International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 07 Issue: 12 | Dec 2020 www.irjet.net p-ISSN: 2395-0072

Real-time Facial Emotion Detection using CNN Architecture and Relevant Recommendations based on Results

Rutuja Tambe[1], Anmol Thete[2], Aniket Gadhave[3], Sumedh Dhengre[4]

[1-4]*Department of Computer Engineering, All India Shri Shivaji Memorial Society's College of Engineering, Pune-01, Savitribai Phule Pune University, Maharashtra, India.*

Abstract - *The interaction between humans and computers will be more natural if computers are able to grasp and acknowledge to human non-verbal communication such as emotions. Although several approaches have been put forth to recognize human emotions based on facial expressions or speech or textual data, it is necessary to enhance the accuracy and robustness of the emotion recognition system. The solution furnished in this paper will not only detect the emotion, but also provide relevant recommendations based on the obtained results. Using real-time images of human faces, several emotions will be classified such as sadness, happiness, anger, fear, surprised, and neutral state. Detailed facial motions will be captured, and appropriate emotion will be detected by using deep learning algorithm such as CNN. To determine the best classifiers for recognizing particular emotions, single and multi-layered networks will be tested. Different resolutions of the images representing faces as well as the images including regions of mouth and eyes will be included. On the basis of the test results, a cascade of the neural networks will be proposed. The cascade will recognize six basic emotions and neutral expression. Depending upon the resultant emotion, the module will suggest songs or suitable tasks, to cheer up a person and enhance his/her mood. Sentiment analysis through CNN will help suggesting some tasks associated to a particular mood. Hence, it will help a person to stay optimistic and happy. The results reveal that the system based on facial expression gives much superior performance than the system based on textual data, or just acoustic information for the emotions considered.*

Key Words: Deep Learning, CNN, Facial Emotion Recognition, Human Computer Interaction, feature extraction, stress detection, Cascade, Classification.

1. INTRODUCTION

A person goes through various emotions depending on situations and the people they interact with and hence may have mood fluctuations which may affect their daily lifestyle. People are experiencing constant emotional dilemma which hampers their productivity and personal growth. As emotions are nothing but the state of mind, they are closely related to the work we do. So it's just a matter of small tasks or activities that can change a person's state of mind.

Inter-personal human communication includes not only spoken language but also non-verbal cues such as hand gestures, facial expressions, which are used to express feeling and give feedback [6]. They also help us understand the intentions of others [7]. According to different surveys [8][9], verbal components convey one-third of human communication, and nonverbal components convey two-thirds. The need for Facial emotion recognition is increasing with the rapid development of artificial intelligent techniques, including in Human-Computer Interaction (HCI) [10][11], Virtual Reality (VR) [12], Augmented Reality (AR) [13], Advanced Driver Assistant Systems (ADAS) [14], and entertainment [15][16].

Humans can detect these signs even if they are being manipulated, by concurrently considering the information obtained by the ears and eyes. If computers can detect these emotional inputs, they can provide clear and appropriate assistance to users in ways that best suit the user's needs and preferences. It is extensively accepted from psychological theory that human emotions can be segregated into seven archetypal emotions: surprise, fear, disgust, anger, happiness, sadness or neutral [6]. Facial motion plays a crucial role in expressing these emotions. The muscles of the face can be changed and can be deliberately altered to communicate different feelings.

Hence, through their facial expressions we can detect emotions of a person and provide them appropriate suggestions to enhance their mood. This paper aims to present an AI/ML module that takes image of a human face as input and returns the emotion that the face expresses. To determine the best classifiers for recognizing particular emotions, single and multi-layered networks will be tested. Different resolutions of the images representing faces as well as the images including regions of mouth and eyes will be included. If computers can detect these emotional inputs, they can provide clear and appropriate assistance to users in ways that best suit the user's needs and preferences, along with keeping them happy and content.

