_____

Facial Emotion Recognition usign CNN

Meghana P G[1], Mohammad Yusuf Khan[2], Mahesh Bharti[3], Kalpesh Mohanta[4], Kanaiya V K[5]

[1,2,3,4,5]*School of Computing and Information Technology, REVA University, Bengaluru, India.*
Email: meghanapg3@gmail.com[1] , askyusuf.khan@gmail.com[2] , maheshbharti3234@gmail.com[3] ,
kalpeshmohanta@gmail.com[4] , kanaiyavk@reva.edu.in[5]

*Corresponding author's E-mail: meghanapg3@gmail.com

Article History	Abstract
	Nowadays where most of the works are being carried online the demand for face recognition technique is elevated. Computerized software is assisting in identifying human feelings such as happiness, sadness, anger, fear, disgust, etc. Over the decades, various research has been carried on the facial expression and emotion recognition. Emotion detection has extended applications. It is not merely related to any specific field nut instead the approach ranges from communication, advertising to hospital requisition and many more. To exist collective mechanisms through which we can accomplish the process of facial emotion recognition. In this paper we are using Convolutional Neural Network for the implementation. Upon exploring numerous datasets for the procedure of experiment we have chosen to go with Kaggle dataset.
	Keywords: *Emotion Detection, Features Extraction, Classifiers, Preprocessing, Segmentation, Convolution Neural Network (CNN)*

1. Introduction

Expression or emotion is the most important factor in any interpersonal communication. They can take many forms that the human eye cannot detect. In the context of interpersonal communication, facial recognition is currently a necessary condition. It helps to understand the intentions of others. Usually, people use expressions such as joy, anger, sadness, and voice to help understand emotional states. Verbal components make up one third of non-verbal components and human communication make up the remaining two thirds. Facial expression is the main channel of interpersonal communication between 2/3. With the help of advanced tools and technologies, they can be identified and recognized. Various functions (such as face, voice, EEG, and text) can be used for emotion detection. Among them, facial expressions are an effective function because they are visible and more information can be extracted from them. Multiple sectors are looking for facial emotion detection for there very own purpose where human emotion can be used to identify desired applications as well as a substitute for solving some problems. document is a template. For questions on paper guidelines, please contact us via e-mail.

Literature Survey

[1]. For emotion recognition, a deep learning framework is proposed. The framework is for classification and feature extraction with the help of Gabor filters and CNN respectively. The experimental results shows the improvement in both the CNN as well as the speed of the recognition accuracy training process. By extracting the image sub feature with the help of the Gabor filter, the neural network is obtained. CNN receives a variety of sub features as a result of this, and moves one step forward to extracting emotions from facial expressions. [2]. With the help of CNN, we using a technique called facial emotion recognition (FERC). It is basically based on a CNN with two parts. The first section removes the image's background, and therefore the second section focuses on the extraction of facial feature vectors. The FERC algorithm has the advantage of working with different orientations due to it's unique EV feature matrix. The ability to accurately determine emotions was greatly enhanced by the removal of the backgrounds. For several emotion-based applications such as polygraph, mood-based learning, etc, FERC could be the first step.

[3]. A 2 layer CNN model for facial emotion recognition is proposed in this paper. With the help of the image dataset, the model classify five different facial emotions. The model is generalised to the data and has the best fit, because of it's validation accuracy and comparable training accuracy. The loss function is scaled back using an Adam optimizer, and the model has been tested to have an accuracy of 78.04 percent. Using a video sequence, it is frequently extended to seek for changes in emotion, which is then used for various real-time applications such as feedback analysis, etc. To achieve effective control facial, this method can also be used in combination with other electronic devices. [4]. We used a nine layer CNN in this paper to train and classify seven different types of standard emotions. To aid in the interpretation and analysis of micro expressions, the percentages of emotions at various stages have been calculated. This experiment makes use of the Extended Cohn Kanade (CK+) and FERC-2013 databases. Before recognising emotion, the Viola Jones algorithm was used to detect faces. In Matsumoto and Hwang's study, traditional accuracy rates for people before training were 48 percent. The system's accuracy have gotten around 90 plus percentage with the help of a face data of a real time emotion recognition system has been developed and proposed using CNN.

