A Deep Learning based Approach for the Recognition of Facial Identity and Expression

Dheeraj Kumar
Department of Computer Science
Pondicherry University
Puducherry, India

Vaitheki K.
Department of Computer Science
Pondicherry University
Puducherry, India

Suresh Joseph K.
Department of Computer Science
Pondicherry University
Puducherry, India

Prasiddha Sarma
Department of Computer Science and Engineering
Tezpur University
Assam, India

ABSTRACT
This work aims to recognize human facial expressions and along with that it tries to incorporate the recognition of the person from the old data which were fed into the system earlier. The developed system uses the bounding box regression technique to perform the object localization task. Various instances from seven types of expressions can be detected using the developed system. From the real-time pictures/videos, the module can identify any of the instances of seven expressions and the person's facial identity simultaneously. The proposed system has been tested across various natural environments and it did perform well in most of the scenarios..

Keywords
Facial identity recognition, CNN, Facial expression recognition.

1. INTRODUCTION
Communication plays a massive part in our daily life and wherein facial expressions are one of the most important aspects of communication. This is because facial expressions not only help in expressing thoughts and ideas but also emotions in a substantial way. The personal emotion of the person standing nearby could be easily identified just by observing their facial expression. Although they are considered to be products of social learning as culture, Darwin mentions that there are expressions that are products of human evolution and technology complements our communication to a great extent. Various sensors are used to identify emotions in human faces and is widely employed in a variety of sectors. The category class of "Neutral" was also included to the sentiment analysis tool in the existing datasets, which automatically detects the six basic universal expressions of happiness, sorrow, anger, surprise, fear, and disgust and including Neutral. The proposed work deals with image datasets, performing data processing, and a convolutional neural network trained using Keras with the idea of recognizing the correct expression of a person. The Facial identity Recognition technology that adopts specific facial points also compliments the expression to identify the person whether is a known or unknown in the respective location.

2. BACKGROUND
In this era of Artificial intelligence, with the advancement of various technologies, human race is gaining new heights with time. It is widely known that, use of face recognition technology is becoming crucial day by day. From office check-in to medical consultations, in every dimension there comes a need of a face recognition technology in some point of time. And with these, researchers are coming up to solve various arising issues relating to these technologies. For further development/improvement of these technologies, numerous works are being carried out in various parts of our world. Relating to this literatures such as [1] can be discussed, where Lawrence developed a Convolutional Neural Network that architecture which outperformed other algorithms in recognizing human faces. GoogleNet-M was also implemented in the literature and it provided some excellent results in face recognition [2]. There will occur some miss prediction for twin person. Some exemplary research were done to solve these issues [3]. Support Vector Machines have shown pretty good results in predicting human faces [4]. Gender recognition from face data is also a challenging task. In [5], authors used face recognition feature vectors to recognise the gender.

The task of understanding human expressions is also gaining a large interest among researchers. With the development of a proper face expression recognition architecture, it yields many benefits to society; such as understanding safety, and security aspects in large area/gatherings, where various automated sensors (i.e., cameras) can help administrators in maintaining a safe and secure environment. The face expression recognition system must able to decode the person's various mental situations such as anger, happiness, sorrow etc., by analysing the picture of the facial expression of the person. The research gon these topics are highly growing among the researcher community [6]. Numerous developments have been made to overcome the key issues and other related issues arising to different factors.

Occlusion, unstable picture information, and illumination are some of the issues. In [7], author proposed a new algorithm "Local Directional Number" to recognise the face and expressions. Estimating and recognising face expressions from in-the-wild videos is a very complex task. In [8], researchers combined various networks such as deep residual network along with modified Recurrent Neural Network to solve this issue of the recognition of face expressions in-the-wild videos.

3. METHOD
The goal of the model is to compile various facial expression images from various datasets and alter them using image augmentation techniques like scaling, zooming, and rotation. dividing the data into training and test sets while also converting it from images to vectors. Then, using various deep learning techniques, select the most appropriate algorithm and train the model twice to account for the volume of data being used. Compare different accuracy and loss with various functions, and then use it to stream data in real-time from a webcam or an embedded camera in a laptop.

Additionally, to train the machine using images of people, it is required to store its facial points, and compare them to real-time facial for recognition. In this work, a image dataset collected from the internet, performing data processing techniques, creating and training a convolutional neural network using Keras TensorFlow 2.X. The idea is to show the correct expression and identify the person in the frame.

