

FACIAL EXPRESSION RECOGNITION USING CONVOLUTIONAL NEURAL NETWORKS

¹Ashlesha Vaidya, ²Hitee Sachdeva, ³R. Brindha^{1,2,3}Computer Science and Engineering

SRM University

¹ashlesha1807@gmail.com, ²hiteesachdeva96@gmail.com,³brindha.ra@ktr.srmuniv.ac.in

Abstract: In this paper we describe the application of facial emotion recognition. This paper delineates various stages involved in the facial emotion or expression recognition. The stages can be modularized into three stages viz. pre-processing module, training module and the testing module. These three stages have been studied in various papers and a comparative analysis has been performed. Various factors leading to degradation and improvement in the performance of the model have been considered. A successful model leads to automated emotion recognition with high accuracy and confidence. Such a model can further can applied to various other applications - an example is suggestive applications - where considering the emotions or mood of a person will tend to give better applicability.

Keywords: Emotion Recognition, face extraction, Histogram Equalization, Convolutional Neural Networks, Softmax Classifier

1. Introduction

Recognition of emotions by computers will give the edge that is much needed in today's research areas of artificial intelligence. Be it machine-human communication or suggestive applications the efficiently and accuracy is only limited by the one aspect that machines cannot identify with - the emotions. These emotions come by in an autonomous way for humans but it becomes a hurdle for a machine. Machines cannot perceive emotions as a human can. Scientists are focusing on changing that and building a model which can in full functionality behave like a human. This research has been a breakthrough in the field of artificial intelligence and machine learning as there are so many applications that can be bettered by implementing this model along with the existing ones.

Facial emotion recognition began in 1991 when A.Pentland and K.Mase used optimal flow method to determine the movement of muscles and then used the feature vectors to automatically determine facial

expressions [1]. As deep learning began gaining momentum and showed performance better than existing methodologies, scientists began experimenting with deep learning in various fields and thus one of the outcome was facial expression recognition obtainable at a higher accuracy.

This paper presents an approach using the artificial neural networks(ANN) for the application of facial emotion detection. A deep network called the Convolutional Neural Network gives a better accuracy for the task. They are basically better feature extractors as compared to other models and are also capable of giving real time results [2]. The aim is to develop a model which detects emotions as categorized into six basic classifications [4] and with a high accuracy. The six defined classifications of emotions are shown in fig.1.

The whole process of expression detection has been divided into three phases and thus elaborated upon in this paper. The three phases include pre-processing phase, training phase (through the deep neural network) and the testing phase. A dataset of images is chosen according to the required application so that the model is able to reach the desired accuracy.

The first phase - the pre-processing module takes the entire dataset as input and gets it ready for the next processing module. The next module is the training or the processing module. This consists of a multilayer neural network. Multiple layers with appropriate training data gives a high accuracy rate [2]. After the Convolutional Neural Network (CNN) is trained it has to be tested for a different or the same data to check the efficiency of the data- this comprises of the third testing module.

2. Dataset for Training

A database has to be designed or acquired such that it is sufficiently apt to train the CNN model. The database comprises of set of pictures of faces with various expressions. An extensive database found to suitable for the purpose of emotion recognition is the Extended Cohn-

Kanade dataset(CK+). This dataset corrects provides samples and evaluations for various expression and also takes into account various environmental factors like the variance in the brightness, thus giving a high accuracy model. In [1] another

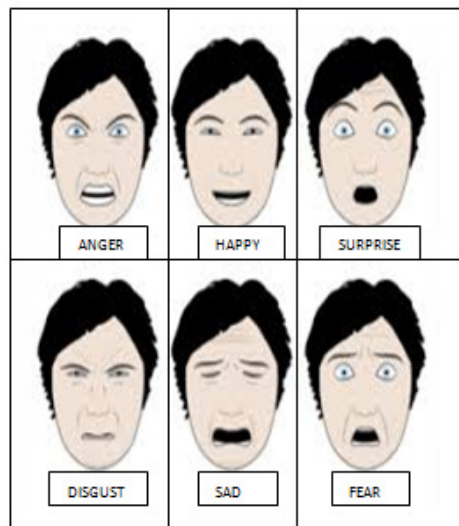


Figure 1. Six Basic Emotions

dataset called the Japanese female Facial Expression Dataset (JAFPE) was also worked on along with the CK+ dataset and it was found that CK+ gave better performance for the CNN model as compared to the JAFPE dataset. A sample set of the data points present in the CK+ dataset are shown in fig. 3 [5][6]. The dataset consists of various facial images depicting different expression in abundant lighting conditions. This variety helps the model to learn better and thus perform better in the testing phase.

