FACIAL EXPRESSIONS RECOGNITION USING CNN BASED ON RAF DATABASE

Archan Agrawal
Dept of E & Tc
MIT , Aurangabad

Prof Akansha Somani
Dept of E & Tc
MIT , Aurangabad

Abstract: Facial expression recognition (FER) is now getting extensively popular because of its ability to predict an unknown data-set, and to its extent with some accuracy. An average human being possesses or shows seven different expressions based on the situation, namely anger, sad, happy, surprise, disgust, neutral and scared. Each individual has a unique way to express the afore-mentioned expressions and hence the term "an unknown data-set". To identify human's present mindset through facial expressions, many data sets are prepared based on face components (such as lips, cheek, nose, eyes and eye brows etc.,) dislocations and elasticity of all the facial parts. Many facial recognition systems are functioning on muscle distribution analysis from the mother image set's pixel parameters. This research paper is going to present about image pre processing, facial expression learning methods, classification methods and implementation of FaceEx algorithm for facial expression analysis through RAF CNN data sets and Viola-Jones Principle.

Keywords: Convolutional Neural Network, Facial Expression Recognition, RAF Dataset, Viola Jones Algorithm.

I) INTRODUCTION

Facial expression is a comprehensive tool that distinguishes an individual from another. Although facial expressions vary from person to person but still the underlying feelings that they showcase are the same. Significant amount of studies have been conducted on the topic facial expression recognition considering its benefits. For example, FER can help identify if a driver is fatigued or not which could prevent a possible cause of an accident , Same with the case in medical treatment. Altogether a human being shows seven different expressions - namely anger, sad, happy, scared, surprise, disgust and neutral, which varies from person to person and is not culture- specific. FER usually has three different stages – pre-processing, facial expression learning, and classification of the faces based on the emotion .

II) Convolutional Neural Network Architectures

CNN architecture is created by combining the layers . The most common form of CNN stacks a number of convolutional and pooling layers together until the input image has been spatially reduced. This intermediate output is followed by fully-connected layers, in which the last one outputs a value such as the class score .

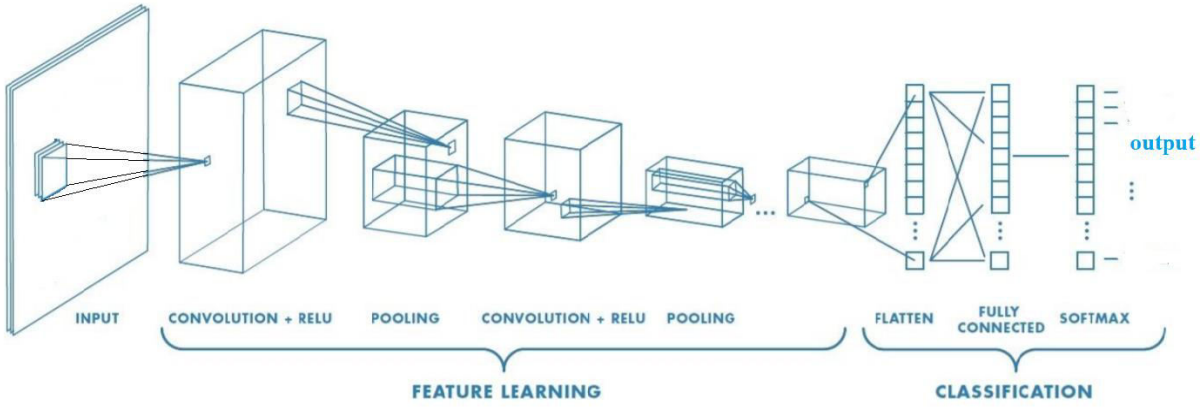

Figure 1: *Convolutional Neural Network Model*

A) Architecture Consideration

CNNs are powerful neural networks, but it is important to consider some elements during their design.

B) Input Layer

The size of the input volume should be divisible by two several times. These size values range from 32 to 512.

C) Convolutional Layer

Small filters are commonly used with a stride of one and zero-padding. This zero- padding is selected using the formula $P = F-1$ so that the size of the input volume is preserved . It is worth nothing that small filters such as 3×3 and 5×5 can be used in any layer, but large filters such as 7×7 should only be used in the first convolutional layer.

It is preferable to stack several small filters to use one equivalent large filter because the small filters express more powerful features of the input by preserving the non- linearities; and require fewer parameters.