3.1 Collection of Dataset

For solving any task using Machine Learning/ Deep Learning, data plays the most vital role. In preparation of the model, two datasets were used. One of the datasets used here is collected from Kaggle[9]. The dataset "FER-2013" contains 32,298 pictures and having seven categories of expressions. All the images are grayscale images. The categories present in the dataset are angry, disgust, fear, happy, neutral, sad, and surprise.

Also, one new image dataset is created, which contains colored images of 10 persons. These 10 persons performed all the seven expressions thrice, and it is also used for the training purpose of the recognition model. All the images are converted into 85x85 pixel size and since few images are coloured images, those are also converted into grayscale images.

Data augmentation was done to increase the size of the dataset, as it helps in training a neural network.

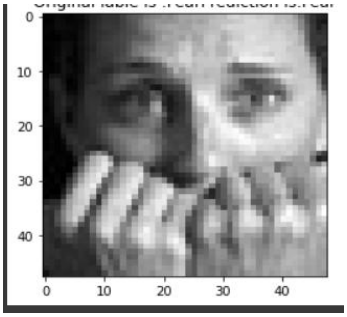


Fig 1(a): Snapshot from the dataset

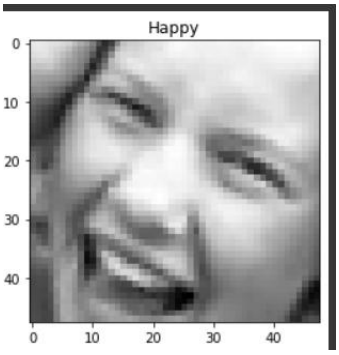


Fig 1(b): Snapshot from the dataset

3.2 Flowchart of the Model

The following figure depicts the flowchart for the proposed model.

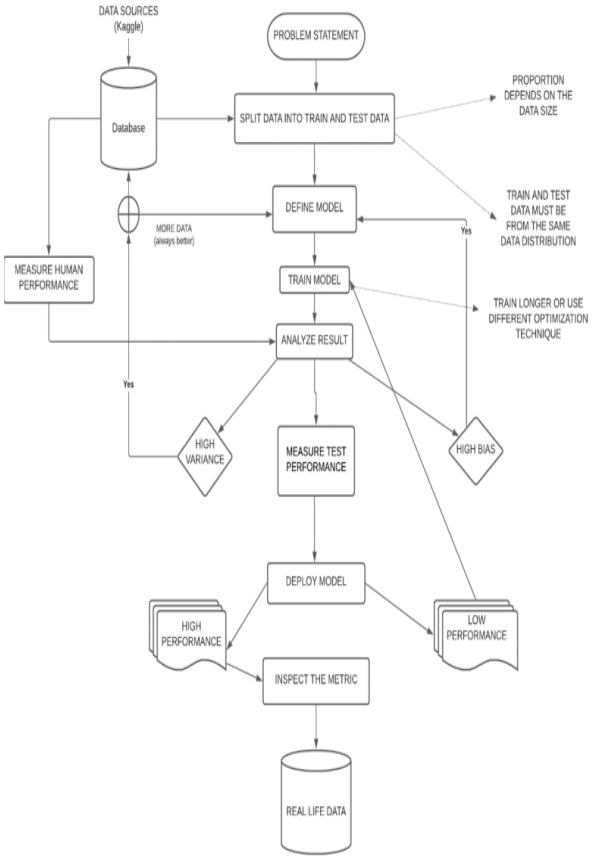


Fig 2: Flowchart of the proposed model

3.3 Face Detection

Initially, the detector will detect the face segments from the picture/video using bounding box regression after that the image features are collected from the grayscale images. The facial landmarks are also collected from the pre-processed grayscale images. '*Viola-Jones Face Detection Technique'*[10] is used to detect the face from the provided images/video frames. In this algorithm, it takes the help of positive and negative images. By positive image it refers to the images of faces and by negative images it refers to the images without having any faces.

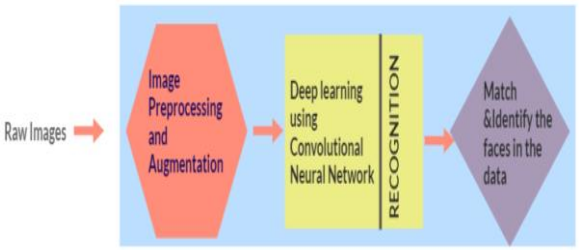


Fig 3: Architecture of face detection module

3.4 Emotion Detection

After detecting the faces successfully, the emotions of the person in the video frame will be predicted by the Convolutional Neural Network (CNN) module which was trained on the given dataset.

