

# FACIAL EXPRESSION RECOGNITION SYSTEM USING CONVOLUTIONAL NEURAL NETWORKS

Somula Ramasubbareddy, K.Govinda, E.Swetha

**Abstract:** Facial expression recognition has been a functioning exploration territory in the previous ten years, with developing application regions including symbol activity, neuromarketing and amiable robots. The acknowledgment of outward appearances isn't a simple issue for AI techniques, since individuals can change altogether in the manner they demonstrate their looks. Indeed, even pictures of a similar individual in a similar outward appearance can shift in splendor, foundation and present, and these varieties are underscored if thinking about various subjects (due to varieties fit as a fiddle, ethnicity among others). Albeit outward appearance acknowledgment is contemplated in the writing, few works perform reasonable assessment abstaining from blending subjects while preparing and testing the proposed calculations. Thus, outward appearance acknowledgment is as yet a difficult issue in PC vision. In this work, we propose a straightforward answer for outward appearance acknowledgment that utilizes a blend of Convolutional Neural Network and explicit picture pre-handling steps. Convolutional Neural Networks accomplish better precision with huge information. Be that as it may, there are no openly accessible datasets with adequate information for outward appearance acknowledgment with profound structures. Subsequently, to handle the issue, we apply some pre-preparing systems to extricate just demeanour explicit highlights from a face picture and investigate the introduction request of the examples amid preparing. An investigation of the effect of each picture pre-preparing task in the precision rate is displayed. The proposed strategy: accomplishes aggressive outcomes when contrasted and other outward appearance acknowledgment techniques – going up to 92% precision - it is quick to prepare, and it takes into consideration ongoing outward appearance acknowledgment with standard PCs.

**Keywords:** Embedded Vision System, MLP, Neural Network, territory saving projection.

## I. INTRODUCTION

The goal of this task is to characterize pictures of human appearances into discrete feeling classifications. Many built up outward appearance acknowledgment (FER) frameworks utilize standard AI and extricated highlights, which don't have critical execution when connected to already inconspicuous information. Inside the previous couple of months a couple of papers have been distributed that utilization profound learning

for FER which have been fruitful at accomplishing about . Taking note of the achievement of CNNs in this area, our goal is to try different things with both new and existing system models to accomplish comparable outcomes on another informational collection. Our primary target is to perceive outward appearance of an individual utilizing neural-AdaBoost acknowledgment framework and guarantee that the acknowledgment framework is exact and the speed of execution is high. We additionally go for utilizing 3-layer feed-forward neural system which is to be prepared by backpropagation calculation. We likewise train the ANN on the presented feeling datasheet without any preparation and to test the strength by testing every single irregular picture from the dataset.

## Literature Review

This paper introduces a Face Detection System with Expression Recognition utilizing Artificial Neural Networks. It is a robotized vision framework planned and executed utilizing MATLAB. The Face Detection with Expression Recognition framework achieves outward appearance acknowledgment through two stages. The caught picture is handled first to distinguish the face, and after that the outward appearance is perceived. These two stages are finished in five phases. The initial two phases of the framework manage recognizing and trimming the face utilizing picture preparing, specifically the Viola-Jones object identification system. The third stage manages changing over the shades of the trimmed picture from RGB into dim scale and applying the proper smoothing channel. The fourth stage comprises of highlight extraction utilizing Artificial Neural Networks, so as the extricated highlights are contrasted and preparing tests. The last stage characterizes the given yields and shows outward appearance acknowledgment results. It at that point decides if the subject is upbeat, furious or in unbiased state [8]. The Artificial Neural Network utilizes Multi-Layer Perceptron (MLP) with back proliferation calculation for highlights extraction and order. It has 4097 information hubs, one concealed layer with 50 neurons, and one yield layer. Testing results demonstrate that this framework can be utilized for translating three outward appearances: satisfaction, outrage and impartial. It removes exact yields that can be utilized in different fields of concentrates, for example, mental appraisal.