D) PoolingLayer

It is common to use 2×2 or 3×3 filters. The reason to prefer small filters is that large filters are too lossy and aggressive which causes poor performance.

E) Strides and Zero-Padding

According to, astride of one is preferred because it preserves the spatial size of the input volume and works better in practice. Similarly, zero-padding maintains the spatial size of the input, and prevent the information at the border to disappear too quickly.

F) Training stage

In order to evaluate the CNN model and the hyper parameters during this phase, the accuracy of the model was tested against popular benchmark facial expression datasets, namely the RAF database.

III) RAF Database

Real-world Affective Face (RAF) Database used in this paper having 15339 RGB images few of them are Gray Scale images with the dimensions 227 x227 for AlexNet and 100 x 100 dimension for proposed CNN, with 96 x 96 dpi, and 24-bit depth. Moreover, the purpose of their work was to recognize the same seven facial expressions recognized in the paper. It has great variations like in the illumination, age, gender, head poses, etc.

Figure 2 : RAF Database

IV) Hyperparameters

Figure 3: Training CNN Hyperparameters

Parameters	Values
Mini Batch Size	10
Max Epochs	12
Initial Learn Rate	10-4
Learn Rate Drop Factor	0.1

Learn Rate Drop Period	20
Optimizer	Adam
Validation Frequency	500
L2 Regularization	10-4
Epsilon	10-8
Momentum	0.9

V) Viola Jones Algorithm

To test the model in real time, we have vision cascade object detector system which usesViola-Jonesalgorithm.Viola-Jonesalgorithm detects the facial features. We have put small theory that feed the image from webcam, preprocessed the image, then feed cropped face image to classify expression. Preprocessing step includes, cropping the face overlapped by bounding box then change the resolution of cropped face image to resolution of the images in dataset. It correctly predicts the expression when all parameters like light intensity, pose of head, distance of face from webcam should be right position. It correctly predicts the neutral, happiness and surprise expressions. Anger and fear expressions some times tend to mix but most of times predicts correctly. Disgust expression is rarely predicted. Most of the time sadness expression goes wrong. While real time testing it is observed that there is no delay to detect the face and classify it. Figure shows the real time testing model samples for correct classification of happiness expression with 77.87%, fear expression with 49.28% and surprise expression with 87.23% using MATLAB2020a.

VI) Results

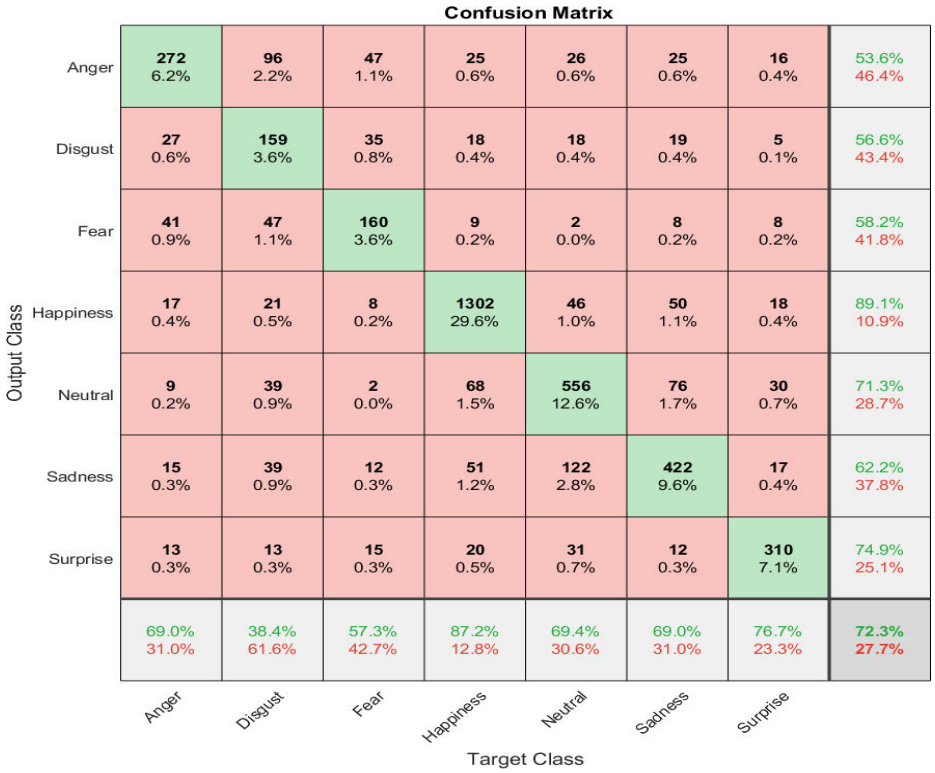

Figure4 : Confusion Matrix of proposed CNN.

