri Inst. of Tech., Guntur, AP, India
[5]Professor, Dept. of Comp. Sc. & Engg., Vasireddy Venkatadri Inst. of Tech., Guntur, AP, India

---

Abstract - *Emotions possessed by humans can be recognized and has a vast scope of study in the computer vision industry. Emotions can be recognized by various means. Facial expression, body language, voice level, etc. can be used for detection. In this paper we used a model to recognize emotion by facial expressions which play very important role in human communication. Since human face is the richest source of emotion, in which conversations depend much on emotions. There are numerous human-machine interactions. Primary objective of this paper is to develop a system which can analyze the image and recognize the expressions of the person by using Deep learning techniques. These expressions can be derived from System's camera or any existing images available. Limitation of this paper is to recognize only seven emotions viz happiness, sadness, anger, disgust, fear, surprise and neutral.*

Key words: Convolution Neural Networks, Emotion detection, Expression detection, Deep learning.

1. INTRODUCTION

The Study of emotions in human-machine interaction has increased in recent years [1], [2]. Computer Systems used mouse and keyboard and the trend was shifted to automatic speech recognition and now human machine interfaces are quite trending all over. The communication between humans and computers will be natural if computers are able to perceive and respond to human nonverbal communication such as emotion [18], [19].

If computers are able to recognize these emotions they could give specific and appropriate help to the users in order to satisfy user's needs and preferences. If systems are successful in classification of emotions, then it gains better understanding of human behaviour by using the information technology. This makes the systems and user interfaces more emphatic and intelligent. Facial Expression recognition problem attracted computer vision community [20]-[27].

Facial emotion detection applications are spread across different fields like medicine, e-learning, marketing, monitoring, entertainment, and law. Counselling. Determination of medical state of person, determining feelings, comfort level to treatment [7], adjusting the learning technique by determining emotion, ATM not dispensing money when person is scared while withdrawing money, prioritizing angry calls in call centres, recognizing mood and satisfying needs, purchasing decisions and improving sales, these are the various areas where emotion recognition technique is used. Music therapy helps patients deal with stress, anxiety, depression and a positive effect to Alzheimer. In learning phase staff understands the student situation and try to give more information effectively and interactive. In monitoring phase if the car driver is sleepy, emotion is detected and alert others. This leads to less number of accidents.

Human emotion detection has been implemented in many areas requiring additional security or information about the person and also used for business promotions [3], [4]. Humans share a rich set of emotions which are expressed through consistent facial expressions. Human Emotions can be classified as happy, sad, neutral, angry, disgust, fear, anxiety etc. [5], [6]. These Emotions are very subtle. Facial muscle contraction for different emotions is different. Detecting these small changes is difficult as even a small contrast results in Different Expressions [28]. Machine Learning have been used for these tasks to obtain good results. There are many methods for facial expression classification like Hidden Markov model, SVM, Adaboost, Artificial Neural Networks [8], [9].

Neural Networks consumes a lot of computational power and they are very slow because of high quality and more input parameters.

Convolution Neural Networks (CNN) uses less preprocessing when compared with other algorithms. The proposed system has different phases such as face detection, emotion recognition, generation of output image with recognized emotions. In this paper we focus on the Facial expression recognition using convolutional neural networks. Data preprocessing and emotion recognition model developed based on CNN are two major phases of the developed system.

2. DATA PREPROCESSING

Original data if directly used for emotion detection, it takes lot of computational power due to more number of input parameters. Model is said to be robust only when it uses less computational power.

To overcome these problems the data is pre-processed. Data preprocessing includes:

a) Face detection
b) Normalization

C) Gray level equalization

A. FACE DETECTION

Open CV by default comprise of a trainer as well as a detector. Classifier of any kind can be generated to detect fire, cars, planes and other objects. Open CV has many pre-defined classifiers for face detection. The XML file is stowed in Open CV/data/Haar cascades folder.

Haar cascades classifier is trained by using the positive and negative images. Edge, line, four rectangle, diagonal are its main features and are shown in Fig.1. Important facial features are extracted from large number of Haar-like features. Haar Classifier is highly efficient and so it is used widely.

XML classifiers that are required in detecting a face are loaded, after which input image is loaded.

