Journal of Applied Computer Science & Artificial Intelligence

Research Paper Vol.-2, Issue- 1, June 2021 ISSN: 2582-8673

Human Facial Expression Detection Using Convolutional Neural Network through Deep Learning

Samarpita Saha[1*], Shambhavi Raj[2], Animish Nishakar[3], Abhishesk Tiwari[4], Amit Majumder[5]

[1] CSE department, JIS College of Engineering, email: samarpitasaha263@gmail.com
2 CSE department, JIS College of Engineering, email: shambhaviraj19@gmail.com
3 CSE department, JIS College of Engineering, email: animish15@gmail.com
4 CSE department, JIS College of Engineering, email: hr0624117@gmail.com
5 CSE department, JIS College of Engineering, email: amit.majumder@jiscollege.ac.in

Available online at: http://jacsai.org/

***Abstract*—** A facial emotion is the noticeable appearance of the emotional situation, psychological feature exertion, purpose, temperament and psychopathology of a discrete and plays a forthcoming role in social connection Automatic recognition of face expressions will be a crucial part of natural human-machine interfaces. We used a deep learning approach using Convolution Neural Network (CNN) for facial expression recognition. CNN model of the work relies on LeNet design. Kaggle face expression dataset with 7 face expression tags as happy, sad, fear,surprise , anger, neutral and disgust is employed in this work.

***Keywords*—** Facial Expression Recognition,Deep Learning , Convolutional Neural Network,

I. INTRODUCTION

Facial emotion recognition is the procedure of recognizing human emotions from facial expressions [1,2,3]. The human brain detects emotions automatically and software has now been progressed that can detect emotions as well. This technology is becoming more precise all the time, and will eventually manage to read emotions the same as our brains do. AI can detect emotions by studying what each facial expression means and put in an application that knowledge to the new information presented to it. Emotional AI or emotion artificial intelligence is a technology that has the ability to read, imitate, interpret, and reciprocate to human facial expressions and emotions [4,5,6].

II. RELATED WORK

Multilayer perceptron is designed to use least amounts of preprocessing. A countenance is that the observable proof of the emotional state, sensory action, purpose, temperament and psychopathology of an individual plays an upcoming part in mutual affairs. Human facial expressions may be categorized into seven primary reactions like happy, sad, surprise, fear, anger, disgust and neutral. Our facial expressionns area unit shows through stimulation of specific parts of face muscles. These generally refined, nonetheless complicated,

signals in associate auditory communication typically contain a plentiful quantity of statistics relating to our state of mind. Reflex acknowledgment of facial expressions may be associated with an important amount of natural human-device interaction; the situation should even be working in psychological knowledge and in experimental application. It's been studied for an outing of mind and brought progress in current decades. Although a lot of progression has been created, recognizing countenance with a high correctness remains to be tough thanks to the problem and variation of facial appearances .On each period to period fundamentals individuals unremarkably get emotions by specific properties, presented such as a vicinity of a feature. Design for manifestations contentment is definitely connected through a beam or associate increasing gesture of the angles of the mouths. Likewise different reactions are distinguished through different distortions characteristic to a selected appearance analysis into machine-controlled identification of facial expressions. It directs the issues related to grouping of static and dynamic options of distortions of facial

*Corresponding Author

appearance. A CNN in machine learning, might be a genre of feed-forward artificial neural network inner which the characteristic design concerning its neurons is fascinated by the society of the animal neuronal area. Specific animal tissue neurons respond to boost in a very constrained range of areas mentioned as the pliable field. The approachable fields of several neurons part overlay and the field of regard tiled by them. The response of a private somatic cell to stimuli at intervals in its receptive field may be approximated mathematically by a convolution operation. Convolutional networks were impressed by biological processes and area unit variations of multilayer perceptrons designed to use least amounts of preprocessing.