2. LITERATURE SURVEY

A detailed study on the facial emotion recognition is discussed in [1] which exposes the properties of a dataset and emotion detection study classifier. Visual attributes of the image are explored and some of the classifier

techniques are discussed in [2] which will be thoughtful in further assessment of the methods of emotion recognition. Neural network emerges extremely, which strives to decode problems in data science. Facial emotion Recognition is drawing its own significance in the research field. Facial emotion recognition is inspected and analyzed on all research areas [3].

Emotion is identified from facial images using filter banks and Deep CNN [4] which gives high accuracy rate with which we had an inference that deep learning can also be used for emotion detection. Facial emotion recognition can be also performed using image spectrograms with deep convolutional networks which is implemented in [5]. Emotion recognition is also possible using different features such as, speech or bi-modal systems, which use both voice and facial expressions [6]. Some researchers have proposed that system should induce and conduct the learner's emotions to the suitable state [7]. But, primarily, the learner's emotions have to be perceived by the system.

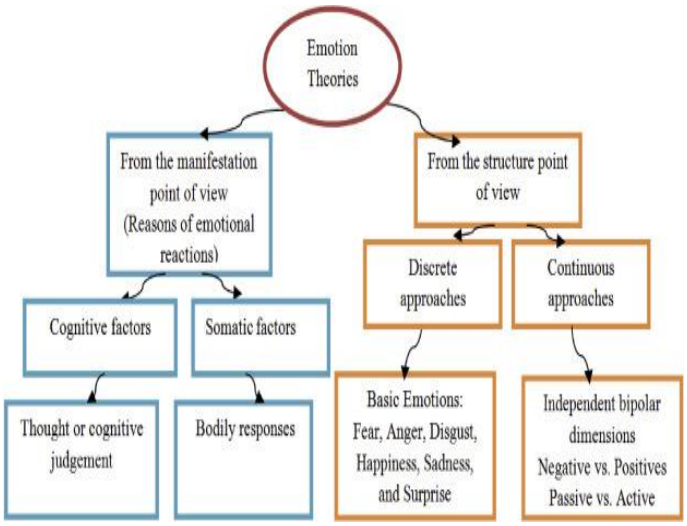

Fig -1: Survey of Emotion Detection Methods [7]

Emotions can be detected by interrogating the user, tracking the unexpressed parameters, voice recognition, facial recognition, important signs and signals or gesture recognition [7]. In addition, hybrid methods have been proposed, which use two or more methods in combination with various emotional cues. This [8] review also throws light on an up-to-date composite deep-learning perspective, combining a convolutional neural network (CNN) for the spatial features of an individual frame and long short-term memory (LSTM) for temporal features of consecutive frames.

3. METHODOLOGY

3.1. Concept

Facial expressions give important clues about emotions. The features used are usually based on local spatial location or displacement of specific points and surface regions, in contrast to sound-based methods, which use global calculations for acoustic features [6]. The face of a person will be captured, and different features will be detected. The facial expression will be recognized using spatial (frame based) and spatio-temporal (sequence based) templates. The aim of this project is to detect the most appropriate human emotion through facial expressions as output, by providing real-time images of a person's face as an input. Also recommend a relevant task or song to enhance the person's mood and relieve them from stressful situations. This will help reduce emotional dilemma and increase the productivity of people.

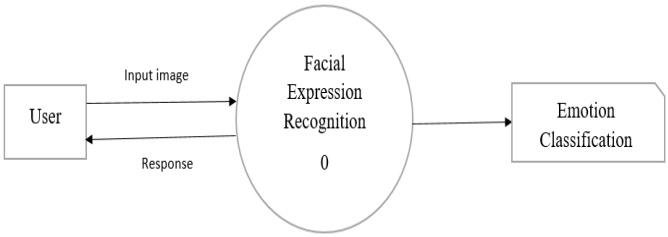

Fig -2: Conceptual Diagram

3.2. Proposed Work

Detailed facial motions will be captured, and appropriate emotion will be detected by using deep learning algorithm such as CNN. The system will determine the best classifiers for recognizing particular emotions, where single and multi-layered networks will be tested. Different resolutions of the images representing faces as well as the images including regions of mouth and eyes will be included. On the basis of the test results, a cascade of the neural networks will be proposed. The cascade will recognize six basic emotions and neutral expression. Depending upon the resultant emotion, the module will suggest songs or suitable tasks, to cheer up a person and enhance his/her mood.