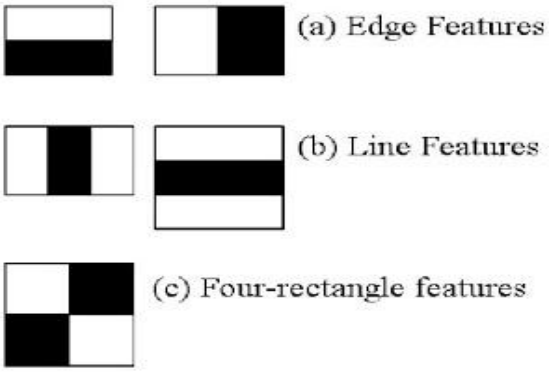

Fig -1: Haar features for face detection

B. NORMALIZATION

The input layer of the network accepts image of specific size, so the original picture is normalized to that size. Scaling ratio of x-axis and y-axis are used to scale the image. Many other normalization techniques can be used but scale normalization is one of the efficient technique and hence used in this model [10].

The points in the original picture are mapped to other points by using scale normalization. In both the directions image is scaled by using the scaling ratio's s_x and s_y.

a) before normalization b) after normalization

Fig -2: Distinction of before and after normalization

C. GRAY LEVEL EQUALIZATION

Images will show state of uneven distribution due to shadows. Factors like illumination, shade effects the feature extraction process, to overcome this problem gray scale values are averaged to get even distribution. In this paper Histogram Equalization (HE) technique is used [11]. The technique used converts the histogram of original image to even distribution.

The results are showed in the fig.3. Gray level Equalization increases the contrast of image. It makes details clearer. The obtained image is conducive for facial features extraction.

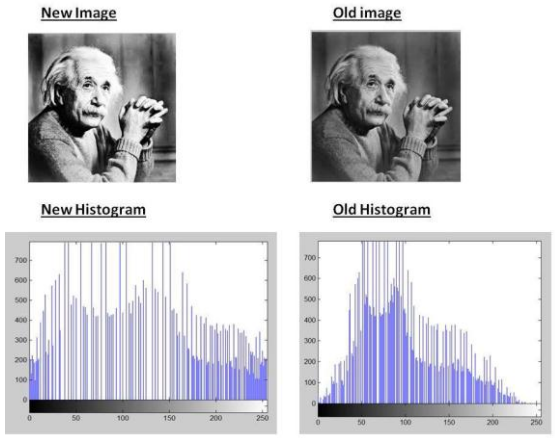

Fig -3: Image before and after gray scale normalization

The image which is passed in the above steps is pre-processed and is now ready to enter into the emotion recognition phase.

3. EMOTION RECOGNITION MODEL BASED ON CNN

Artificial neural networks inspired by human brain learn from large amounts of data. Convolution neural networks assign importance to objects and differentiate one from other. Convolution neural network is feed forward neural network. It extracts features from image which are important for emotion detection and classification. Back propagation algorithm is used to optimize network parameters. The model will be trained to recognize seven emotions as shown in the figure 4

Fig -4: Training Process

The nature of deep learning method is to build neural networks which learns features. The rules are developed at the end of training as shown in the figure 5. Rules are nothing but the weights. In the training phase, the network is initialized with random weights. Training pattern is feed in to get output. The obtained output is compared to target output. Adjust weights based on error. This process is repeated until all patterns are passed one time which is called as epoch. Like this the process is carried out for every epoch.

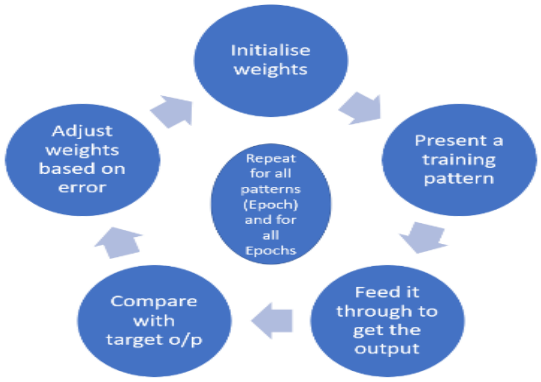

Fig -5: Detailed training view

Initially the pre-processed image is passed to the network's input layer. Layer by layer processing of the image is done. Every layer has two-dimension planes. They are nothing but feature maps, it has several neurons. Computer understands the image by its pixel values. Neural networks interpret the image from pixel to the lines, curves, edges, finally objects understood by human brain, then the emotion is recognized. This paper designs CNN structure with the following layers:

