iJRASET
International Journal For Research in
Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH
IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 9 Issue: VI Month of publication: June 2021

DOI: https://doi.org/10.22214/ijraset.2021.34945

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.429
Volume 9 Issue VI Jun 2021- Available at www.ijraset.com

Emotion Detection from Facial Expressions using Deep Learning

Saraswan Sinha[1], Devansh Verma[2], Jaideep Sunil Chungi[3]
[1, 2, 3]Undergraduate Student, SRM Institute of Science and Technology, Chennai, India

Abstract: Emotion recognition is the process of detecting andclassifying human emotion. The aim of facial emotion recognition (FER) is to, recognize various types of human emotion from an individual's facial expression. The paper discusses the previous approaches and methods applied to perform emotion detection. A model has been developed using Convolution Neural Network for real time emotion detection. The accuracy of the model has been discussed andsuggested the means to further develop the model.
Keywords: Emotion, Computer Vision, Convolution Neural Network, Emotion detection through facial expression, Deep Learning, FER2013.

I. INTRODUCTION

Emotion detection is the process of recognizinghuman emotions from facial expressions. Automatic emotion recognition is one of the most largely researched topics. An emotional state of an individual can be deciphered mostly by the careful observation of changes in one's facial expressions, tone of voice, etc. A study conducted by Mehrabian proved that we gather 55% of the emotional information visually, 38% of the emotional information is interpreted vocally and 7% verbally. [6]. When a person changes the state of emotion, it first gets reflected in the change of his/her facial expression. Thus, mapping facial expressions with various emotional states has become an area of research interest. Scholarly works manifested into the creation of many databases with effective mapping of emotions and facial expressions. The use of many machine learning and deep learning models hasbeen discussed in the extant literature where accuracy scoresare found to be widely varying. In this paper, we have discussed a deep learning model that we have developed andchecked its accuracy using a database named FER2013. While developing the model, we have classified theemotions in seven states, namely neutral, happy, sad, surprise, fear, anger, disgust, and contempt.

While studying the facial expressions various features of the face are examined. Nose, eyes, the endpoint of the mouth, eyebrows, etc are some of many features which are used to detect the face in the image, and depending upon the variations of these features in one particular image helps in detecting the emotion. The challenge lies in extracting features from one face and classifying the facial expression into one particular state of emotion. Another challenge is to detect the same emotion from a different set of people as different people have different ways of expressing their emotions. As a result, the expressions may also vary based on different situations suchas the person's mood, skin color, age, culture, etc [2].

II. LITERATURE REVIEW

"Discovering Psychology" depicts emotion as "a complex psychological state that involves three distinct components: a subjective experience, a physiological response, and a behavioral or expressive response." [16] Apart from deriving a basic understanding of emotions, a plenty of researches have been conducted to understand the basic raw emotions that anyone may go through.

Paul Eckman (1972) classified emotions in six basic categories. They are "anger, surprise, happiness, sadness, fear, disgust". [10].

Robert Plutchik, a renowned psychologist, developed another form of the emotion classification system. This was termed as "emotion wheel" and also termed as the "Plutchik wheel". [11].

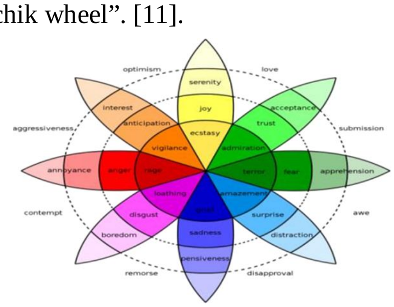

Figure 2.1: Plutchik wheel

This theory depicts that four sets of core emotions are being felt by people. Each set contains two contrasting emotions and hence they are being put in opposite pairs of wheels. Sadness and joy sits opposite to each other in the wheel. Similarly, anger and fear faces each other. The third set consist of expectation and surprise and the fourth set has the contrasting pair of disgust and trust. [11]

One research study used skin color range to detect the face region in the image. After face detection, the ratio of the widt of multiple regions in the human face is calculated to identify the facial features. Partitioning is doneusing the Aw-SpPCA algorithm. "Given as input any emotion of face, this pattern training set will classify the particular emotion". [12]

Another approach widely used is called the Geometric Approach. Components of the face are extracted and divided into various sections like the eye's section, mouth section, etc. Once these features are extracted, Euclidean distance between these sections is calculated, which in turn is used for the classification of emotions. [13]. Many other techniques are developed for facial emotion detection. All of them are mostly divided into 2 categories: Feature-based and View based. [14], [15]. Gabor filters that use the Appearance Based approach are used to classify emotions. These filters use a "multi-layer perceptron neural network to classify emotions." "Principal Component Analysis" is used for feature and dimensionality reduction.