**Revised Manuscript Received on July 05, 2019.**

Somula Ramasubbareddy, Asst.Prof, VNRVJIET, Hyderabad, Telangana,

K.Govinda, Associate Professor, VIT University, Vellore, Tamilnadu

E.Swetha, S V College of Engineering, Tirupati, Andhra Pradesh

Mail: svramasubbareddy1219@gmail.com.

## FACIAL EXPRESSION RECOGNITION SYSTEM USING CONVOLUTIONAL NEURAL NETWORKS

The present establishments are confronting real security issues; therefore, they need a few extraordinarily prepared staff to accomplish the ideal security. These work force, as people, commit errors that may influence the dimension of security [9].

A proposed answer for the previously mentioned issue is a Face Recognition Security System, which can identify gatecrashers to confined or high-security regions, and help in limiting human mistake. This framework is made out of two sections: equipment part and programming part. The equipment part comprises of a camera, while the product part comprises of face-identification and face-acknowledgment calculations programming. At the point when an individual enters to the zone being referred to, a progression of previews are taken by the camera and sent to the product to be examined and contrasted and a current database of confided in individuals. A caution goes off if the client isn't perceived.

Recognizing lies is urgent in numerous regions, for example, airplane terminal security, police examinations, counter-fear mongering, and so on. One system to identify lies is through the distinguishing proof of facial miniaturized scale demeanors, which are brief, automatic looks appeared on the substance of people when they are attempting to disguise or subdue feelings. Manual estimation of miniaturized scale articulations is hard work, tedious, and mistaken. This paper shows the Design and Development of a Lie Detection System utilizing Facial Micro-Expressions. It is a robotized vision framework planned and executed utilizing LabVIEW. An Embedded Vision System (EVS) is utilized to catch the subject's meeting. At that point, a LabVIEW program changes over the video into arrangement of casings and procedures the edges, each at once, in four continuous stages. The initial two phases manage shading change and sifting. The third stage applies geometric-put together unique layouts with respect to each casing to indicate key highlights of the facial structure. The fourth stage separates the required estimations so as to recognize facial small scale articulations to decide if the subject is lying or not. Testing results demonstrate that this framework can be utilized for deciphering eight outward appearances: satisfaction, bitterness, happiness, outrage, dread, amazement, nauseate, and hatred, and recognizing facial small scale articulations. It extricates exact yield that can be utilized in different fields of concentrates, for example, mental evaluation. The outcomes demonstrate high exactness that permits future improvement of utilizations that react to unconstrained outward appearances continuously [10].

[11]The point of this investigation is to reestablish some vision to the outwardly weakened individual through non-obtrusive advanced incitement. The goal is achieved by transmitting prepared data to the vision cortex, through non-obtrusive Trans-cranial Magnetic Stimulation (TMS) procedure to energize the basic unit of the sensory system. The nourishing of the vision neural framework is done through an Artificial Vision Neural Network (AVNN) of interconnected cells that produce pictures which help the vision impeded individual to perceive objects. The AVNN adjusts the prepared data to a structure appropriate for transmission. The said data is remotely sent by a variety of

terminals to the cerebrum; that is, a neuron to neuron correspondence is connected [11].

Programmed outward appearance acknowledgment is an intriguing and testing subject in flag handling, design acknowledgment, man-made brainpower, and so forth. In this paper, another strategy for outward appearance acknowledgment dependent on nearby paired examples (LBP) and neighborhood Fisher discriminant investigation (LFDA) is displayed. The LBP highlights are initially extricated from the first outward appearance pictures. At that point LFDA is utilized to deliver the low dimensional discriminative installed information portrayals from the extricated high dimensional LBP highlights with striking execution enhancement for outward appearance acknowledgment undertakings. At last, bolster vector machines (SVM) classifier is utilized for outward appearance arrangement. The test results on the prominent JAFFE outward appearance database show that the introduced outward appearance acknowledgment strategy dependent on LBP and LFDA acquires the best acknowledgment exactness of 90.7% with 11 decreased highlights, outflanking the other utilized strategies, for example, key segment investigation (PCA), direct discriminant examination (LDA), territory saving projection (LPP)[12].