Sentiment analysis through CNN will help suggesting some tasks associated to a particular mood. Hence, it will help a person to stay optimistic and happy. Interaction implemented between user and the system will be stored in the database. This data will be used in implementing collaborative filtering using CNN algorithm. The result of this will be shown to the user through reliable user interface. This application will not only detect the person's mood, but also provide them with suitable recommendations based on the output. Our system gets trained with dataset "*fer2013.csv*", with every epochs it increases systems accuracy. Then serialize the model to JSON and save the model weights in an hd5 file so that we can make use of this file to make predictions rather than training the network again.

3.3. System Design

The user interface includes a login window as soon as the application starts with a text box for user-name and a password box for the password. User can access the system through android and web application also. Then the user will be prompted to capture a selfie, and after the image is captured, it will be preprocessed and the features extracted will be stored into the image database. These features obtained will then be sent to the trained neural network, which will predict features and use them to detect the emotion and obtain the results. Based on these results, the system will provide relevant recommendations to the user. The user will then find some recommended tasks or some videos on his screen, as per the resultant mood, in order to improve their mood.

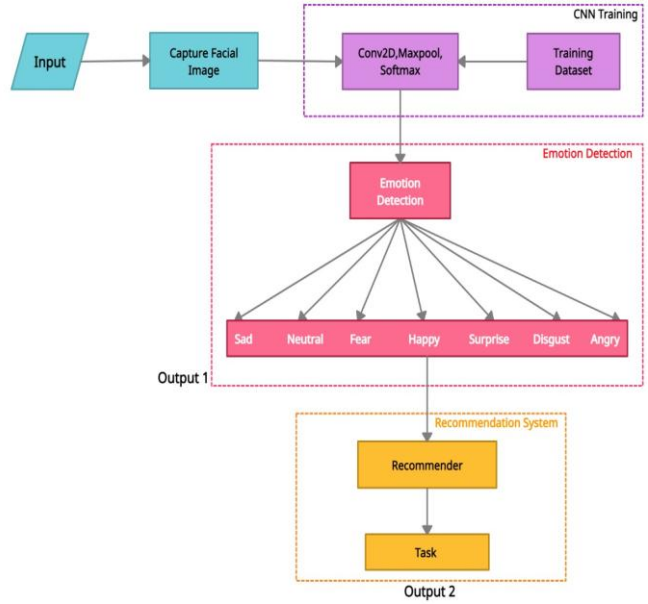

Fig -3: System Architecture

The general system design comprises of the following modules:

a) Input Image

b) Training using Convolutional Network

c) Analysis and Classification

d) Recommendation kernel

e) Result

First of all when the user registers into the application, it requests the user to capture a selfie. This image is then divided into different sections of the face, such as forehead, eyebrows, lower eye, right check and left check. After all the pre-processing is done, then with Convolutional neural network it trains the given dataset and with every epoch accuracy increases. Then with the user's image it detects emotion and give accordingly suggest tasks to change mood of sad, depressed person.

4. ALGORITHM USED

The Convolutional Neural Network (CNN) is a Deep Learning algorithm that can capture images, assign significance value (readable and discriminatory metrics) to the various aspects / elements in the image and are capable to discriminate one from the other. The previous processing required in the Convolutional Network is very trivial compared to other classification algorithms. While ancient filtering methods are hand-engineered, with adequate training, Convolutional Networks have the ability to study these filters / features.