Some of the pre-exisiting work done on the topic are mentioned below:

Reference and year	Approach and Method	Performance
Wei-Long Zheng and Bao-Liang Lu (2016)	EEG-based affective models without labeled target data using transfer learning techniques (TCA-based Subject Transfer)	Positive (85.01%) emotion recognition rate is higher than other approaches but neutral (25.76%) and negative (10.24%) emotions are often confused with each other.
Zixing Zhang, Fabien Ringeval, Fabien Ringeval, Eduardo Coutinho, Erik Marchi and Björn Schüller (2016)	Semi-Supervised Learning (SSL) technique	Delivers a strong performance in the classification of high/low emotional arousal (UAR = 76.5%), and significantly outperforms traditional SSL methods by at least 5.0% (absolute gain).
Y. Fan, X. Lu, D. Li, and Y. Liu. (2016)	Video-based Emotion Recognition Using CNN-RNN and C3D Hybrid Networks	Achieved accuracy 59.02% (without using any additional Emotion labeled video clips in training set) which is the best till now.
A. Yao, D. Cai, P. Hu, S. Wang, L. Shan and Y. Chen (2016)	HoloNet: towards robust emotion recognition in the wild	Achieved mean recognition rate of 57.84%.
Yelin Kim and Emily Mower Provos (2016)	Data driven framework to explore patterns (timings and durations) of emotion evidence,	Achieved 65.60% UW accuracy, 1.90% higher than the baseline.

Table 1: Various methods for FER [5]

A. Systems and Methods

1) *Deep Learning:* Machine Learning has various branches, deep learning is one of them.Artificial neural networks is theunitary part of deep learning. For deep learning models to perform with high levels of accuracy, large datasets are required. These deep learning models learn from the data fed to them. These deep learning models automatically extract necessary features from the input data. Typically, deep learning models, in terms of accuracy, perform more efficiently than machine learning models but require more data and time to be trained.
2) *Database:* The reason deep learning models are considered more efficient than and generally preferred over machine learning models is because of the presence of neural networks cascading one after the other. These cascaded neural networks account for automatic feature extraction. These neural networks need to be trained on large datasets. Several FER databases are now publicly available to researchers for creating and training their model based on neural networks. Some are presented in Table 2.

Databases	Descriptions	Emotions
MultiPie [10]	More than 750,000 images captured by 15 view and 19 illumination conditions	Anger, Disgust, Neutral, Happy, Squint, Scream, Surprise
MMI [11]	2900 videos, indicate the neutral, onset, apex and offset	Six basic emotions and neutral
GEMEP FERA [12]	289 images sequences	Anger, Fear, Sadness, Relief, Happy
SFEW [13]	700 images with different ages, occlusion, illumination and head pose.	Six basic emotions and neutral
CK+ [14]	593 videos for posed and non-posed expressions	Six basic emotions, contempt and neutral
FER2013 [15]	35,887 grayscale images collect from google image search	Six basic emotions and neutral
JAFFE [16]	213 grayscale images posed by 10 Japanese females	Six basic emotions and neural
BU-3DFE [17]	2500 3D facial images captured on two view -45°, +45°	Six basic emotions and neutral
CASME II [18]	247 micro-expressions sequences	Happy, Disgust, Surprise, Regression and others
Oulu-CASIA [19]	2880 videos captured in three different illumination conditions	Six basic emotions
AffectNet [20]	More than 440.000 images collected from the internet	Six basic emotions and neutral
RAFD-DB [21]	30000 images from real world	Six basic emotions and neutral

Table 2: Description of available Databases [7]

ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.429
Volume 9 Issue VI Jun 2021- Available at www.ijraset.com

Images in grayscale of faces having size 48x48 pixel constitutes the FER2013 data set. The portion of the faces inthe images are positioned in a manner that they are almost centered, all the faces have almost equal occupancydimension wise in the image and they can be detected automatically. As discussed above, the objective is to detect one of the seven facial expressions and thereby classifying the image under any one of the seven emotions (happy, sad, angry, afraid, surprise, disgust, and neutral).