a) Convolution layer
b) Pooling layer
c) Fully connected layer
d) SoftMax layer

Layers involved in this model are: **Input->Conv(C1)->Pool(S1)->Conv(C2)->Pool(S2)->Conv(C3)->Pool(S3)->FC->Softmax**

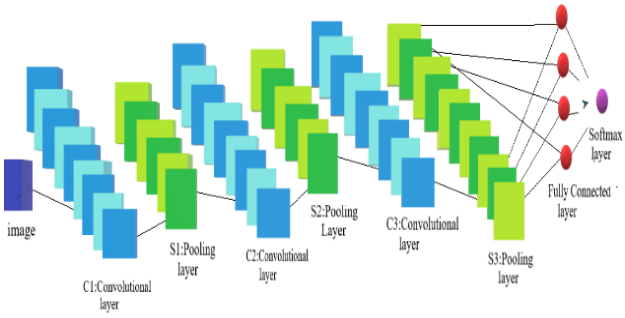

Fig -6: CNN Structure for emotion recognition

Input layer is two-dimensional matrix. It is composed of image pixels. The model presented in this paper used four convolution layers. The gray scale image with 48 x 48 pixel matrix is the input to this layer. Every feature map is connected to its previous map. In every layer there are several feature maps. The convolution layer C1 uses 64 convolutional nuclei. The size of convolution nuclei is 3 x 3. Layer C2 and C3 uses 128 nuclei. Softmax layer contains 7 neurons. The feature of the output layer is classified among the seven emotions. .Layers as shown in figure 9.

A. CONVOLUTION LAYER

Convolution layer is the basic building block of a Convolutional Network. It does most of the computational work to the model. Convolution is a mathematical operation on two functions to produce a third function. The convolution operation is associated with local connectivity, which learns correlations among neighboring pixels and process. Image is matrix of numbers as seen by computer. These numbers represents pixel values.

In Convolution layer every node is not connected to every other neuron. Connections are localized. The convolution is applied on the input image using a convolution filter. It is also called as kernel is used to produce a feature map. A filter passes over the input image, [12] scanning a few pixels at a time and creating a feature map that predicts the class to which each feature belongs [13]. Padding and stride are two main properties of every filter. Activation Function is used.

In the proposed model a kernel of 3x3 is used to detect the features like edge detection, curved and sharpen features by applying different filters. Relu is an Activation Function which performs non-linearity. In the past, nonlinear functions like tanh, sigmoid are used now Relu is used as it is far better because the network is able to train a lot faster without making a significant difference in accuracy. Relu function operates as below:

$$f(x) = \max(0, x).$$

This layer changes the negative values to 0.

It was already proved that the network using Relu activation function has moderate sparsity. It solves the problem of gradient disappearance. This may occur in the process of modifying back propagation parameters. This accelerates convergence of the network. The features extracted by convolution operation can be directly used to train classifiers, but it has lot of computational challenges. To reduce the parameters, down sampling operation is proposed after the convolution. The average or maximum value of a specific feature in an image are computed. This

statistical dimensionality reduction method reduces the number of parameters, prevents fitting.

B. POOLING LAYER

After Convolutional layer, the process undergoes to pooling layer [14], [15]. Pooling reduces the amount of information in each feature obtained in the above layer and also maintains the most important information. Pooling reduces the dimensions and so it is also called as dimensionality reduction. In pooling layer there are different types. They are MaxPooling, MinPooling and AveragePooling. Mostly used pooling is MaxPooling. This layer normally takes a filter (normally size of 2x2) and a stride of same length. It then applies it to the values and place the maximum number in every region that the filter moved around into the output matrix.