3. Architecture of the Model

The over model designed for facial emotion recognition is divided into three modules. The control flow between these modules and the overall architecture is represented in fig. 2.

The CK+ dataset is made available to the pre-processing module. The face of a person is detected in the stage. Since the emotion of a person is particularly concluded with the face expression, the face needs to be distinguished from the other features available in a picture. Once the face is detected this module is also responsible for histogram equalization of the image in order to deal with various environmental conditions.

Once the facial image is pre-processed the dataset is sent to the processing unit which mainly comprises of the CNN. This module is where the learning takes place and it performs the classification of emotions. A CNN model gives better performance with smaller dataset. Flexibility of adding more layers to the model enhances our capacity to increase the performance of the model.

Finally the model is ready to be tested and the CNN can be tested with newer images to find its accuracy in recognizing images. This can be accomplished with the testing module. This paper further describes all three modules in greater detail.

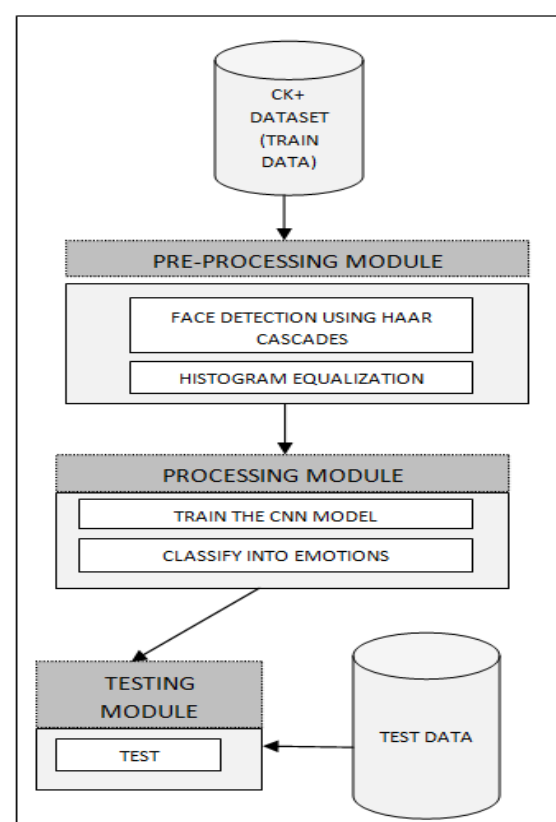


Figure 2. Architecture Diagram for the emotion recognition model

4. Pre-Processing Module

Before a data point can be entered into the model for either learning or classification it needs to be made suitable for processing.

The first step involved is the detection of faces from images. Extraction of faces from images can be performed with various algorithms. Haar Cascades found

by the Viola and Jones [8] are used for face detection or extraction. This algorithm works by evaluating haar-like features instead of pixel values and AdaBoost is used as the boosting algorithm [8]. Other algorithms such as simple vector machines can also be used for face detection but their performance is less accurate as delineated in [8]. Various other algorithms which can be used for face recognition and verification include Principle Component Analysis (PCA), Individual Component Analysis (ICA), Linear Discriminate Analysis (LDA), Local Binary Pattern (LBP) etc. [8]. Haar-like features are extracted from an image Figure 4 shows image of a person and the face extracted from that image using haar-like features.

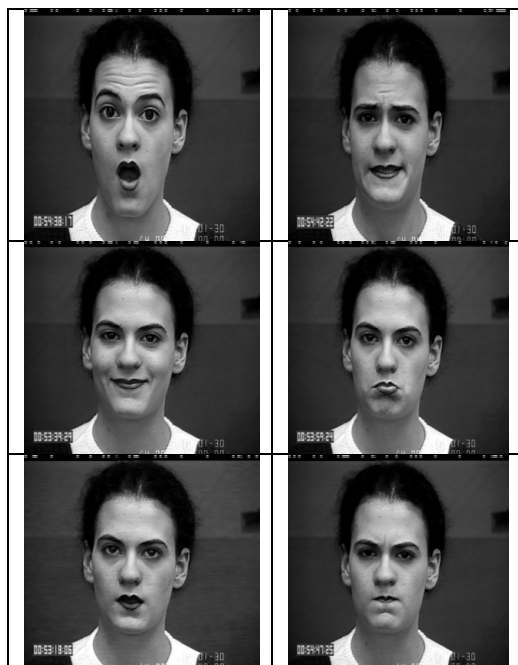


Figure 3. Dataset samples

The cropped and extracted face shown in fig. 4 is obtained by the opencv library in python.

The next step in this module is the histogram equalization of the cropped images of faces obtained in the previous step of face extraction.