“train.csv” mainly depicts two fields, namely “emotion” and “pixels”. The column containing “emotion” classifies seven types of emotions shown in the image and they are mapped through numbers 0 to 6. The other field ie, pixel contains double-quoted strings for each image. These stringsare made up of pixel values separated by space where order wise row is major. “test.csv” containing a single column of pixel values and emotions are being predicted from these pixel values. ”train.csv” contains 28,709 image pixel values. This data set, actually a preliminary version of a researchproject database had been used earlier in a workshop with thedue consent of the researchers Pierre-Luc Carrier and Aaron Courville. [8]

III. METHODOLOGY

The process of emotion recognition involves processing images and detecting faces in these images.Once a face has been detected, various facial features are extracted by the model. Facial Expression Recognition consists of thefollowing five steps [6]:-

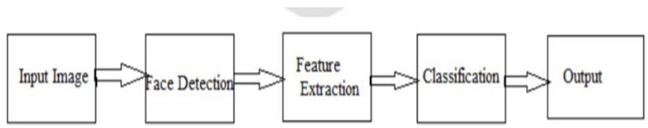

Fig. 1: General step of emotion recognition

A. Face Detection

The process involves detecting faces from an image. Different images have different scales and lighting conditions. This is accomplished using Haar cascade frontal face classifier. This step is required because images can have different lighting conditions and scales. Face expression recognition tends to fail if the test image is not similar to the training images. Hence, instead of feeding the entire image to a neural network, it is preferred to only feed that portion of the image which has the face in it.

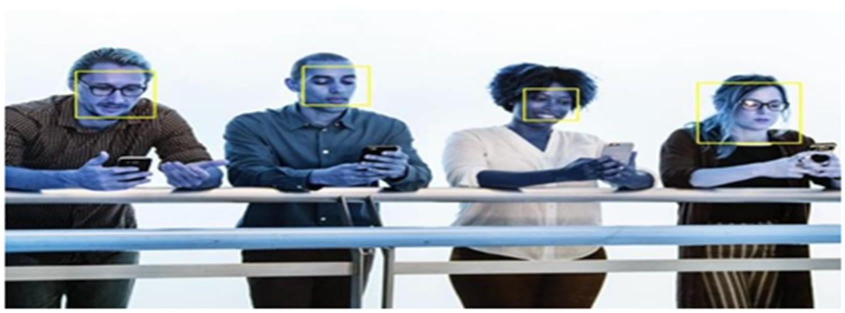

Fig.2: Example of facial detection

B. Facial feature Extraction

Many applications of the field computer vision extracts landmark features of the face at the preliminary level of execution. This being one of the most crucial part of the application bears high importance for the successful execution of these applications. Extracting the typical features of the face is thus bears high importance for achieving the desired success rate. [4]Facial feature extraction is identifying the landmark feature of the face which includes eyes, mouth, nose, etc.

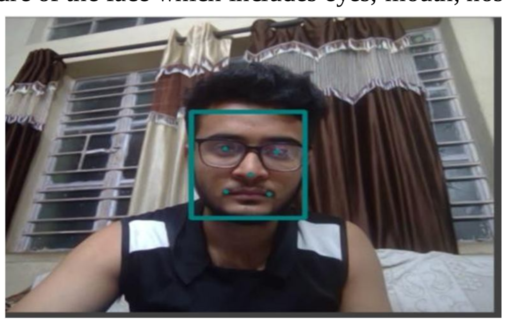

Fig. 3: Landmark feature extraction

C. Emotions Classification

The set of features that are extracted are used in theclassification stage. The mean distance between the facial features is used to determine the emotion and random valuesare set as thresholds for each emotion. During training, the model learns from different input images that it is provided with and sets the threshold accordingly.

As we have already discussed, a deep learning model has various neurons in it. These neurons have their own set of weights associated with them. When the model isin its training phase, it sets different weights for different neurons.

As the model is fed with more input images, these weights are updated. Those set of weights with which the model has the highest accuracy score is retained.

Fig.3: Example of Emotion Classification

D. Model

The model has 2 set of convolutional layers and 2 Fully connected layers.

