DELHI TECHNOLOGICAL UNIVERSITY
DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING
(Formerly Delhi College of Engineering)
Bawana Road, Delhi- 110042

CO-324 Pattern Recognition

Project Report

Submitted to

Asst Prof Kavinder Singh

Submitted by

SOURABH
2K18/CO/355
3rd Year
Batch-PR D1

VISHRUTH KHARE
2K18/CO/393
3rd Year
Bach- PR D1

INDEX

Emotion Recognition by Analyzing Facial Expressions

Abstract

An important aspect of human communication is the non-verbal interaction. It comprises mainly of facial expressions along with body language. A slew of information is conveyed visually in form of facial expressions, making emotion recognition system crucial for interaction with machines. However, detecting an emotion is an exacting task due to flexibility residing over different faces like position, orientation, colour or pose. This paper encompasses various techniques for facial recognition to analyze key human emotions: happiness, sadness, anger, surprise, fear, disgust and neutral. Further we investigated them to predict the best model among them.

Introduction

Expressions are the perceptible demonstration of the sentimental, cognitive, intentional, personal and psychopathological traits. Currently a mammoth research is being carried forward in the field of digital image and image processing. Facial expression recognition being one of the most important applications of image processing opens a wide horizon of scope.

[1] Facial emotions often articulated through creation of explicit facial muscles. Emotions could be faint, yet intricate. Signals of the expressions frequently contain a copious amount of information about our state of mind. [2] While conversing only 7% consequence of

messages have a say by verbal part as a whole, 38% by vocal part and 55% effect of the speaker's message is contributed by facial expression which comprises of

I. Happiness

II. Sadness

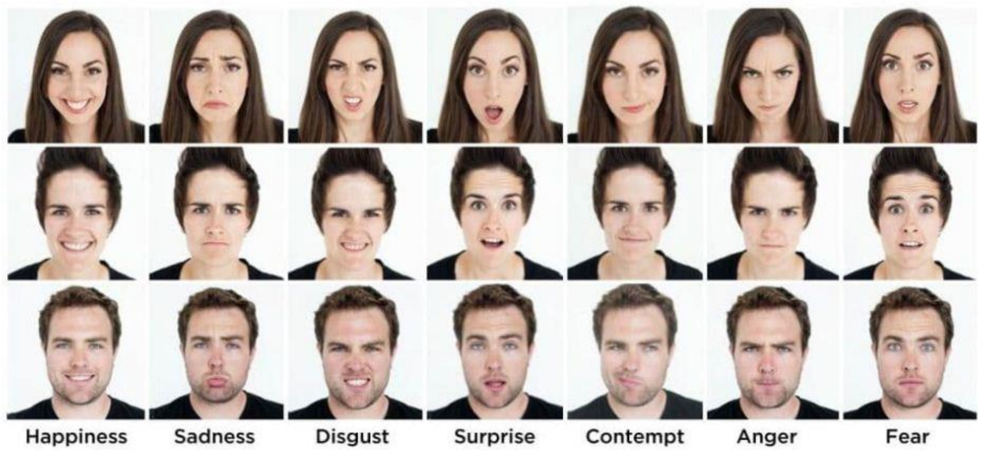
III. Disgust

IV. Surprise

V. Contempt

VI. Anger

VII. Fear



Only if machines were able to identify these emotional inputs, they could possibly provide precise and suitable hand to users in different possible behaviors which are further in harmony with the user's wants and desires.

Investigational outcomes from existing explanation demonstrate that in some conditions accurateness is superior, however time exhausted is also high, like Salman et al [8], developed Decision Trees to spot Facial Expressions. Explanations have dissimilar restrictions; like Chu et al [9] offered inexpensive facial expressions on portable devices by just extracting features, expenditure of a smaller amount time, bar the accuracy though. For the paramount way out for emotion recognition, focus needs to be laid upon accuracy and efficiency both.