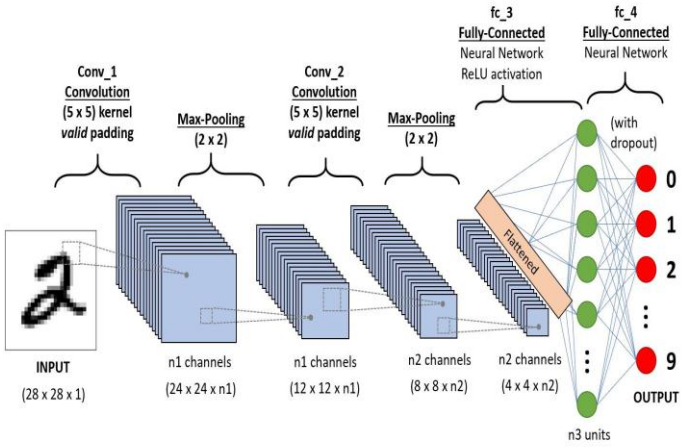

Fig -4: CNN Architecture [18]

The structure of CNN is similar to the connectivity patterns of neurons in the Human Brain and was inspired by the Visual Cortex organization [18]. Individual neurons respond to stimuli only in the restricted region of the viewing field known as the Reception Field. The collection of such fields pierces to cover the entire visible surface.

Convolutional Network is proficient in capturing Spatial and Temporary dependencies on an image using relevant filters [18]. The architecture creates a better balance in the image database due to the reduction in the number of parameters involved and the reuse of weights. In other words, the network can be trained to better understand image perception.

5. MATHEMATICAL MODEL

Let the proposed system be described by the set theory as follows:

S = {DC, REC, S, F, O, I, Q, q0, qf, NDD, D}

Where,

I = input: {image of user}

O = output: {emotion detection and recommendation}

q0 = initial state: {system starts and asks for user input image}

qf = final state: {system displays the resultant mood and recommendations}

S = success: {If the system works accurately without any halt or error.}

F = failure: {System halts due to some error or doesn't predict accurately.}

Q = set of states: {q0, q1, q2, ….. , qf}

D = deterministic data: {Null}

NDD = non deterministic data: {All states resulting output is non deterministic}

DC = data collection: {q : already trained dataset.}

REC = recommendation: {q : results in expected output.}

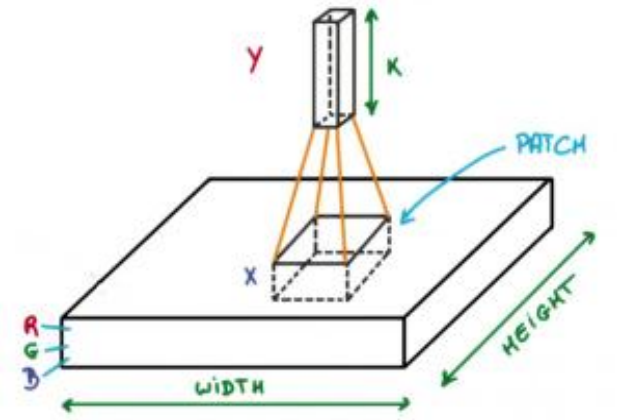

Fig -5: Learnable Filters (patch) in CNN [19]

Mathematics involved in the whole convolution process comprises of the following [19]:

1. Convolution layers contain a collection of readable filters (the patch above). Each filter has a minimum width, height and depth equal to the input volume (3 if the input layer is image input).
2. For example, if we have to apply convolution to an image in size 34x34x3. The possible size of the filter can be a x a x 3, while the 'a' can be 3, 5, 7, etc. but relatively smaller in size.
3. During progression, we load each filter at all volume input distances step by step where each step is called a stride (which can values as 2 or 3 or even 4 in larger images with large dimensions) and calculate the dot product between the weights of the filters and patch from the input volume.
4. As we load our filters we will get 2-D output from each filter and we will blend them together and as a result, we will get a volume output equal to the number of filters. The network will read and memorize all the filters.