Each set contains 2 convolutional layers having 32 neurons and 64 neurons respectively. A dropout of 0.25 is added and “relu” is used as an activation function. Thedropout ensures there is no overfitting. The value 0.25 indicates that one-fourth of the neurons are randomly turned off. Max Pooling is used to downsample the input, by carrying forward the largest information available amplitude-wise. On the event of the kernel focusing on the first 2*2 Grid of the input image which has say distinct values 3,5,7 and 9. The highest among these four numbers, which is 9 in this case will be picked up by the process of Max pooling. [9]

There are two Fully-Connected (FC) layers. The first fully connected layer has 1024 neurons with “relu” as the activation function with a drop out of 0.25 and the second FC layer has 7 neurons. Each neuron turns on when emotion is detected. The latter is the output layer of our model and gives out a value in [0,1,2,3,45,6]. Since this layer does the classification softmax activation function is added to it[3].

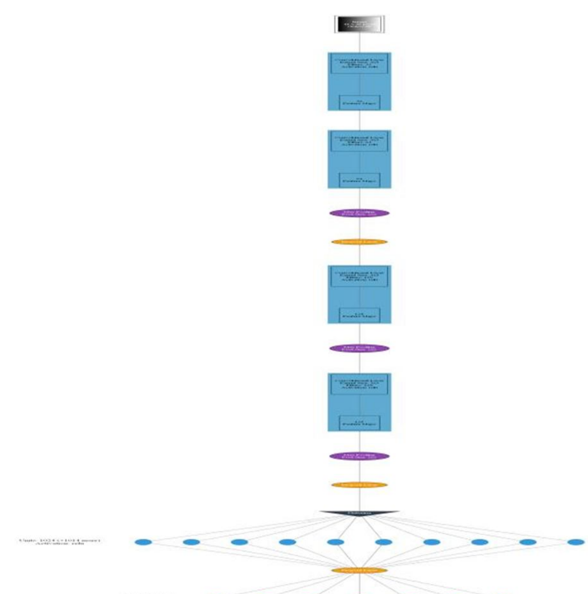

Fig.4: Model visualization

IV. FINDINGS AND ANALYSIS

This section discusses about the performance of the model on testing data. Here we provide a detailed report of the training and testing accuracy of our classifier on various images.

A. Training and Testing Accuracy

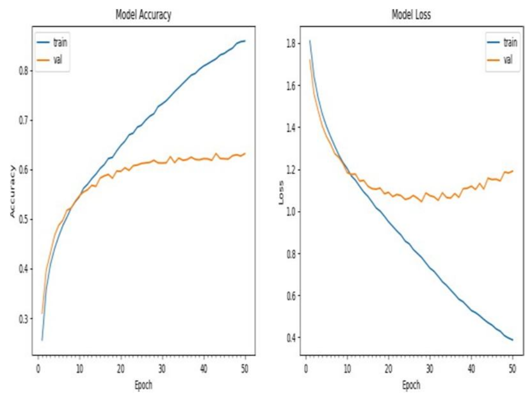

Fig. 5 : Accuracy and loss plot

The model shows an accuracy of over 80% on the training data and 63.2% in 50 epochs on the testing data.Theloss on the data falls significantly below 0.4 on the training data however on the testing data the model seems to have been overfitted with the loss hovering around 1.0-1.2 from epoch 20-50.

B. Model Performance

The model was the most successful in detecting the emotion states of happyness, neutral, and sadness. We infer that the availability of a large number of images associated with smiling, normal, and sad states is the reason for the high rate of success. However, it did not do very well in detecting the emotion of disgust. lesser number of training images mapped to the class disgust may have manifested this. The model could, however, detect the sad and fearful faces with acceptable rates of accuracy. However, the response in some cases were wrongly showing as surprise initially and then were changing to fear. People have different ways of facially expressing themselves to show fear or surprise may have resulted into this momentary error.Another plausible cause could be that the mean distance between the landmark features in these cases must be very close to each other.

C. Applications of FER

FER is used in the car board system. This detects the emotional state of the driver which can be analyzed and used to ensure the safety of the driver and the passengers. This technology has been adopted in various domains of works in industry. Unilever has explored the possibilities of judging the overall confidence level of an interviewee by using this technology. This information helps them in screening the candidate in more efficient manner. [1]

While video games are in the trail phase, the facial expression of the gamers are being observed. The emotions thus captured real time. This feedback along with other observations help in designing the final product.