The addition of a pooling layer after the convolutional layer is a common pattern used for ordering the layers within a convolutional neural network that may be repeated one or more times in the give model. The pooling layer operates on each feature map separately. It creates a new set of the same number of pooled feature maps. Pooling involves selecting a pooling operation. It is like a filter to be applied to feature maps. The size of the pooling operation or filter is smaller than the size of the feature map, specifically, it is almost 2x2 pixels applied with a stride of 2 pixels. For example if we take feature map of size (224x224) then after pooling is applied the output matrix will be of size (112x112) as shown in figure 7.

It halves the dimensions of the next feature map. Computational complexity is for the next convolution operation is reduced. Training speed is increased. If the softmax layer is trained directly with the learnt features with reducing dimensions, then dimension problem occurs and computations will be high. So to avoid this pooling layer is used after convolution layer.

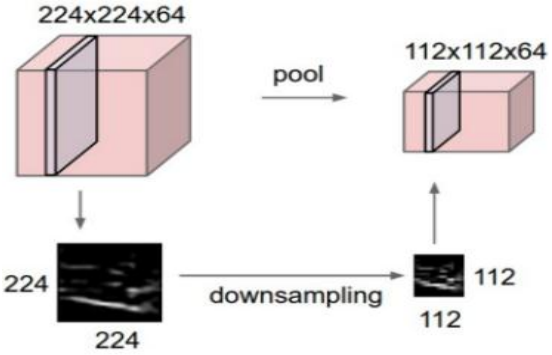

Fig -7: Pooling Procedure

C. FULLY CONNECTED LAYER

Fully connected layers are an essential component of Convolutional Neural Networks (CNN), which have been proven very successful in recognizing and classifying images for computer vision. Convolutional layer and pooling layer do feature extraction whereas fully connected layers do classification based on the features by previous layer.

Fully Connected input layer(flatten): This layer takes the output of the previous layers i.e., (Convolutional layers and pooling layers) flattens them and returns them into a single vector that can be input to the next stage. Flatten means the matrix is reduced to a single dimension vector.

First Fully Connected layer: This layer takes the input from the feature analysis. It applies weights to predict the correct label.

Fully Connected output layer: This layer gives the final probabilities for each label. The 1D data acts as the input to the neurons of this layer which performs a dot product of this input data and the neurons weights to produce a single number as output (probability values) single value per neuron.

In fully connected layer every neuron in one layer is connected to every other neuron in another layer. In general neural networks the input image is directly given to the network in which every neuron is connected to every other neuron, there the computations are high, the speed becomes low and the results will be accurate as every parameter is considered which improves the quality.

In CNN before coming to this layer, dimensions are reduced and the actual training is done in this layer. The main difference of NN and CNN is that all the layers before this layers handle preprocessing in CNN, but for traditional NN separate preprocessing is done.

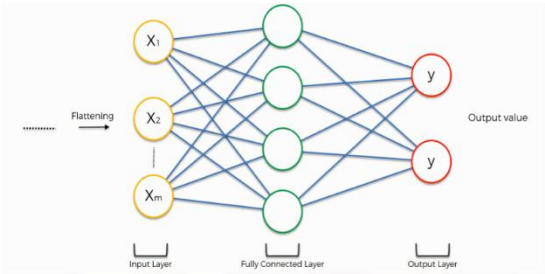

Fig -8: Fully Connected layer functionality

D. SOFTMAX LAYER

This is the last layer in the process to detect the emotion. It takes the output from the previous layer i.e., Fully Connected layer, the output is probability values, after applying softmax function the values are in between (0, 1). For the given image

the probability for every emotion is calculated. All probability values are compared the one which is having highest is declared as the emotion state for the given input.

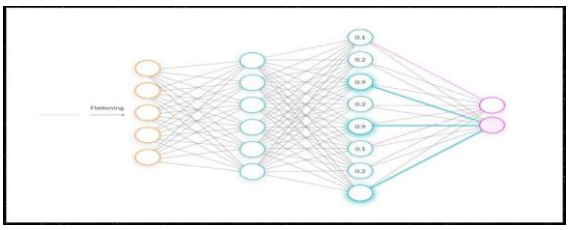

Fig -9: Softmax layer with probabilities detected at each neuron

E. DATASET STUDY

The data used for training and testing is ICML 2013 dataset. It comprises of 35, 917 images. This dataset has images with emotions like happy, sad, disgust, angry, fear, surprise, neutral. 80 percent of images in the dataset are used for training and 20 percent of images are used for testing. The dataset contains 48x48 pixel gray scale images of faces. Neural networks performs well when more data is used for training. The images have front viewed faces. This data set when compared to FER 2013 has more number of disgust images. So this dataset is chosen. Figure 10 shows the dataset images.