III. PROPOSED APPROACH

In this work of Facial Expression Recognition we use the Deep Learning approach with the Convolutional Neural Networks model .The format with the huge amount of dataset is in CSV (Comma Separated Values). To develop Facial Expression Recognition there are two phases –training and testing. The Convolutional Neural Networks model is applied to train and test dataset . One model is built using a training dataset and then predict the relationship using a test dataset which is used to validate the model. In training, the FER system collects a training data 48*48 greyscale pictures of faces, with the expression tag and learns a connection weights for the network. We are going to read an image with a face from a training and testing dataset. Normalization is applied to change the range of pixel intensity in the image. Input an image with a face in the training step. Normalization is an important step to train the Convolutional Network [7,8], which ensures that each input pixel in the training dataset has a similar data distribution. Validation dataset is used to compare the performance and decide to select the set of weights with different numbers of hidden layers during training. In testing , the FER system collects a greyscale image of a face from a test set using the final performance obtained during training it and predicts the output by considering one of the seven basic expressions.

Convolution neural networks(CNN) [3,4,5] cope with the image process. Machine acquires the knowledge from the info given and supports that device acknowledges the pictures. For example, we offer an exact set of various pictures of cats and dogs to a CNN model, that model properly acknowledges the pictures of dog and cat.

Let's see however a machine processes the pictures before making a Convolutional neural network model. Suppose that we've given an emoticon face to the laptop then, the device initially splits the given image into element size so a matrix is produced to acknowledge the colours of the image on the premise of RGB combination.

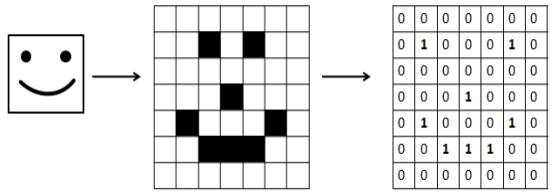

Figure 1. Image represented as multidimensional array

In Figure-1 the machine has represented emotion faces by a multidimensional array, in this array 0represents the white colour and 1represents the black colour. Convolution operations apply a filter to create a feature map to cut back the scale of the picture so the more processes on pictures are going to be straightforward and quicker. Each image has its own characteristics so convolution operation helps to detect traits of pictures exploitation trait detector.

To generate a convolutional neural network(CNN) [11, 12] model we have performed four operations. These are:

- a)Convolution operation
- b)Max pooling
- c)Flattening
- d)Full connection

A. Convolution operation

By Convolution operation a feature map is created to scale back the scale of the picture so the further process on pictures are going to be simple and faster. Each picture has its individual options thus to find the actual characteristic of a picture this operation finds a feature of picture victimisation feature detector.

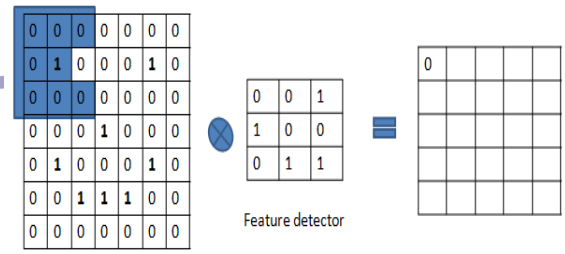

Figure 2. Feature Detector

From the above figure we can observe that a feature map is created by convolution operation merely increasing the image matrix to the feature detector. It retains choosing a 3*3 array from the pictures and detects any common range amid a specially picked matrix and detector. Mainly the feature detector matrix are often of any size here only for calculation simplicity we tend to square measure victimisation 3*3 feature detectors. The method is followed till each 3*3 combinations of input images are completed. Concluding feature map is shown in Figure 3.

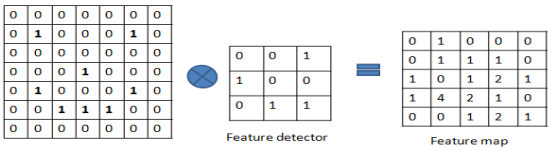

Figure 3. Feature Map

A. *MaxPooling*

We noticed that each picture has its own characteristics . But, a picture can have some special characteristics than other pictures given to the machine. Detecting distinct characteristics in a picture is called Max pooling. A model is generated by us that has ductility to recognize such unique characteristics in pictures.We have picked a 2*2 array from the above map and a maximum number from the map has been chosen and located in the pooled feature map as shown in below.

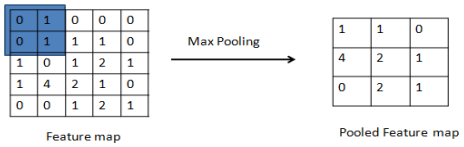

Figure 4. MaxPooling

From this above picture we can notice that from the picked 2*2 array machines have chosen the max number and located it in the pooled feature map on the other part of the figure. A last pooled feature map created after the machine has done the process.