This is necessary in order to take into account the various environmental constraints that may cause variance in the facial images. If these conditions are not taken into account then it becomes difficult for the CNN model to learn appropriately due to varying conditions or extremely varying pixel values. A comparison among the histograms of an image before and after the histogram equalization is shown in fig. 5.



Figure 4. Face extraction using haar-cascades.

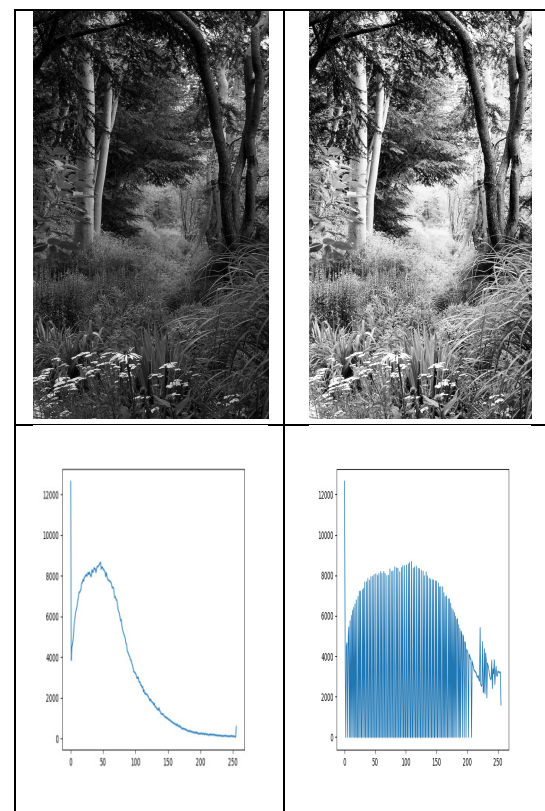


Figure 5. Comparison of histograms before and after histogram equalization

Figure 5 shows that histogram before the equalization is concentrated at one end and is not distributed. This means that the number of pixels for a particular pixel value is more. The particular picture shown in fig. 5 is chosen because it clearly shows the difference in the two

histograms. In practical terms this concludes that the image is either bright or dull. In order to compensate for this variation, we perform the histogram equalization. After this it can be seen that the histogram formed is more distributed than before and this concludes that the number of pixels for all pixel values are more or less equally distributed.

5. Processing Module

This module involves the CNN and the pre-processed images are given to the CNN for training it. Next section of this paper describes CNN and its architecture.

A. Convolutional Neural Networks

One layer of the CNN is divided into two namely convolutional layer and subsampling layer[1].

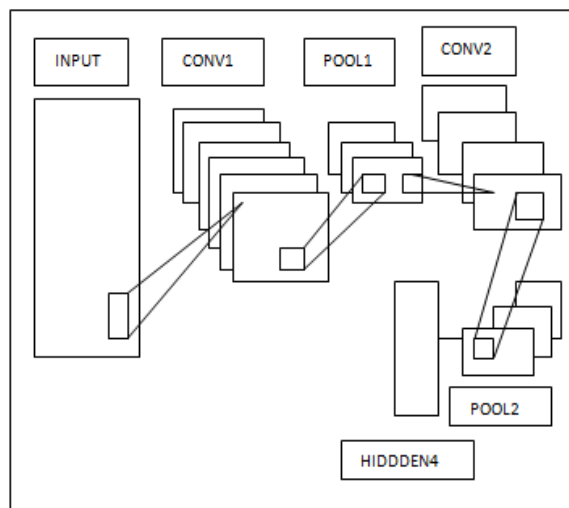


Figure 6. CNN Architecture

The input to the CNNs is the 2D images of faces, these images are then convoluted to extracted various features from the image. This forms the convolutional layer. This layer thus basically consists of features broken down from an image. Further extraction of features into their components results in the formation of the subsampling layer.

More than one convolutional and subsampling layers can be added to a network depending upon the application and the need for accuracy. Once the features are extracted from the image they are passed to some basic neural network classifier (in this paper we have discussed about softmax classifier). This classifier helps in the classification into emotions using the extracted features

from the images. Thus we can detect the expression from the image of a face of a person. A basic architecture model of the CNN structure is shown in fig. 6

In fig. 6 the input denotes the 2D pre-processed images. There are 2 layers of CNN in fig. 6- conv1 and pool1 form one pair of convolutional and subsampling layer and conv2 and pool2 form the second layer. As the image passes from conv1 to pool1 to conv2 and then to pool2 the level of features extracted becomes more and more intricate which in turn gives better accuracy. After pool2 there is a layers named as hidden4 which is a fully connected neural network used for classification purpose namely the softmax classifier. This classifier layer can classify from a number of options- it is multinomial. For a particular node in the network the output given b the softmax classifier can be represented by the equation 1.

$$y_i = \frac{e^{x_i}}{\sum_i e^{x_i}} \quad (1)$$

In the above equation 1 'i' represents a particular node, x_i terms are the inputs to the softmax classifier and the y_i term represents the output obtained from the classifier.