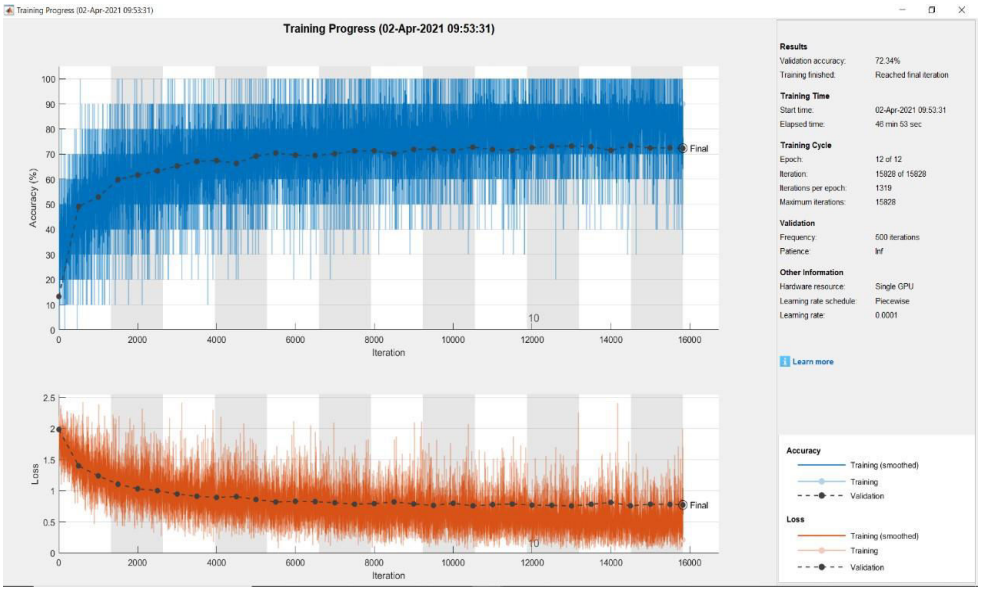

Figure 5: Training progress for proposed AlexNet CNN model

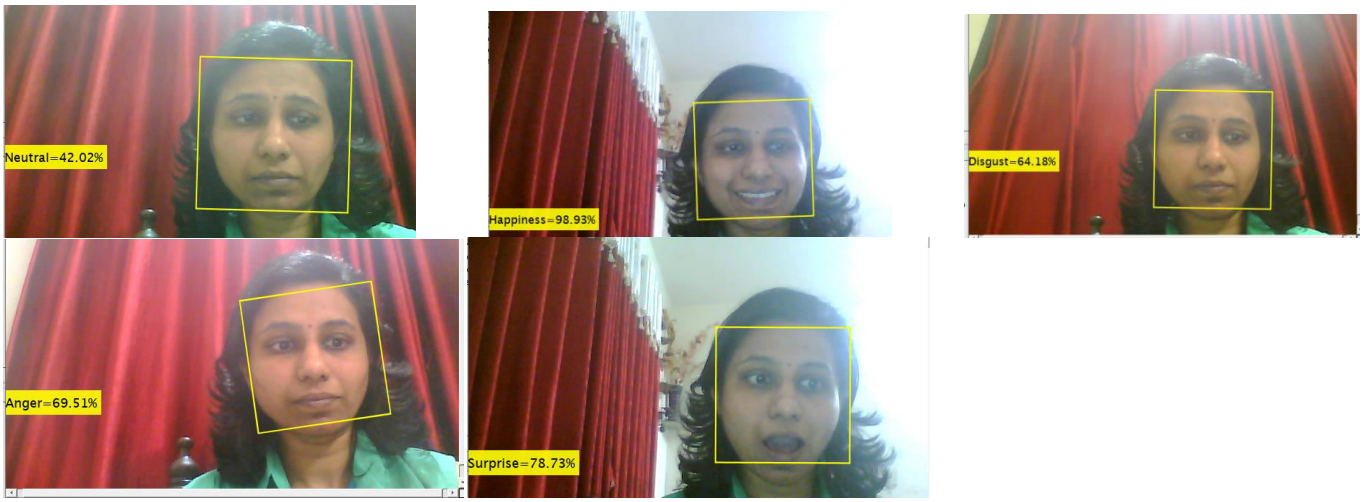

Figure 6 : Real-Time testing Sample Expressions

The training progress for pretrained AlexNet CNN model with the training accuracy 72.77%. Required training time is 193 min 38 sec, total number of epochs used are 12 and it takes 16884 total number of iterations with 1407 iteration per epoch.Validation frequency used is 500. Figure6.2shows the Confusion matrix for the AlexNet CNN model.