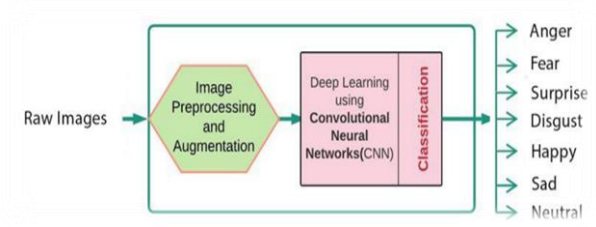


Fig 4: Architecture of emotion detection module

3.5 Combined Model for Face and Emotion Detection

The module for face detection as well as emotion detection were developed separately. Later these two modules are combined into a single module which will detect the emotion of the person in the frame and identify the name of the person if his facial information is available in the dataset and if not it will detect the emotion and name the person will be set as unknown.

4. CNN ARCHITECTURE

In our work, CNN is implemented to recognise the persons and to extract facial features. From these feature, the CNN is trained and in the later part the facial expression model is incorporated which is able to recognise seven emotions.

The CNN is a generalized deep learning architecture for understanding, predicting grid-like structures such as images.

The following figure is used to illustrate a CNN structure and it is collected from "https://www.analyticsvidhya.com/blog/2021/05/convolutional-neural-networks-cnn/"

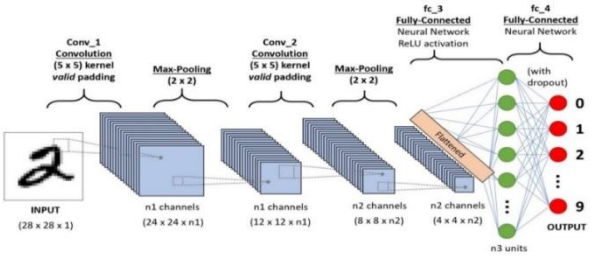


Fig 5: . A generalized architecture of a CNN

The CNN architecture incorporated in the system have the following layers – 4 Convolutional layers, 2 Max pooling layers, 4 dropout layers, 1 flatten layer, 2 dense layers.

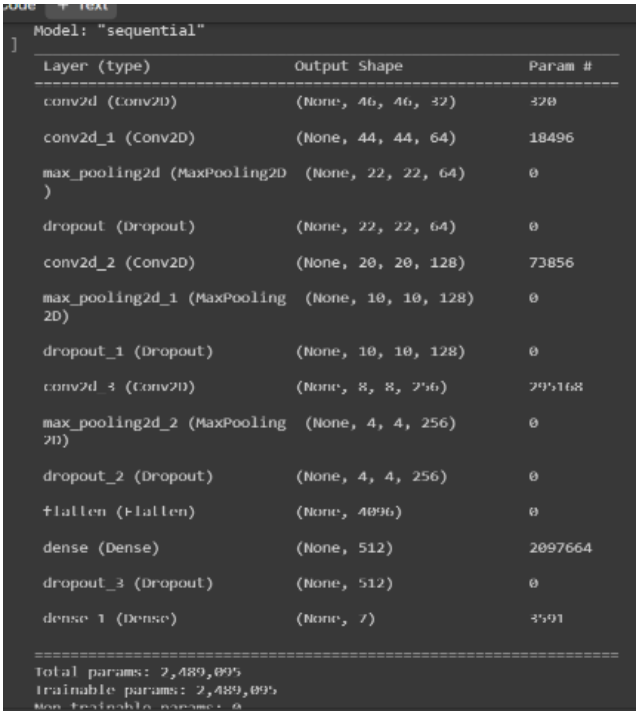
Convolutional layers uses the concept of filters to extract the features from the images. It computes the kernel filters and provides the feature information.

Max pooling layers deals with reducing the size or dimension of the feature set. It helps in reducing the complexity at the time of learning from the features.

Dropout layers are mainly used for reducing the overfitting issue of the training data.

Flatten layers are used to reduce the matrix of filters into one-dimensional array, and it reduces the computation cost of the model.

Dense layer is a totally connected layer from its previous layer. This layer mainly deals with the classification of the images.