In this paper, two novel techniques for outward appearance acknowledgment in facial picture arrangements are exhibited. The client needs to physically put some of Candide lattice hubs to confront tourist spots portrayed at the principal casing of the picture grouping under examination. The lattice following and misshapening framework utilized, in light of deformable models, tracks the network in sequential video outlines after some time, as the outward appearance develops, until the casing that compares to the best outward appearance force. The geometrical removal of certain chose Candide hubs, characterized as the distinction of the hub arranges between the first and the best outward appearance force outline, is utilized as a contribution to a novel multiclass Support Vector Machine (SVM) arrangement of classifiers that are utilized to perceive either the six essential outward appearances or a lot of picked Facial Action Units (FAUs)[13].

PC vivified operators and robots convey a social measurement to human PC collaboration and power us to ponder how PCs could be utilized in day by day life. Up close and personal correspondence is a constant procedure working at an a period scale in the request of 40 milliseconds. The dimension of vulnerability right now scale is impressive, making it vital for people and machines to depend on tactile rich perceptual natives as opposed to moderate emblematic derivation forms. In this paper we present advancement on one such perceptual crude. The framework naturally distinguishes frontal faces in the video stream and codes them as for 7 measurements progressively: nonpartisan, outrage, sicken, dread, happiness, misery, shock. The face discoverer utilizes a course of highlight locators prepared with boosting procedures [15, 2]. The appearance recognizer gets picture patches situated by the face identifier. A Gabor portrayal of the fix is shaped and after that handled by a bank of SVM classifiers. A tale mix of



Adaboost and SVM's upgrades execution. The framework was tried on the Cohn-Kanade dataset of presented outward appearances [14]

### Proposed Methods

This proposed work implements recognition of facial expressions using convolutional neural networks in which we use two approaches and compare the obtained results. Artificial Intelligence frameworks to perceive human expressions have pulled in much research interests and potential utilizations of such frameworks proliferate, spreading over spaces for example, client mindful advertising, observing health, what's more, emotionally keen mechanical interfaces. In light of the imperative job that outward appearance plays in imparting feeling in people, there has been generous research enthusiasm for computer vision frameworks to perceive human feeling. Certain outward appearances have widespread significance. In a 1971 paper titled "Constants Across Cultures in the Face furthermore, Emotion", Ekman et al. distinguished six outward appearances that are all inclusive over all societies: disgust, fear, happiness, anger, surprised, sadness. These are the same feelings that cutting edge outward appearance analysts mean to recognize utilizing PC vision. The ongoing accomplishment of convolutional neural systems (CNNs) in assignments, for example, object grouping reaches out to the issue of facial expression recognition. In the accompanying segments, we will display a diagram of our concern, a writing survey, and a report of our work.

### Approach1

1. In this project we are using the Tensorflow backend and Keras model to create the machine learning model and creating the convolution neural network.

2. At first we are creating the CNN model.

3. Convolution Neural Network(CNN) has basically 4 STEPS:

1. 2D Convolution
2. 2D Maxpooling
3. Flattening
4. Artificial Neural Network

4. First we are convolving the image using a 3X3 matrix and we are using the relu activation function.

5. Then we are Maxpooling the image and at last flattening the image.

6. This flattened array becomes the input layer of the ANN.

7. Now we are creating the dense layers and using softmax as the activation function.

8. Atlast, we are compiling the whole model.

9. Now, for the testing purpose, we are using the keras model again.

10. Now, we are using OpenCV to read an image and send the image to the CNN model to test the emotion.

11. According to the emotion, after feature extraction, we are predicting the emotion, and till now, the maximum accuracy we achieved is about 92%.

12. The picture is also displayed in the result.

- The images are converted to grayscale for detailed extraction.
- For the filtering of edges and pattern detection we will be

using 2D-Convolutional filtering.

### Approach2

1. In this project we are using the Tensorflow backend and Keras model to create the machine learning model and creating the convolution neural network.

2. We are using tlearn module of python to introduce a high level API to train the neural network faster and in a very easy manner.

3. Numpy has also been used in this project for the proper functioning and also for the scaling and image processing of the live image.