B. *Flattening*

We used flattening to convert the data into a one-dimensional array .It is the most straightforward function in convolutional neural networks. During this step the pooled feature map is flatten in a single column. We have done this because the 2D array goes to be AN input layer to the long run artificial neural network of the CNN model. The pooled feature map looks like beneath after flattening .

Figure 5. Flattening

Convolutional Neural Network uses a fully connected artificial neural network to predict the pictures.

C. *Full Connection*

Full connection layer connects each neuron in one to another layer. To complete the classification task the fully connected layer is used where the number of nodes in the layer is set by the number of varieties used in the task. This layer is shown in Figure 6.

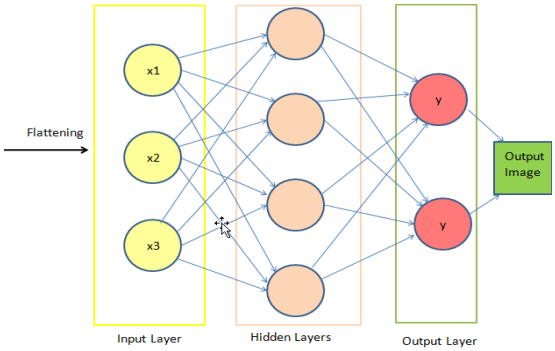

Figure 6. Fully Connected Network

IV. IMPLEMENTATION STEPS

In our approach for identifying facial expression we have used Deep Learning technique based on Convolution Neural Network (CNN). Two phases are there to develop the facial expression identification system. These are the training phase and testing phase.

D. *Training Phase*

1. During the training phase, the training dataset is provided to the system for learning purposes by inputting one image at a time. Each image is considered as a matrix of feature values. The following steps are performed during the training phase.

i. Sequential model is created by considering the required layers (e.g. Convolution, MaxPooling, Dense etc.) and initialize the neural network

ii. The images from the training image dataset are passed to the input layer of the deep learning neural network.

iii. The input data is then passed through the other intermediate layers of the network.

iv. The softmax activation function is used by the final layer after we pass between the fully connected layers. This activation function is used to acquire probabilities of the input taking place in a particular class which helps to do classification of the images.

E. Testing Phase

During the testing phase we provide a test dataset of images as input to the model generated in the training phase. Performance of the model generated after training (also called classifier) can be measured by finding recall, precision and f-score values. More f-score indicated better performance of the system. Recall, precision and f-score values for any category of images (say type *i* in general) are measured using the following formulas as mentioned in equations 1,2 and 3.

$$recall = \frac{\text{Number of images (type i) correctly identified by system}}{\text{Number of images(type i) in the standard test dataset}} \quad (1)$$

$$precision = \frac{\text{Number of images correctly identified (type i) by system}}{\text{Number of images identified (type i) by the system}} \quad (2)$$

$$F1\_score = \frac{2 \times recall \times precision}{recall + precision} \quad (3)$$

V. DATASET AND EXPERIMENTAL RESULTS

The dataset Facial Expression Recognition Challenge (FER2013) from Kaggle is employed for the coaching and testing. It holds pre-cropped, 48×48-pixel grayscale pictures of faces each tagged with seven feeling categories: happiness ,neutral, fear,anger, sadness, surprise, and disgust in all of them. The training set has 28,709 sample images. The public test has 3,589 examples. The final test set for evaluation purposes has another 3,589 examples.

VI. RESULTS AND ANALYSIS

CNN design for human facial expression detection as mentioned higher than was enforced in Python. Together with the Python programming language, pandas, numpy, sequential, flatten ,dense, dropout, convolution2D, maxpooling2D, zeropadding2D, SGD, cv2 libraries were used. We begin with the initialisation of the model followed by a batch normalisation layer with ReLu as associate degree activation and dropouts to try learning with efficiency. Experimental results for validation dataset and test dataset are shown below Table-1 & 2 respectively.