Model: "sequential"

Layer (type)	Output Shape	Param #
conv2d (Conv2D)	(None, 46, 46, 32)	320
conv2d_1 (Conv2D)	(None, 44, 44, 64)	18496
max_pooling2d (MaxPooling2D)	(None, 22, 22, 64)	0
dropout (Dropout)	(None, 22, 22, 64)	0
conv2d_2 (Conv2D)	(None, 20, 20, 128)	73856
max_pooling2d_1 (MaxPooling2D)	(None, 10, 10, 128)	0
dropout_1 (Dropout)	(None, 10, 10, 128)	0
conv2d_3 (Conv2D)	(None, 8, 8, 256)	295168
max_pooling2d_2 (MaxPooling2D)	(None, 4, 4, 256)	0
dropout_2 (Dropout)	(None, 4, 4, 256)	0
flatten (Flatten)	(None, 4096)	0
dense (Dense)	(None, 512)	2097664
dropout_3 (Dropout)	(None, 512)	0
dense_1 (Dense)	(None, 7)	3591

Total params: 2,489,095
Trainable params: 2,489,095
Non-trainable params: 0

Fig 6: . Screenshot of the layer information

5. MODEL EVALUATION

The model is developed as a combined predictor for both facial recognition and expression recognition. It first perform localization of the objects to locate the face in the image and after extracting the face data it proceeds to the next step. If there exists a face in the image, it again compute its features and tries to match the facial information with the dataset. After finding a match it results the person name, otherwise if it unable to find person's match from the dataset, it results as a unknown person. After executing these phase of processes, the model will perform the operations to predict the facial expression. The resultant facial expression information will be one class of expressions among the seven classes of expressions.

The following figures shows the accuracy and loss of the model respectively.

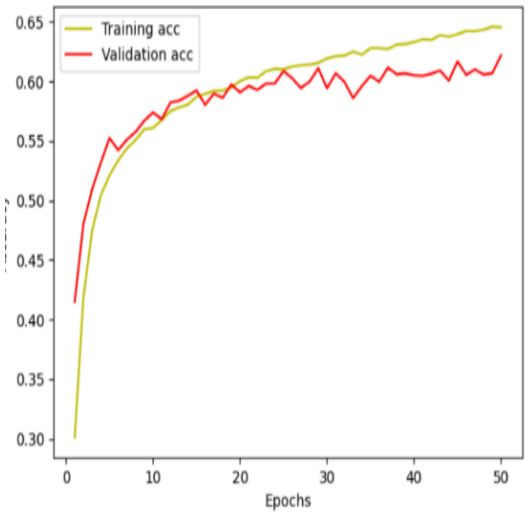


Figure 7: Accuracy of the model

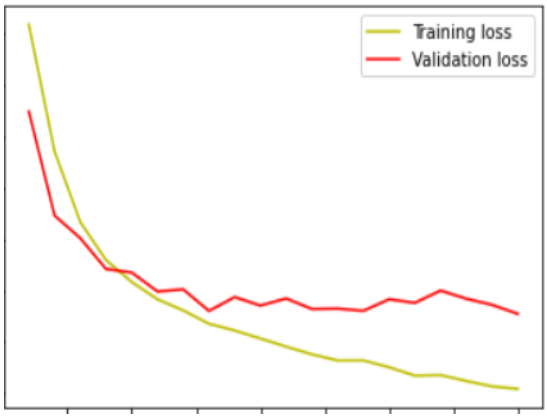


Figure 8: Model loss

6. SYSTEM SNAPSHOTS

Here few snapshots of the working model are shown..

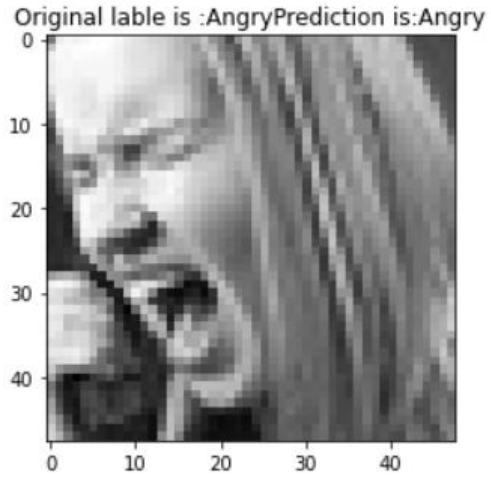


Figure 9(a). Snapshot of a predicted image (Emotion module)

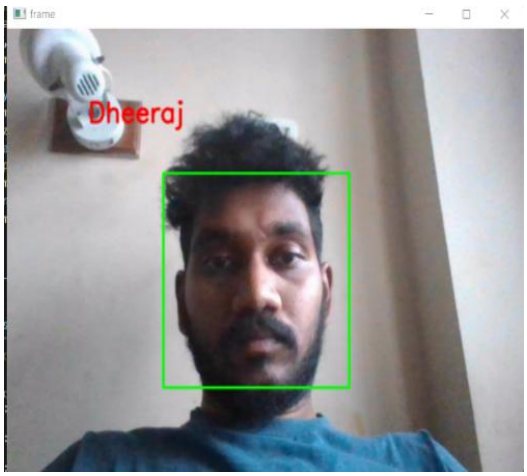


Figure 9(b). Snapshot of a predicted image(Face recognition module)