6. RESULTS

The main component of this paper is the module that we are designing to identify a person's mood and help him enhance it, if he is sad or stressed. It is an Android based system with python and AI module in background. Beyond that, the application is a self-contained unit and will not rely on any other Desktop O.S related software components or any extra storage space. This application will be interacting with user when user registers on it through a reliable user interface. For the system to be more precise and accurate with results, modern deep learning algorithms like Neural networks will be used for classification.

The application doesn't affect any other features on the machine and no hardware other than camera is used. The application will be able to detect emotions of a person through his facial expressions, quickly and efficiently. Application will help business professionals of any company to manage stress levels. This will create a healthy work environment and surge productivity. Application will provide relevant recommendations to the user based on the obtained result, so that their mood will be enhanced. The CNN algorithm used will work efficiently with the used operating system, to provide much accurate results.

7. ADVANTAGES

- User Friendly.
- Access to authorized personnel only.
- Detect emotion instantly and obtain personalized recommendations easily using internet/mobile.
- Does not require any additional expensive hardware to adopt.
- It helps employees and HR (Human Resource) team of any company to manage stress levels.
- This will create healthy work environment and increase productivity.
- HR and management will also bring back the positive and negative feelings of employees and customers who help businesses grow.

8. DRAWBACKS

- Validation of emotion dataset is a challenge in order to have accurate emotion recognition system.
- User may get irrelevant recommendations.
- Absence of internet connection may affect the performance.
- Redundant data may exist on multiple iterations.

9. APPLICATIONS

- Human behavior understanding for counselors
- Emotional stereotypes of learners
- First Impression test for web-page usability
- To answer the question of whether and how emotional state affects the productivity of a person.
- The emotional reaction on advertisements will be calculated in response to numerous formats of online ads.
- In website customization and gaming platforms

10. CONCLUSION

Emotion sensing technologies are evolving at this stage. Once they are fully-fledged, advantages of this technology will assist mankind in various aspects of life. The most beneficial is that it will avoid accidents or tragedies resulting due to driver's state of mind, such as frustration, stress, depression etc. Our goal was to recognize current mood of users based on their facial expression and then to provide appropriate recommendations so that they can enhance their mood and have a good time. The system attempts to provide stress-relief videos or tasks, in order to calm the person in distress, or anger.

It also recommends motivational content for users, to overcome depression and sadness and stay delighted. This research is meant to offer a brief solution that is faster, easier and concise. With the help of simple UI like Android app it can take picture and then perform relevant tasks, which won't irritate the user also. This system is more reliable because it uses modern Deep learning technique like Convolutional Neural Network which increases accuracy. This will be of major help in stressful situations and plays a crucial role to spur the graph of a person's productivity.

11. FUTURE SCOPE

Facial Expression Recognition (FER) will be an important topic in the field of computer vision and artificial intelligence due to its crucial educational and commercial potential [7]. The more challenging task of capturing salient visual information directly from conventional videos is a topic for future work in this research. In future, platform will be able to capture and detect facial expressions through small videos as input.

ACKNOWLEDGEMENT

We would like to acknowledge the partial support of our project guide Prof. S. G. Dhengre, serving as a Professor in the All India Shri Shivaji Memorial Society's College Of Engineering, Pune-01. He provided proper guidance time to time and his suggestions stood noteworthy in making this paper complete.

REFERENCES

[1] G. Hintin, Greves and A. Mohemed, "Emotion Recognition with Deep Recurrent Neural Networks," in 2013 IEEE International Conference on Acoustics, Speech and Signal Processing, 2013, pp. 6645–6649.

[2] K. Weint and C.-W. Huaang, "Characterizing Types of Convolution in Deep Convolutional Recurrent Neural Networks for Robust Speech Emotion Recognition," pp. 1–19, 2017.

[3] A. Routrey, M. Swaen and P Kabisetpathy, "Database, features and classifiers for emotion recognition: a review," Int. J. Speech Technology, vol. 0, no. 0, p. 0, 2018.