Quite a lot of projects have previously been carried in this field [3], [4], [5], [6], [7]. Nevertheless, none of these papers attempted to compare different existing models for the purpose of emotion recognition. Our aspiration will be to develop an Emotion Recognition System by analyzing various approaches along with the comparison to the other available systems. Thus, we would be able to conclude with the model of best possible accuracy.

The paper at hand would provide with concise introduction to facial expression. Then describe various all-inclusive facial expressions along with their features. Detail on comparative study of popular techniques proposed earlier for Automatic Facial Emotion Recognition System would then be studied. Furthermore, different phases of Emotion Recognition System using different models have been briefly explained. Then a comprehensive comparison of different models is carried out to come up with the best possible model from the lot with its accuracy.

Classifying Facial Expressions and its Features:

Facial expression presents key mechanism to describe human emotion. From starting to end of the day human changes plenty of emotions, it may be because of their mental or physical circumstances. Although humans are filled with various emotions, modern psychology defines six basic facial expressions: Happiness, Sadness, Surprise, Fear, Disgust, and Anger as universal emotions [2]. Facial muscles movements help to identify human emotions. Basic facial features are eyebrow, mouth, nose & eyes.

Emotion	Meaning	Features
Happiness	Happiness is most desired expression by human. Secondary emotions are cheerfulness, pride, relief, hope, pleasure, and thrill.	Open Eyes, mouth edge up, open mouth, lip corner pulled up, cheeks raised, and wrinkles around eyes.
Sadness	Sadness is opposite emotion of Happiness. Secondary emotions are suffering, hurt, despair, pity and hopelessness.	Outer eyebrow down, inner corner of eyebrows raised, mouth edge down, closed eye, lip corner pulled down.
Disgust	Disgust is a feeling of dislike. Human may feel disgust from any taste, smell, sound or tough.	Lip corner depressor, nose wrinkle ,lower lip depressor, Eyebrows pulled down
Surprise	This emotion comes when unexpected things happens. Secondary emotions of surprise are amazement, astonishment.	Eyebrows up, open eye, mouth open, jaw dropped
Anger	Anger is one of the most dangerous emotions. This emotion may be harmful so, humans are trying to avoid this emotion. Secondary emotions of anger are irritation, annoyance, frustration, hate and dislike.	Eyebrows pulled down, Open eye, teeth shut and lips tightened, upper and lower lids pulled up.
Fear	Fear is the emotion of danger. It may be because of danger of physical or psychological harm. Secondary emotions of fear are Horror, nervousness, panic, worry and dread.	Outer eyebrow down, inner eyebrow up, mouth open, jaw dropped

The Dataset

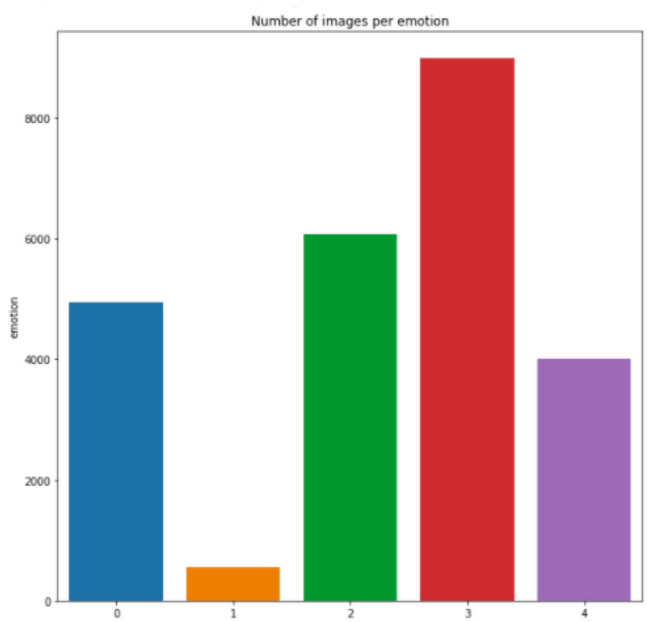
The data consists of 96 x 96 pixel grayscale images of faces. The faces have been automatically recorded as a result that the face is additionally or a smaller amount centered and lives in about the same quantity of space in every image. The job is to classify all faces on the basis of the emotion revealed in the facial expression using different classifiers in one of the seven categories as written (0=Angry, 1=Disgust, 2=Fear, 3=Happy, 4=Sad, 5=Surprise, 6=Neutral).