[5]. In this paper deep learning has been used to arrive at aimed results. Basically there are 3 facial emotion classification stages such as emotion classification, face detection and feature extraction. Face detection in real time employs algorithms such as the Haar classifier, adaptive skin colour, and Adaboost contour points. Multiple techniques namely Linear Discriminant Analysis (LDA), Fisher Face method, Principle Component Analysis (PCA), Local Binary Pattern (LBP) were used to extract features. Neural Network (NN), Hidden Markov Model (HMM), Bayesian Network (BN), Support Vector Machine (SVM) and other algorithms were used to classify emotions or expressions. This paper has reviewed the experiment conduction in a controlled environment. [6Karolinska Directed Emotional Faces (KDEF) and The Kaggle datasets were the two main focus of this paper. The model Inception Net V3 Model is used for training purposes. The process of image labelling and automatic image classification can be performed with the help of Inception Net V3 Model. The accuracy rate in this paper was approximately 75.6 percent.

[7]. The author creates a fully connected deep neural network model for facial emotion detection in this article. To analyze the performance of the expected model, the model is explained in two public data sets. A unique DNN convolutional neural network upholds the improved model. It contains deep residual blocks and convolutional layers.In the extended model, a label is first prepared for training, with pictures from various aspects. Second, the extended DNN model of the image. In particular, the combination of residual block clouds and FCN was found to significantly improve overall results, confirming the advanced model's effectiveness. [8]. In this paper, to carry out an end-to-end spontaneous emotion prediction project from speech and image data, the author created a multimodal system that works at the raw signal. LSTM network was used to take into account the associated facts withinside the data. To trigger the learning of the model author trained the speech and optic networks, separately. Experimenting with unimodal modality indicates that the model get incredible results on the test set in compassion with different models. Further studies on the subject is the software of comparable architectures for conduct evaluation withinside the wild. Facial expression is the most simple and common signal for all humans to convey the mood. There are several tries to form an automatic face expression analysis tools [9] because it has application in several fields comparable to robotics, medicine, driving assist systems, and polygraph [10-11].

Objective
Due to its great academic and commercial potential, facial emotion recognition (FER) is a crucial topic in computer vision and artificial intelligence. Although multiple techniques are there to recognize emotion, it can be the video, audio, text, this article only focuses on research using only facial images. This research work aims to develop an application that uses video to classify emotions.
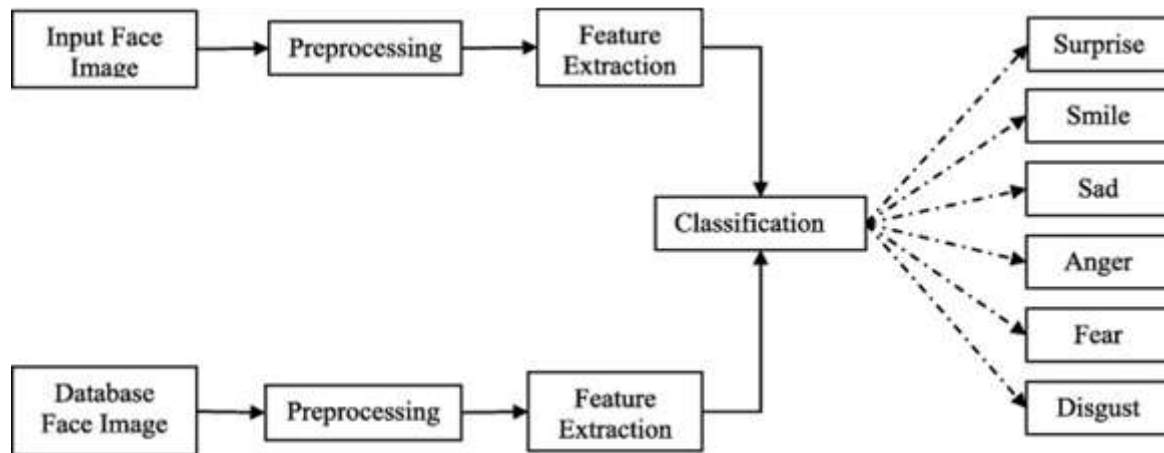
2. Materials And Methods
FER System makes use of the following steps for training and recognizing the emotions:

- Image retrieval,

- Image Pre-processing,

- Face Detection,

- Face Cropping,

- •   Image Segmentation,
- •   Masking
- •   Training and recognition,

The image dataset is bifurcated in training and testing data, and same steps are followed.

A. *Image Retrival*

The source of the image is from webcam or a camera, the livestreaming video is captured and converted in frames of images and then forwarded for Pre-Processing.