4. At first we are creating the CNN model.

5. We are then using the keras model for the training of the neural network and also the input of the image using a live neural network.

6. Then we are Maxpooling the image and at last flattening the image.

7. At the end of the program using all the previously trained keras and tlearn python models the code is run.

8. The camera is switched on and it takes a live image and gives the output according to the change in the image.

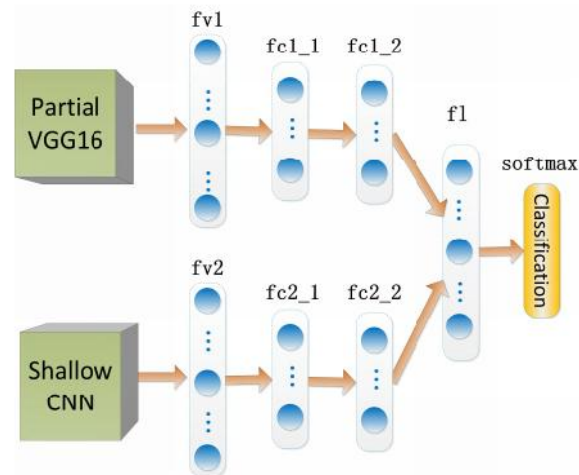


Fig 5: Weighted fusion network of outputs.

Fig. 5 demonstrates the proposed weighted fusion network. Expression related element vectors fv1 is removed from facial grayscale pictures utilizing the partial VGG16 coordinate with the adjusting procedure. Highlight vector fv2 is removed from LBP facial pictures utilizing the shallow CNN. Each component vector is associated with two cascaded full interface layers for dimension decrease. These full associate layers are  $fc1_1 = \{s1, s2, \dots, sm\}$  (m is experimentally to 100),  $fc1_2 = \{s1, s2, \dots, s6\}$  for fv1 and  $fc2_1 = \{l1, l2, \dots, lm\}$  (m = 100),  $fc2_2 = \{l1, l2, \dots, l6\}$  for fv2.

Weighted combination system of binary outputs. Facial highlights are consequently caught by the system and are uncovered through fc1\_2 and fc2\_2. Further, fc1\_2 and fc2\_2 are combined in a weighted manner to build an fused vector  $f1 = \{p1, p2, \dots, p6\}$ .

## FACIAL EXPRESSION RECOGNITION SYSTEM USING CONVOLUTIONAL NEURAL NETWORKS

The  $i$ th element  $p_i$  can be calculated as follows:

$$p_i = \alpha \cdot s_i + (1 - \alpha) \cdot l_i$$

where  $\alpha$  loads the commitments of facial grayscale pictures and LBP facial pictures to FER assignments;  $\alpha$  is determined tentatively by cross validation. Softmax order with a component of 6 (6 fundamental articulations) is utilized to perceive the given articulation dependent on the melded highlight vector. The softmax work creates a straight out likelihood conveyance, when the information is a lot of multi-class logits as:

$$Y_i = \frac{e^{x_i}}{\sum_{j=1}^K e^{x_j}} \text{ for } i = 1, \dots, K$$

where input  $x$  is  $K$ -dimensional vector and yield  $y$  is additionally a  $K$ -dimensional vector with genuine qualities in the range (0,1) and that signify 1 as standardization happens by means of the entirety of type terms that partition the real exponentiation term. The likelihood that the class  $y = k$  for a given info  $x$  and with  $j = 1, \dots, K = 6$  in our work) can be written in a framework structure as pursues:

$$\begin{bmatrix} P(y = 1|x) \\ \vdots \\ P(y = K|x) \end{bmatrix} = \begin{bmatrix} \zeta(x)_1 \\ \vdots \\ \zeta(x)_K \end{bmatrix} = \frac{1}{\sum_{j=1}^K e^{x_j}} \begin{bmatrix} e^{x_1} \\ \vdots \\ e^{x_K} \end{bmatrix}$$

where  $P(y = k|x)$  is the likelihood that the class is  $k$  given that the info is  $x$ . The cross entropy is utilized as the cost capacity, which is characterized as

$$Loss(y, z) = - \sum_{i=1}^K Z_i \cdot \log(y_i)$$

where  $z_i$  shows the genuine mark and  $y_i$  refers to the yield of softmax work. In the present examination, we utilize the backpropagation (otherwise known as backprops) in light of angle relative enhancement calculation.