Fig -10: Dataset Images

For the training and testing the network. Dataset plays a very important role. Even though the model developed is robust if the dataset is not good, it leads to wrong outcomes.

4. RESULT ANALYSIS

The model developed takes the image as input and detects the emotion of every person in the given image and produces output image with emotions. To get how the model performs on each emotion it is depicted as follows. It is found that happiness emotion is obtained with high accuracy and neutral with low accuracy as it is matched with sad expression until carefully examined.

Happiness is the most desired expression of a human. The path of the image is set to trained data model to recognize the emotions. Model discussed in this paper can detect a maximum number of images in a group as shown in fig 11.

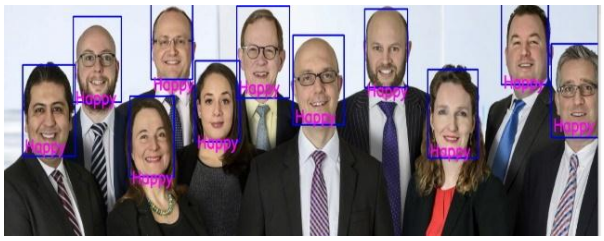

Fig -11: Happy emotion detection

A group of expressions in an Image showing different kind of emotions is shown in fig12.

Fig -12: Model output detecting different emotions

In the below fig13 group of 4 images, the side faces are not recognized because the model detects only images capturing face at least 70 Percentage as in second left image.

Fig -13: Face detection

Image showing different kind of emotions. Neutral is not satisfactory result of our model as the edge detection of this emotion is not perfect.

5. CONCLUSION

The system proposed in this paper detects emotion of a group of persons in a given image by their facial expression using convolutional neural network. It primarily comprises of three parts i.e., face representation, emotion pattern recognition and classification. CNN detects the patterns of emotion as we trained the model with large amount of data. The proposed system detects seven major emotions they are happy, sad, angry, fear, disgust, surprise and neutral.

The data with images of even size collected are fed to model. Compared to past literature this model is able to recognize disgust emotion accurately and this model is

robust to noise. We achieved a performance of detecting expression with 80 percent accuracy. The proposed model detects the emotion of a person if face is recognized more than 70% in an image.

Future works should attempt to the development of model to detect emotion of a person even with half faces, detecting emotion in video's and dynamic recognition of emotion with 3D technology.

REFERENCES

[1] R. M. Mehmood, R. Du, and H. J. Lee, ``Optimal feature selection and deep learning ensembles method for emotion recognition from human brain

[2] T. Song, W. Zheng, C. Lu, Y. Zong, X. Zhang, and Z. Cui, ``MPED: A multi-modal physiological emotion database for discrete emotion recognition,''IEEE Access, vol. 7, pp. 12177_12191, 2019

[3] E. Batbaatar, M. Li, and K. H. Ryu, ``Semantic-emotion neural network for emotion recognition from text,'' IEEE Access, vol. 7, pp. 111866_111878, 2019.

[4] H. Meng, N. Bianchi-Berthouze, Y. Deng, J. Cheng, and J. P. Cosmas,``Time-delay neural network for continuous emotional dimension prediction from facial expression

[5] sequences,'' IEEE Trans. Cybern., vol. 46, no. 4, pp. 916_929, Apr. 2016.

[6] X. U. Feng and J.-P. Zhang, ``Facial microexpression recognition: A survey,''*Acta Automatica Sinica*, vol. 43, no. 3, pp. 333_348, 2017

[7] M. S. Özerdem and H. Polat, ``Emotion recognition based on EEG features in movie clips with channel selection,'' Brain Inf., vol. 4, no. 4, pp. 241_252, 2017.

[8] F. Vella, I. Infantino, and G. Scardino, ``Person identi_cation through entropy oriented mean shift clustering of human gaze patterns,'' Multime-dia Tools Appl., vol. 76, no. 2, pp. 2289_2313, Jan. 2017.