[4] K. Y. Hueng, C. H. Wiu, T. H. Yieng, M. H. Sha and J. H. Chiu, "Emotion recognition using auto-encoder bottleneck features and LSTM," in 2016 International Conference on Orange Technologies (ICOT), 2016, pp. 1–4.

[5] M. N. Sttilar, M. Leich, R. S. Bolie and M. Skinter, "Real time emotion recognition using RGB image classification and transfer learning," in 2017 11th International Conference, Signal Processing Communication Systems, pp. 1–8, 2017.

[6] C. Busso, Z. Deng, S. Yildirim, M. Bulut, C. M. Lee, A. Kazemzadeh, S. Lee, U. Neumann, S. Narayanan, "Analysis of Emotion Recognition using Facial Expressions, Speech and Multimodal Information" in 2004, ICMI'04, October 13–15, 2004, State College, Pennsylvania, USA.

[7] M. Imani, G. A. Montazer, "A survey of emotion recognition methods with emphasis on E-Learning

environments", in Journal of Network and Computer Applications, 2019.

[8] Byoung Chul Ko, "A Brief Review of Facial Emotion Recognition Based on Visual Information", in 2018, in MDPI journal.

[9] A. Mehrabian, "Communication without words", in Psychology Today 1968, 2, 53–56.

[10] K. Kaulard, D. W. Cunningham, H. H. Bülthoff, C. Wallraven, "The MPI facial expression database — A validated database of emotional and conversational facial expressions", in PLoS ONE 2012, 7, e32321.

[11] F. Dornaika, B. Raducanu, "Efficient facial expression recognition for human robot interaction", In Proceedings of the 9th International Work-Conference on Artificial Neural Networks on Computational and Ambient Intelligence, San Sebastián, Spain, 20–22 June 2007; pp. 700–708.

[12] C. Bartneck, M. J. Lyons, "HCI and the face: Towards an art of the soluble", In Proceedings of the International Conference on Human Computer Interaction: Interaction Design and Usability, Beijing, China, 22–27 July 2007; pp. 20–29.

[13] S. Hickson, N. Dufour, A. Sud, V. Kwatra, I. A. Essa, "Eyemotion: Classifying facial expressions in VR using eye-tracking cameras", arXiv 2017, arxiv:1707.07204.

[14] C. H. Chen, I. J. Lee, L. Y. Lin. "Augmented reality-based self-facial modeling to promote the emotional expression and social skills of adolescents with autism spectrum disorders", in Res. Dev. Disabil. 2015, 36, 396–403.

[15] M. A. Assari, M. Rahmati, "Driver drowsiness detection using face expression recognition", In Proceedings of the IEEE International Conference on Signal and Image Processing Applications, Kuala Lumpur, Malaysia, 16–18 November 2011; pp. 337–341.

[16] C. Zhan, W. Li, P. Ogunbona, F. Safaei, "A real-time facial expression recognition system for online games" in International Journal of Computer Games Technology. 2008.

[17] A. Mourão, J. Magalhães, "Competitive affective gaming: Winning with a smile", In Proceedings of the ACM International Conference on Multimedia, Barcelona, Spain, 21–25 October 2013; pp. 83–92.

[18] https://towardsdatascience.com/a-comprehensive-guide-to-convolutional-neural-networks-the-eli5-way-3bd2b1164a53

[19] https://www.geeksforgeeks.org/introduction-convolution-neural-network/

BIOGRAPHIES

RUTUJA TAMBE,
Student at AISSMS COE, Pune-01.
Savitribai Phule Pune University,
B.E. in Computer Science.

ANMOL THETE,
Student at AISSMS COE, Pune-01.
Savitribai Phule Pune University,
B.E. in Computer Science.

ANIKET GADHAVE,
Student at AISSMS COE, Pune-01.
Savitribai Phule Pune University,
B.E. in Computer Science.

SUMEDH DHENGRE,
Assistant Professor at AISSMS COE, Pune-01.
Savitribai Phule Pune University,
B.E. in Computer Science.