The Convolutional Neural Network will contain an input layer, some convolutional layers, some dense layers (aka. fully-connected layers), and an output layer.

- The input layer has pre-determined, fixed dimensions, so the image must be **pre-processed** before it can be fed into the layer. The cropped face is then converted into grayscale and resized to 64-by-64 pixels. This step greatly reduces the dimensions.
- In convolutional layers, it passes into the layers where we specify the number of filters as one of the hyperparameters. **Convolution** generates feature maps that represent how pixel values are enhanced, for example, edge and pattern detection.

- The dense layer is inspired by the way neurons transmit signals through the brain. It takes a large number of input features and transform features through layers connected with trainable weights.
- Then we use **Back propagation** and it starts from evaluating the difference between prediction and true value, and back calculates the weight adjustment needed to every layer before. We can control the training speed and the complexity of the architecture by tuning the hyper-parameters, such as **learning rate** and **network density**.
- The sigmoid activation function is used. This output presents itself as a probability for each emotion class.

### Block Diagram

Fig 1. Convolutional Neural Network Architecture for facial emotion recognition(Approach 1)

Fig 2. Training and output of CNN for different facial expressions

Fig 3. Overview of the network architecture of the CNN using TFLearn ( Approach 2)

### RESULT

Fig 4. Graphical Comparison between two approaches adapted.

From this we can conclude that the final accuracy from approach 1 we used in solving the facial emotion recognition system using Convolutional Neural Network is 85 – 92% which is very much high compared to the approach we have taken to solve the facial emotion recognition system using CNN and tflern.

#### Series 2 – Convolution Neural Network

#### Series 1 – Convolutional Neural Network using tflern

Surprisingly, the second, much larger, network learns quickly as well, but converges to an accuracy of about 54%. Apparently reducing the network size breaks down the promising performance of the original network more than expected. Together with the much higher computational intensity, and therefore slower live performance, this model is not a worthy challenger of the other two architectures.

live emotion recognition through video is a standout amongst the most critical key-focuses in human-machine connection. To demonstrate the capacities of the got system, an application is built up that can legitimately process webcam film through the final show. With utilization of the previously mentioned OpenCV face acknowledgment program, the greatest showing up face from constant video is followed, extricated, and scaled to usable 48x48 info. This information is then bolstered to the contribution of the neural system demonstrate, which in its turn restores the estimations of the yield layer. These qualities speak to the probability

that the every feeling is portrayed by the client. The yield with the most noteworthy esteem is thought to be the present feeling of the client, and is portrayed by an emoji on the left of the screen.

## CONCLUSION

In this paper the proposed work uses Convolution Neural Network and Deep Learning technologies to detect the emotion of the face. We have used the Keras model, Tensorflow backend for the machine learning part to train and then test the dataset. Next we are using OpenCV and matplotlib to test a given input image. Thus achieving a decent accuracy in detecting the emotion of a human face using Convolution Neural Network. So, comparing the two approaches we have taken, the first approach using CNN gives us an accuracy of 92% whereas the other approach using CNN and TFlearn provides an accuracy of 54%. Hence, we can conclude that the first approach gives us better results as compared to the other.