Using CNN is advantageous as these networks can learn better from a smaller dataset as compared to other available neural networks. Also Flexibility of adding more layers to the model enhances our capacity to increase the performance of the model.

Nonetheless there are some disadvantages to the discussed model. One of them being the distinction among the sad, disgust and angry expressions. The reason behind it is similarity among the features of faces for these expressions. The pixel values tend to resemble and so it becomes difficult for the model to learn from the data.

6. Testing Module

This module is used to test the performance of the model built. This is done by giving in test cases (images of faces) and checking the accuracy of detection. The test dataset can be different or similar to the train dataset. A different dataset will give more validity to the model. An accuracy of around 80% is obtained with the CK+ dataset and using two layers in the CNN[1].

7. Conclusion

A model to automatically detect emotions from facial expressions can be build using neural networks giving an accuracy of about 80%. This paper discussed about the various modules involved in the process of emotion recognition namely the pre-processing, processing and the

testing module. Acquisition of appropriate dataset is a prior necessity in order to proceed to operation of the various modules. CK+ dataset has been used in this paper. This contains a large number of subjects with various images under a variety of conditions. This makes the dataset an apt choice for dealing with certain problems with inaccuracy in predictions due to lightning conditions. After the three modules perform their operations the overall result is that the machine is able to judge the emotion of the person. This can be advantageous for various applications as it can help a machine to work with a human in a better and a more efficient way.

Further, this model can be intertwined with other practical models like suggestive applications- where a machine makes suggestions to a human- this can be improved along with this model as better suggestions can be made with better understanding of the likes and dislikes of the subject.

References

- [1] Ke Shan, Junqi Guo, Wenwan You, Di Lu, Rongfang Bie, "Automatic Facial Expression Recognition Based on a Deep Convolutional-Neural-Network Structure" , SERA 2017, London, UK, IEEE, June 2017, pp: 123-128
- [2] Peter Burkert_z, Felix Trier_z, Muhammad Zeshan Afzal, Andreas Dengel and Marcus Liwicki, "DeXpression: Deep Convolutional Neural Network for Expression Recognition", German Research Centre for Artificial Intelligence, Germany, arXiv:1509.05371v2, August 2016
- [3] Tanner Gilligan, Baris Akis, "Emotion AI, Real-Time Emotion Detection using CNN", Stanford University
- [4] Monika Dubey, Prof. Lokesh Singh, " Automatic Emotion Recognition Using Facial Expression: A Review", International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395 -0056, 3(2), February 2016, pp: 488-493
- [5] Kanade, T., Cohn, J. F., & Tian, Y. (2000). Comprehensive database for facial expression analysis. Proceedings of the Fourth IEEE International Conference on Automatic Face and Gesture Recognition (FG'00), Grenoble, France, pp: 46-53
- [6] Lucey, P., Cohn, J. F., Kanade, T., Saragih, J., Ambadar, Z., & Matthews, I. (2010), The Extended Cohn-Kanade Dataset (CK+): A complete expression dataset for action unit and emotion-specified expression. Proceedings of the Third International Workshop on CVPR for Human Communicative Behavior Analysis (CVPR4HB 2010), San Francisco, USA, pp: 94-101.
- [7] Faizan Ahmad , Aaima Najam, Zeeshan Ahmed, "Image-based Face Detection and Recognition: "State of the Art"", IJCSI International Journal of Computer Science Issues, 9(6) , arXiv:1302.6379, November 2012
- [8] Rajashree Tripathy , R N Daschoudhury , "Real-time Face Detection and Tracking Using Haar Classifier on SoC", IJECSE, 3(2), pp: 175-181
- [9] T. Padmapriya and V. Saminadan, "Improving Throughput for Downlink Multi user MIMO-LTE Advanced Networks using SINR approximation and Hierarchical CSI feedback", International Journal of Mobile Design Network and Innovation- Inderscience Publisher, ISSN : 1744-2850 vol. 6, no.1, pp. 14-23, May 20 15.
- [10] M. Rajesh, Manikanthan, "ANNOYED REALM OUTLOOK TAXONOMY USING TWIN TRANSFER LEARNING", International Journal of Pure and Applied Mathematics, ISSN NO: 1314-3395, Vol-116, No. 21, Oct 2017.