V. CONCLUSION

Among the various available classifications of emotions we have chosen the one proposed by Paul Eckman. The seven basic emotional states has been depicted in the model. In accordance with that the database chosen to work with was FER2013. We developed a deep learning model which successfully detected the emotions and classified them in seven categories on real time basis. The accuracy rate achieved by the model was 63.2% which is acceptableas per given standards. The areas where the model performed satisfactorily and the events where there was scopes of improvement has been discussed in the result section.

The FER on real time basis is becoming very prevalent in real life situations like car board system, interview processes, designing of video games, etc. Our model thus offers a feasible alternative to the available methods. We propose further research to improve upon the success rate and eliminate certain bugs. Futher to this an area of improvement was found for the database that has been used by increasing the sample size of facial images mapped to the emotional states of fear, surprise and disgust.

REFERENCES

[1] Picard, R., 1997. Affective computing. cambridge,massachustes institure of technology.
[2] Calvo, R.A. and D'Mello, S., 2010. Affect detection: An interdisciplinary review of models, methods, and their applications. IEEE Transactions on affective computing, 1(1), pp.18-37.
[3] Greco, A., Valenza, G. and Scilingo, E.P., 2016. Advances in Electrodermal activity processing with applications for mental health. Berlin, Germany: Springer.
[4] Zhao, L., Huang, H., Li, X., Ding, S., Zhao, H. and Han, Z., 2019. An accurate and robust approach of device-free localization with convolutional autoencoder. IEEE Internet of Things Journal, 6(3), pp.5825-5840.
[5] Khan, R. and Sharif, O., 2017. A literature review on emotion recognition using various methods. Global Journal of Computer Science and Technology.
[6] Marechal, C., Mikolajewski, D., Tyburek, K., Prokopowicz, P., Bougueroua, L., Ancourt, C. and Wegrzyn-Wolska, K., 2019. Survey on AI-Based Multimodal Methods for Emotion Detection.
[7] Mellouk, W. and Handouzi, W., 2020. Facial emotion recognition using deep learning: review and insights. Procedia Computer Science, 175, pp.689-694.
[8] Challenges in Representation Learning: FacialExpression Recognition Challenge: Learn facial expressionsfrom an image, viewed on 30 April, 2021,https://www.kaggle.com/c/challenges-in-representation- learning-facial-expression-recognition-challenge/data
[9] Plutchik, R., 1984. In search of the basicemotions. PsycCRITIQUES, 29(6), pp.511-513.
[10] Alom, M.Z., Piao, M.L., Alam, M.A., Kim, N. and Park, J.H., 2012. Facial Expressions Recognition from Complex Background using Face Context and Adaptively Weighted sub-Pattern PCA. World Academy of Science, Engineering and Technology, 6, pp.3-20.
[11] Agushinta, D., Suhendra, A., Madenda, S. and Suryadi, H.S., 2011. Face component extraction using segmentation method on face recognition system. Journal of EmergingTrends in Computing and Information Sciences, 2(2), pp.67-72.
[12] Tian, Y.L., Kanade, T. and Cohn, J.F., 2000, March. Recognizing lower face action units for facial expression analysis. In Proceedings Fourth IEEE International Conference on Automatic Face and Gesture Recognition(Cat. No. PR00580) (pp. 484-490). IEEE.
[13] Sunita, R. and Sudipta, R., 2012. Prof. Bandyopadhyay Samir K. A tutorial review on face detection. On International Journal of Engineering Science and Technology (IJEST), 1(8).
[14] Pantic, M. and Patras, I., 2006. Dynamics of facialexpression: Recognition of facial actions and their temporal segments from face profile image sequences. IEEETransactions on Systems, Man, and Cybernetics, Part B (Cybernetics), 36(2), pp.433-449.
[15] Lajevardi, S.M. and Lech, M., 2008, December. Facial expression recognition using neural networks and log-gabor filters. In 2008 Digital Image Computing: Techniques and Applications (pp. 77-83). IEEE.
[16] Shin, D. and Wang, Z., 2015. The experimentation of matrix for product emotion. Procedia Manufacturing, 3, pp.2295-2302.