B. *Pre_processing*

This stage consists of a series of process carried out on each frame. The motive of pre-processing is to improve the picture quality. While recording, unnecessary blur and noise will appear in the video due to blur. This noise is eliminated by the filter. While pre-processing, Wiener Gaussian filter, median filter, and filter will be applied to reduce image noise, and the filtered output will be analyzed using the "peak signal" image quality indicator. The noise figure (PSNR) and root mean square error (RMSE) use these filters to pre-process the input image.

C. *Face Detection*

It is used for facial recognition. Create a system object known as a detector to discover the object. The cascaded object detector is used to locate the human face, mouth, eyes, nose, and upper body. By default, the detector is configured to locate faces. The feature displays the prolonged discipline used to outline the face area.

D. *Face Cropping*

After the face location is identified through the bounding box, the crop feature imcrop() is used to crop the face location from the enter portrait. The coordinates of the rectangle decide the cropping location. %Rect is a role vector with 4 elements [xmin, ymin, width, height] [10], which determines the dimensions and role of the crop location and crop location. Use Viola Jones algorithm to hit upon this reduce face in an iterative way to take away heritage effects (which includes T-shirts, etc.).

MP: Max Pooling   RS: Reshape   C: Calm   H: Happy   S: Sad   E: Energetic
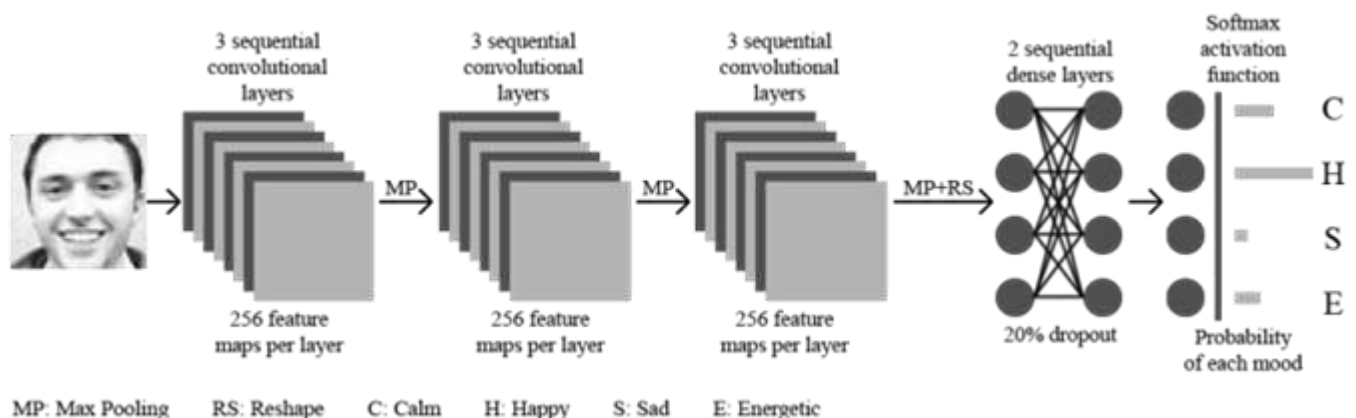
E. *Image Segmentation*

Skin colour is the most critical characteristic of a human face. Due to the want to understand feelings in non-pores and skin regions, the principle motive of this work is to split non-pores and skin regions inclusive of mouth and eyes from the image. The segmentation calls for a appropriate colour area. The RGB colour area is one of the most generally used rendering colour spaces. RGB image, pores and skin and no-pores and skin pixel threshold approach is simple to distinguish[11]. The thresholds carried out are indexed below, and the non-pores and skin elements are efficaciously identified.