Link to dataset: https://www.kaggle.com/c/challenges-in-representation-learning-facial-expression-recognition-challenge/data?select=fer2013.tar.gz

1. Key facial features

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 2140 entries, 0 to 2139
Data columns (total 31 columns):
 #   Column                     Non-Null Count  Dtype
---  ------                     --------------  -----
 0   left_eye_center_x          2140 non-null   float64
 1   left_eye_center_y          2140 non-null   float64
 2   right_eye_center_x         2140 non-null   float64
 3   right_eye_center_y         2140 non-null   float64
 4   left_eye_inner_corner_x    2140 non-null   float64
 5   left_eye_inner_corner_y    2140 non-null   float64
 6   left_eye_outer_corner_x    2140 non-null   float64
 7   left_eye_outer_corner_y    2140 non-null   float64
 8   right_eye_inner_corner_x   2140 non-null   float64
 9   right_eye_inner_corner_y   2140 non-null   float64
 10  right_eye_outer_corner_x   2140 non-null   float64
 11  right_eye_outer_corner_y   2140 non-null   float64
 12  left_eyebrow_inner_end_x   2140 non-null   float64
 13  left_eyebrow_inner_end_y   2140 non-null   float64
 14  left_eyebrow_outer_end_x   2140 non-null   float64
 15  left_eyebrow_outer_end_y   2140 non-null   float64
 16  right_eyebrow_inner_end_x  2140 non-null   float64
 17  right_eyebrow_inner_end_y  2140 non-null   float64
 18  right_eyebrow_outer_end_x  2140 non-null   float64
 19  right_eyebrow_outer_end_y  2140 non-null   float64
 20  nose_tip_x                 2140 non-null   float64
 21  nose_tip_y                 2140 non-null   float64
 22  mouth_left_corner_x        2140 non-null   float64
 23  mouth_left_corner_y        2140 non-null   float64
 24  mouth_right_corner_x       2140 non-null   float64
 25  mouth_right_corner_y       2140 non-null   float64
 26  mouth_center_top_lip_x     2140 non-null   float64
 27  mouth_center_top_lip_y     2140 non-null   float64
 28  mouth_center_bottom_lip_x  2140 non-null   float64
 29  mouth_center_bottom_lip_y  2140 non-null   float64
 30  Image                      2140 non-null   object
dtypes: float64(30), object(1)
memory usage: 518.4+ KB
```

2. Facial Emotion Recognition



Different Approaches

1. Decision Tree

It is an arrangement within which every inner node stand for a "test" on top of a characteristic, all branches symbolizes the result of the experiment and every leaf node corresponds to a group label (choice taken following calculating all traits). The trails from root to leaf signify categorization policy.