## REFERENCES:

- [1] C. M. Lee, S. Yildirim, M. Bulut, A. Kazemzadeh, C. Busso, Z. Deng, S. Lee, & S. Narayanan, Emotion recognition based on phoneme classes, Proc. ICSLP, Jeju, Korea, 2004, 889-892.
- [2] T. Vogt & E. André, Improving Automatic Emotion Recognition from Speech via Gender Differentiation, Proc. Language Resources and Evaluation Conference, Genoa, Italy, 2006, 1123-1126.
- [3] D. Jiang & L. Cai, Speech Emotion Classification with the Combination of Statistic Features and Temporal Features, Proc. IEEE International Conference on multimedia, Taipei, Taiwan, China, 2004, 1967-1970.
- [4] B. Schuller, S. Reiter, R. Muller, M. Al-Hames, M. Lang, & G. Rigoll, Speaker Independent Speech Emotion Recognition by Ensemble Classification, Proc. IEEE International Conference on Multimedia and Expo, Amsterdam, the Netherlands, 2005, 864-867.
- [5] M. Kurematsu, J. Hakura, & H. Fujita, The Framework of the Speech Communication System with Emotion Processing, Proc. WSEAS International Conference on Artificial Intelligence, Knowledge Engineering and Data Bases, Corfu Island, Greece, 2007, 46-52.
- [6] F. Dellaert, T. Polzin, & A. Waibel, Recognizing Emotion in Speech, Proc. ICSLP, Philadelphia, PA, USA, 1996, 1970-1973.
- [7] J. Ang, R. Dhillon, A. Krupski, E. Shriberg, & A. Stolcke, Prosody-Based Automatic Detection of Annoyance and Frustration in Human-Computer Dialog, Proc. ICSLP, Denver, Colorado, USA, 2002, 2037-2040.
- [8] S. Yacoub, S. Simske, X. Lin, & J. Burns, Recognition of Emotions in Interactive Voice Response Systems, Proc. European Conference on Speech Communication and Technology, Geneva, Switzerland, 2003, 729-732.
- [9] J. Liscombe, Detecting Emotion in Speech: Experiments in Three Domains. Proc. HLT/NAACL, New York, NY, USA, 2006, 231-234.
- [10] S. Liu, Landmark detection of distinctive featurebased speech recognition. Journal of the Acoustical Society of America, 100(5), 1996, 3417-3430.
- [11] H.J. Fell & J. MacAuslan, Automatic Detection of Stress in Speech, Proc. of MAVEDA, Florence, Italy, 2003, 9-12.
- [12] Linguistic Data Consortium, "Emotional Prosody Speech", www ldc.upenn.edu/Catalog/CatalogEntry.jsp?catalogId=LDC2002S28, University of Pennsylvania.
- [13] R.P. Hobson, The autistic child's appraisal of expressions of emotion. Journal of Child Psychology and Psychiatry, 27, 1986, 321-342.
- [14] R.P. Hobson, The autistic child's appraisal of expressions of emotion: A further study. Journal of Child Psychology and Psychiatry, 27, 1986, 321-342.
- [15] D. Tantam, L. Monaghan, H. Nicholson, & J. Stirling (1989). Autistic children's ability to interpret faces: A research note. Journal of Child Psychology and Psychiatry, 30, 1989, 623-630.
- [16] K.A. Loveland, B. TUNALI-KOTOSKI, Y.R. Chen, J. Ortegón, D.A. Pearson, K.A. Brelsford, & M.C. Gibbs, Emotion recognition in autism: Verbal and nonverbal information, Development and Psychopathology, 9(3), 1997, 579-593.
- [17] A.L. Bacon, D. Fein, R. Morris, L. Waterhouse, & D. Allen, The responses of autistic children to the distress of others. Journal of Autism and Development Disorders, 28, 1998, 129-142.
- [18] M. Sigman, & E. Ruskin, Continuity and change in the social competence of children with autism, Downs syndrome, and developmental delays. Monographs of the Society for Research in Child Development, 64 (1, Serial No. 256), 1999
- [19] T. Ahsan, T. Jabid, and U.-P. Chong. Facial expression recognition using local transitional pattern on gabor filtered facial images. IETE Technical Review, 30(1):47-52, 2013.
- [20] P. Lucey, J. F. Cohn, T. Kanade, J. Saragih, Z. Ambadar, and I. Matthews. The extended cohn-kanade dataset (ck+): A complete dataset for action unit and emotion-specified expression. In Computer Vision and Pattern Recognition Workshops (CVPRW), 2010 IEEE Computer Society Conference on, pages 94-101. IEEE, 2010.
- [21] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei. Imagenet: A large-scale hierarchical image database. In Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on, pages 248-255. IEEE, 2009.