$$Red > 95 \ \& \ Green > 40 \ \& \ Blue > 20,$$

$$max \ (Red, Green, Blue) - min \ (Red, Green, Blue) >= 15,$$

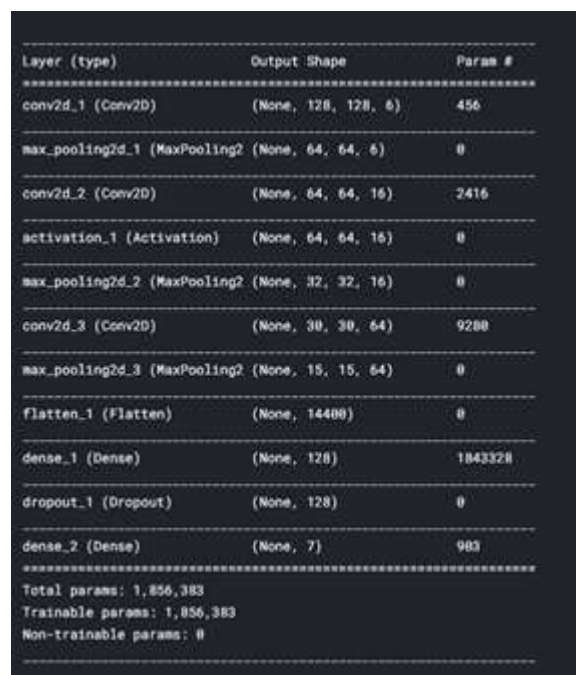$$|Red - Green| >= 15 \ \& \ Red > Green \ \& \ Red > Blue \ [9]$$

F. *Masking*

Masking is also called spatial filtering. The edges are masks. In this step, the extruded image is applied to the cropped face image to verify that the mask exactly matches the cropped image.

G. *Emotion Detection*

The essence of deep learning is to build a deep neural network similar to the structure of the human brain, and check the expression of more complex data functions layer by layer through several hidden non-linear structures. This machine learning mechanism of big data internal rules forces the extracted features to have more substantial data attributes, so the classification results can be greatly improved.

For 2D image input, the neural network model can interpret the entire layer from the pixels initially perceived by the computer to the edges, parts, object outlines, and objects understandable by the human brain, and then can be directly classified in the model to obtain the recognition results. CNN is a closed-loop neural network that can extract elements from 2D images and use backpropagation algorithms to optimize network parameters. Traditional CNN usually has three main levels: Convolution Layer, Pooling Layer followed by Connecting Layer. Each layer consists of several two-dimensional planes.

In CNN, the input layer is a two-dimensional arrangement of image pixels. The nesting of the convolutional layer C and the grouping layer S is the central module for performing feature extraction of the convolutional network. This paper develops a CNN framework for facial expression detection.

Layer (type)	Output Shape	Param #
conv2d_1 (Conv2D)	(None, 128, 128, 6)	456
max_pooling2d_1 (MaxPooling2	(None, 64, 64, 6)	0
conv2d_2 (Conv2D)	(None, 64, 64, 16)	2416
activation_1 (Activation)	(None, 64, 64, 16)	0
max_pooling2d_2 (MaxPooling2	(None, 32, 32, 16)	0
conv2d_3 (Conv2D)	(None, 30, 30, 64)	9280
max_pooling2d_3 (MaxPooling2	(None, 15, 15, 64)	0
flatten_1 (Flatten)	(None, 14400)	0
dense_1 (Dense)	(None, 128)	1843328
dropout_1 (Dropout)	(None, 128)	0
dense_2 (Dense)	(None, 7)	903

Total params: 1,856,383
Trainable params: 1,856,383
Non-trainable params: 0

Fig 3. CNN architecture

After training the model on the dataset,
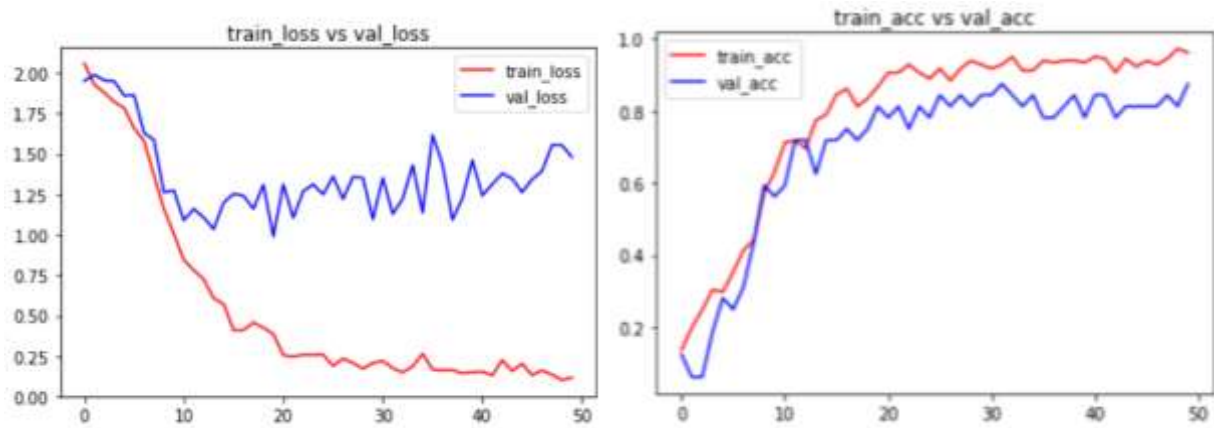
Fig 4: Final result after training the model