[9] S. K. A. Kamarol, M. H. Jaward, H. Kälviäinen, J. Parkkinen, and R. Parthiban, ``Joint facial expression recognition and intensity estimation based on weighted votes of image sequences,'' Pattern Recognit. Lett., vol. 92, pp. 25_32, Jun. 2017.

[10] J. Cai, Q. Chang, X.-L. Tang, C. Xue, and C. Wei, ``Facial expressionrecognition method based on sparse batch normalization CNN,'' in *Proc.37th Chin.* Control Conf. (CCC), Jul. 2018, pp. 9608_9613.

[11] M. Takalkar, M. Xu, Q. Wu, and Z. Chaczko, ``A survey: Facial micro-expression recognition,'' Multimedia Tools Appl., vol. 77, no. 15, pp. 19301_19325, 2018.

[12] Magudeeswaran and J. F. Singh, ``Contrast limited fuzzy adaptive histogram equalization for enhancement of brain

[13] images,'' Int. J. Imag. Syst. Technol., vol. 27, no. 1, pp. 98_103, 2017.

[14] F. Zhang, Q. Mao, X. Shen, Y. Zhan, and M. Dong, ``Spatially coherent feature learning for pose-invariant facial expression recognition,'' ACM Trans. Multimedia Comput., Commun., Appl., vol. 14, no. 1s, Apr. 2018, Art. no. 27.

[15] H. Ma and T. Celik, ``FER-Net: Facial expression recognition using densely connected convolutional network,'' *Electron. Lett.*, vol. 55, no. 4, pp. 184_186, Feb. 2019.

[16] L. Wei, C. Tsangouri, F. Abtahi, and Z. Zhu, ``A recursive framework for expression recognition: From Web images to deep models to game dataset,'' *Mach. Vis. Appl.*, vol. 29, no. 3, pp. 489_502, 2018.

[17] S. Li andW. Deng, ``Reliable crowdsourcing and deep locality-preserving learning for unconstrained facial expression recognition,'' *IEEE Trans.Image Process.*, vol. 28, no. 1, pp. 356_370, Jan. 2018

[18] A.Mehrabian, "Communication without words," Psychology today, vol.2, no.4, pp.53-56, 1968.

[19] R.W. Picard, Affective Computing. Cambridge.MA : MIT Press, 1997.

[20] D. Beymer, A. Shashua, and T. Poggio, Example Based Image Analysis and Synthesis, M.I.T. A.I. Memo No. 1431, 1993.

[21] I.A. Essa and A. Pentland, "A Vision System for Observing and Extracting Facial Action Parameters", Proc. IEEE CVPR, pp.76-83, 1994.

[22] H. Li, P. Roivainen, and R. Forcheimer,"3D Motion Estimation in Model-Based Facial Image Coding", IEEE Trans. Pattern Analysis and Machine intelligence, vol. 15, pp. 545-555,1993.

[23] K. Mase, "Recognition of Facial Expression from Optical Flow", IEICE Trans., vol. E 74, pp. 3474-3483, 1991.

[24] K. Matsuno, C. Lee, and S.Tsuji, "Recognition of Human Facial Expressions without Feature Extraction", Proc. ECCV, pp. 513-520, 1994.

[25] M. Rosenblum, Y. Yacoob, and L.S. Davis, "Human Emotion Recognition from Motion Using a Radial Basis Function Network Architecture", IEEE Workshop Motion of Non-Rigid and Articulated Objects, Austin, Texas, pp. 43-49, Nov. 1994.

[26] D. Terzopoulos and K. Waters, "Analysis and Synthesis of Facial Image Sequences Using Physical and Anatomical Models", IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 15, pp. 569-579, 1993.

[27] Y.Yacob and L Devis, "Recognizing Human facial expression from long image sequences using optical flow", IEEE transaction on Pattern Analysis and Machine Intelligence [PAMI], 18{6}: 636-642, 1996.

[28] J. N. Bassili, Emotion recognition: The role of facial movement and the relative importance of upper and lower areas of the face, J. Personality and Social Psych., vol. 37, pp. 204