Figure shows the training progress for proposed CNN model with the training accuracy72.34%. Required training time is 46 min53sec, total number of epochs used are 12 and it takes 15828 total number of iterations with 1319 iteration per epoch.

Validation frequency used is 500. Figure 6.4 shows the Confusion matrix for the Proposed CNN model.

To test the model in real time, we have vision cascade object detector system which uses Viola-Jones algorithm. Viola-Jones algorithm detects the facial features. We have put small theory that feed the image from webcam, preprocessed the image, then feed cropped face image to classify expression. Preprocessing step includes, cropping the face overlapped by bounding box then change the resolution of cropped face image to resolution of the images in dataset. It correctly predicts the expression when all parameters like light intensity, pose of head, distance of face from webcam should be right position. It correctly predicts the neutral, happiness and surprise expressions. Anger and fear expressions sometimes tend to mix but most of times predicts correctly. Disgust expression is rarely predicted. Most of the time sadness expression goes wrong. While real time testing it is observed that there is no delay to detect the face and classify it. Figure , shows the real time testing model samples for correct classification of happiness expression with 77.87%, fear expression with 49.28% and surprise expression with 87.23% using MATLAB2020a.

VII) Conclusion

The approach of detecting facial expression was through a design and development of a Convolutional Neural Network (CNN) capable of predicting human facial expressions. The trained CNN model was also used for Real- Time testing of facial expressions. For training both the models that is pretrained AlexNet model and proposed CNN model, Real-World Affective face database has been used and it gives the

better performance in both training and while testing the model. 72.34% training accuracy is acquired from the proposed CNN model which is very similar to the training accuracy of pretrained AlexNet model which is 72.77%. we have also tested the model in real-time using the vision cascade object detector system which uses the Viola-Jones algorithm to detect the face in real-time and gives the better performance in the classification stage. RAF database used has great variations like in the illumination, age, gender, head poses, etc. For the better performance of a convolutional neural network, the total number of images present should be high. The performance of the model also varies from system to system.

VIII) REFERENCES

[1] Chibelushi, C. and Bourel, F. (2016). Facial Expression Recognition: A Brief TutorialOverview.

[2] Hinton, G. (2012). *Neural Networks for Machine Learning*. [online] Coursera. Available at: https://www.coursera.org/course/neuralnets [Accessed 8 Mar.2016].

[3] Y.Lv,Z. FengandC.Xu,"Facialexpressionrecognitionviadeep learning,"Smart Computing(SMARTCOMP),2014InternationalConferenceon,HongKong,2014, pp. 303-308. doi:10.1109/SMARTCOMP.2014.7043872.

[4] Bishop,C.(2006).*Patternrecognitionandmachinelearning*.NewYork:Springer, pp.227 - 249 and 256 -272.

[5] Cs231n.github.io. (2016). *CS231n Convolutional Neural Networks for Visual Recognition*. [online] Available at:http://cs231n.github.io/convolutional-networks/ [Accessed 15 Apr.2016].

[6] Murphy, K. (2012). *Machine learning*. Cambridge, Mass.: MIT Press, pp. 563 – 579.

[7] Cs231n.github.io. (2016). *CS231n Convolutional Neural Networks for Visual Recognition*. [online] Available at: http://cs231n.github.io/neural-networks-1/ [Accessed 26 Jul.2016].

[8] Poczos, B.and Singh, A. (2016). *Introduction to Machine Learning. Deep Learning*.CMU.

[9] Goodfellow, I., Warde-Farley, D., Mirza, M., Courville, A. and Bengio, Y. (2013). Maxout Networks. *JMLR WCP* 28, pp.1319-1327. arXiv:1302.4389[stat.ML].

[10] Ufldl.stanford.edu.(2016).*UnsupervisedFeatureLearningandDeepLearning Tutorial*. [online] Available at:http://ufldl.stanford.edu/tutorial/supervised/ConvolutionalNeuralNetwork/ [Accessed 15 Apr.2016]