3. Results and Discussion

In this paper, we recommend a facial features recognition technique that uses a CNN model which extracts facial expression and features effectively. Compared to conventional methods, the proposed technique can robotically examine patterns, features and reduce the incompleteness due to artificial layout features. The proposed technique without delay inputs the picture pixel value via training pattern picture data. Facial expressions captured in fact might also additionally have various noises, together with face posture, occlusion, and blurring. To cope with this concern, as a destiny work, we are able to investigate extra strong fashions which fulfill actual conditions. We will additionally cognizance on the way to lessen the complexity of network structure, and could try and apprehend dynamic expressions with 3-D convolution technology.

4. Conclusion

Using convolutional neural networks, this paper proposes a brand new framework for recognition and facial expression detection. We believe that face is an vital a part of defining facial expressions. Facial expressions can enable neural networks with fewer than ten layers to compete (and possibly outperform) deeper networks in detecting emotions. Extensive experimental evaluation of our work on 4 famous facial reputation databases has proven encouraging results, In addition, we've got carried out a show technique to focus on the maximum seen regions of facial photos, which can be the maximum vital elements of photos that understand diverse facial expressions.

References:
[1] B. Majidi, M. M. Taghi Zadeh and M. Imani, "Fast Facial emotion recognition Using Convolutional Neural Networks and Gabor Filters," 2019 5th Conference on Knowledge Based Engineering and Innovation (KBEI), 2019, pp. 577-581, doi: 10.1109/KBEI.2019.8734943.
[2] Mehendale, N. Facial emotion recognition using convolutional neural networks (FERC). SN Appl. Sci. 2, 446 (2020). https://doi.org/10.1007/s42452-020-2234-1
[3] M. H. Supriya, E. Pranav, S. Kamal and C. Satheesh Chandran, "Facial Emotion Recognition Using Deep Convolutional Neural Network," 2020 6th International Conference on Advanced Computing and Communication Systems (ICACCS), 2020, pp. 317-320, doi: 10.1109/ICACCS48705.2020.9074302.
[4 G. Sanyal, G. A. R. Kumar and R. K. Kumar, "Facial emotion analysis using deep convolution neural network," 2017 International Conference on Signal Processing and Communication (ICSPC), 2017, pp. 369-374, doi: 10.1109/CSPC.2017.8305872.
[5] Ko BC. A Brief Review of Facial Emotion Recognition Based on Visual Information. Sensors (Basel). 2018;18(2):401. Published 2018 Jan 30. doi:10.3390/s18020401
[6] D Y Liliana "Emotion recognition from facial expression using deep convolutional neural network"2019 J. Phys.: Conf. Ser. 1193 012004
[7] Paramjit Sehdev, Deepak Kumar Jaina and Pourya Shamsolmoali, Extended deep neural network for facial emotion recognition, Pattern Recognition Letters, Volume 120, 2019,
Pages 69-74, ISSN 0167-8655, https://doi.org/10.1016/j.patrec.2019.01.008.
[8] Stefanos Zafeiriou, Panagiotis Tzirakis, George Trigeorgis, Mihalis A. Nicolaou, Björn Schuller " End-to-End Multimodal Emotion Recognition using Deep Neural Networks" Computer Vision and Pattern Recognition 10.1109/JSTSP.2017.2764438
[9] Zafar B, Ashraf R, Ali N, Iqbal M, Sajid M, Dar S, Ratyal N (2018) A novel discriminating and relative global spatial image representation with applications in CBIR. Appl Sci 8(11):2242
[10] Ali N, Zafar B, Riaz F, Dar SH, Ratyal NI, Bajwa KB, Iqbal MK, Sajid M (2018) A hybrid geometric spatial image representation for scene classifcation. PLoS ONE 13(9):e0203339
[11] Ali N, Bajwa KB, Sablatnig R, Chatzichristofs SA, Iqbal Z, Rashid M, Habib HA (2016) A novel image retrieval based on visual words integration of SIFT and SURF. PLoS ONE 11(6):e0157428