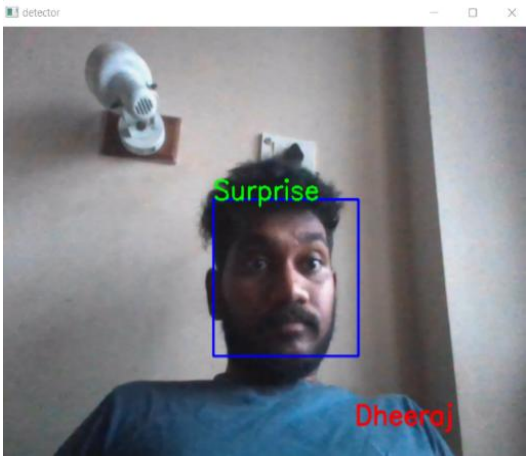


Figure 9(c). Snapshot of a predicted image(Combined module)

7. CONCLUSION AND FUTURE WORK

In the field of Computer vision, facial recognition plays an enormous role. While dealing with estimating the human postures, faces; it is very important to estimate it correctly. Also tremendous research works are going on in the area of emotion recognition. The facial expression recognition system can be further extended to development of a proper system for understanding non-verbal feedback from persons. By understanding the facial expression, machine can generate its meanings and it is very crucial to generate these meanings correctly. Further improvement can be made to the system by adding large amount of data – images/videos, that will help to execute Neural networks in a more sophisticated way.

8. REFERENCES

[1] S. Lawrence, C. L. Giles, Ah Chung Tsoi and A. D. Back, "Face recognition: a convolutional neural-network approach." in IEEE Transactions on Neural Networks, vol. 8, no. 1, pp. 98-113, Jan. 1997, doi: 10.1109/72.554195.I. S. Jacobs and C. P. Bean, "Fine particles, thin films and exchange anisotropy," in Magnetism, vol. III, G. T. Rado and H. Suhl, Eds. New York: Academic, 1963, pp. 271–350.

[2] Z. Yu, Y. Dong, J. Cheng, M. Sun, και F. Su, 'Research on Face Recognition Classification Based on Improved GoogleNet', *Security and Communication Networks*, 2022.

[3] R. Angeline, Krishna Chaitanya Poladi, and Rudra Dev Mishra , "Leverage and discriminant analysis of face recognition", AIP Conference Proceedings 2405, 020020 (2022) https://doi.org/10.1063/5.0072956.

[4] Guodong Guo, S. Z. Li and Kapluk Chan, "Face recognition by support vector machines," *Proceedings Fourth IEEE International Conference on Automatic Face and Gesture Recognition (Cat. No. PR00580)*, 2000, pp. 196-201, doi: 10.1109/AFGR.2000.840634.

[5] Y. Lin and H. Xie, "Face Gender Recognition based on Face Recognition Feature Vectors," 2020 IEEE 3rd International Conference on Information Systems and Computer Aided Education (ICISCAE), 2020, pp. 162-166, doi: 10.1109/ICISCAE51034.2020.9236905.

[6] B. He, "Deep learning algorithm for Facial Expression Classification" 2022 3rd International Conference on Computer Vision, Image and Deep Learning &

International Conference on Computer Engineering and Applications (CVIDL & ICCEA), 2022, pp. 386-391, doi: 10.1109/CVIDLICCEA56201.2022.9824661.

[7] A. Ramirez Rivera, J. Rojas Castillo and O. Oksam Chae, "Local Directional Number Pattern for Face Analysis: Face and Expression Recognition," in IEEE Transactions on Image Processing, vol. 22, no. 5, pp. 1740-1752, May 2013, doi: 10.1109/TIP.2012.2235848.

[8] H. Liu, J. Zeng and S. Shan, "Facial Expression Recognition for In-the-wild Videos," 2020 15th IEEE International Conference on Automatic Face and Gesture Recognition (FG 2020), 2020, pp. 615-618, doi: 10.1109/FG47880.2020.00102.

[9] M. Sambare, "FER-2013" https://www.kaggle.com/datasets/msambare/fer2013 (accessed Feb. 5, 2022).

[10] J. M. Al-Tuwaijari and S. A. Shaker, "Face Detection System Based Viola-Jones Algorithm," *2020 6th International Engineering Conference "Sustainable Technology and Development" (IEC)*, 2020, pp. 211-215, doi: 10.1109/IEC49899.2020.9122927.