Table 1. Validation Result (0=Angry, 1=Disgust, 2=Fear, 3=Happy, 4=Sad, 5=Surprise, 6=Neutral)

Expression Type	precision	recall	f1-score
0	0.6	0.51	0.55
1	0.67	0.71	0.69
2	0.55	0.48	0.51
3	0.8	0.86	0.83
4	0.48	0.59	0.53
5	0.83	0.8	0.81
6	0.6	0.55	0.57

Table 2. Test Result (0=Angry, 1=Disgust, 2=Fear, 3=Happy, 4=Sad, 5=Surprise, 6=Neutral)

Expression Type	precision	recall	f1-score
0	0.6	0.46	0.52
1	0.65	0.69	0.67
2	0.56	0.52	0.54
3	0.77	0.87	0.82
4	0.48	0.59	0.53
5	0.8	0.7	0.75
6	0.6	0.55	0.57

From the result of experiment it is clear that our system provides best result on 'Happy' facial expression (83% for validation dataset and 82% f-score for test data).

VII. CONCLUSION

In our work, we used the Convolution Neural Network to categorize human facial emotions i.e. happy, sad, surprise, anger, fear, disgust & neutral. Our system provides accuracy of 66.7% on validation dataset after 120 epochs during training. In the future ,we thought to apply another deep learning approach to check whether performance of the facial expression recognition system improves or not. During training we have considered many parameters with their default values as specified in the implementation of the methods. Later we will try to modulate the parameters used in the system for improvement.

REFERENCES

[1] Shan, C., Gong, S., & cOwan, P. W. (2005, September). Robust facial expression recognition using local binary patterns. In Image Processing, 2005. ICIP 2005. IEEE International Conference on (Vol. 2, pp. II-370). IEEE.

[2] Chibelushi, C. C., Bourel, F. (2003). Facial expression recognition: A brief tutorial overview. CVonline: On-Line Compendium of Computer Vision, 9.

[3] C. Zor, “Facial expression recognition,” Master’s thesis, University of Surrey, Guildford, 2008.

[4] Suwa, M.; Sugie N. and Fujimora K. A Preliminary Note on Pattern Recognition of Human Emotional Expression, Proc. International Joint Conf, Pattern Recognition, pages 408-410, 1978

[5] Raghuvanshi, Arushi, and Vivek Choksi. "Facial Expression Recognition with Convolutional Neural Networks." Stanford University, 2016

[6] Alizadeh, Shima, and Azar Fazel. "Convolutional Neural Networks for Facial Expression Recognition." Stanford University, 2016

[7] Krizhevsky, Alex, Ilya Sutskever, and Geoffrey E. Hinton. ”Imagenet classification with deep convolutional neural networks.” Advances in neural information processing systems. 2012.

[8] Simonyan, Karen, and Andrew Zisserman. ”Very deep convolutional networks for large-scale image recognition.” arXiv preprint arXiv:1409.1556 (2014).

[9] Z. Yue, F. Yanyan, Z. Shangyou and P. Bing, "Facial Expression Recognition Based on Convolutional Neural Network," 2019 IEEE 10th International Conference on Software Engineering and Service Science (ICSESS), 2019, pp. 410-413, doi: 10.1109/ICSESS47205.2019.9040730.

Authors Profile

Samarpita Saha is currently doing her B.Tech from Computer Science and Engineering department, JIS College of Engineering, West Bengal, India. She has interest in Machine Learning, Deep Learning, and Image Processing.

Shambhavi Raj is currently doing her B.Tech from Computer science and Engineering department, JIS College of Engineering, West Bengal, India. She has interest in Machine Learning, Deep Learning, and Image Processing.

Animish Nishakar is currently doing his B.Tech from Computer science and Engineering department, JIS College of Engineering, West Bengal, India. He has interest in Machine Learning, Deep Learning, and Image Processing.

Abhishesk Tiwari is currently doing his B.Tech from Computer Science and Engineering department, JIS College of Engineering, West Bengal, India. He has interest in Machine Learning, Deep Learning, and Image Processing.

Amit Majumder has received B.Tech degree in Computer Science and Engineering from Kalyani Govt. Engineering College, Kalyani, West Bengal, India, and ME degree in Computer Science and Engineering from Jadavpur University, Kolkata, West Bengal, India. Currently he is working as Assistant Professor in CSE department at JIS College of Engineering, Kalyani, West Bengal, India. He has interest in areas of Artificial Intelligence, Soft Computing, Natural Language Processing, Machine Learning, Information Extraction. Currently he is doing his work on extracting bio-molecular event using Bio-NLP technique