It is a non- recurring leaning flowchart like plan having leaves and branches. It gets a JSON file as an input by way of chosen attributes. It comprises of aloofness flanked by sentiments and Bezier. Algorithm translates JSON to data frame kind to formulate it for the sake of straightforward assortment of diverse datasets. Preparation and test set are divided by means of the ratio of 8:2. Decision Tree Classifier approach since the sk-learn files. Subsequent to generating the training and test sets are kept in place keen on Decision Tree Classifier.

Diverse collection of points demonstrates so as to there is refusal required to choose added characteristics to classifier since they don’t provide original data. The data sets descend is keen on the similar category seeing that the first four characteristics along with the accuracy score say about the same. Count extra traits just augment the calculating time which is not needed.

2. Linear SVM

It is a machine learning classification technique which is straightforward to put into practice and offers high-quality simplification presentation, and by means of a little modification, it can resolve an assortment of troubles.
This technique will begin training on 2 lots of vectors within an n-dimensional space and concludes the hyper plane which makes best use

of the border amid two secure locations in the training set; they are generally called as support vectors. Additional computations occupy only support vectors.

Talking about the kernel function, it acts in an input space, but not inside the attribute space, therefore it generally show signs of superior dimensionality.
This function plots attributes of the input space on the way toward the attribute space which degrades the complexity.

A rightful interior produce in the attribute space is corresponded in the direction through a kernel. It does not divide teaching set linearly, but for attribute space, it will be linearly divisible. Popularly it is called as the "Kernel trick".

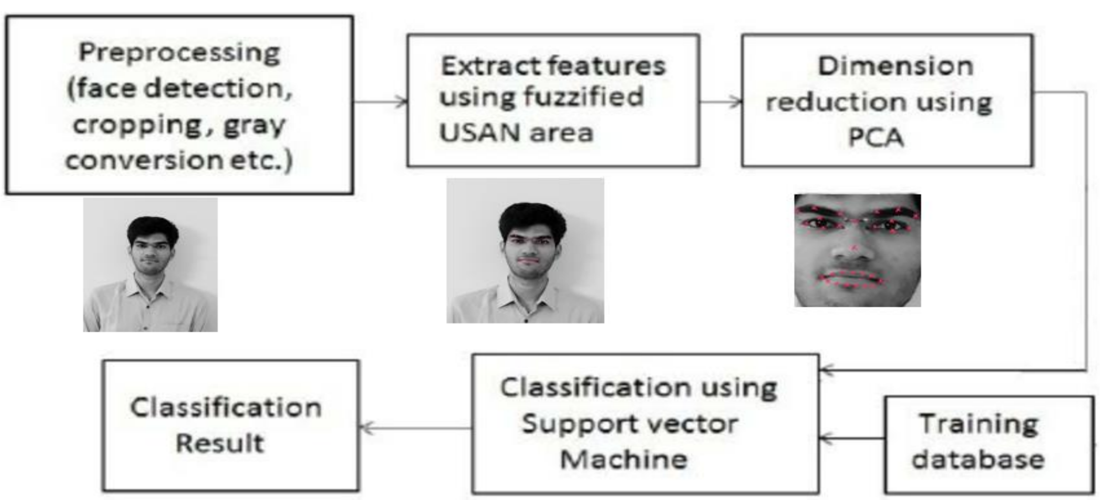
The subsequent SVM kernels used:

1. Gaussian radial basis function:

$$K(x, x') = exp(-\frac{\|x-x'\|^2}{2\sigma^2})$$

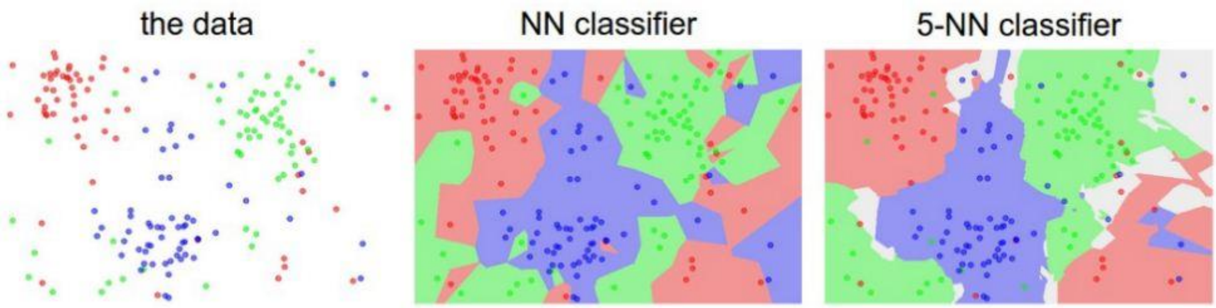
2. Exponential radial basis function:

$$K(x, x') = exp(-\frac{\|x-x'\|}{2\sigma^2})$$



3. KNN Based

The thought is as an alternative to find the sole nearby image in the set, we resolve to find the top k closest images, and contain the vote on the tag of the analysis image. In scrupulous, at the instance k = 1, we utilize the Nearest Neighbor classifier. Instinctively, elevated values of k have a smoothing consequence with the intention of making the classifier further opposed to outliers:



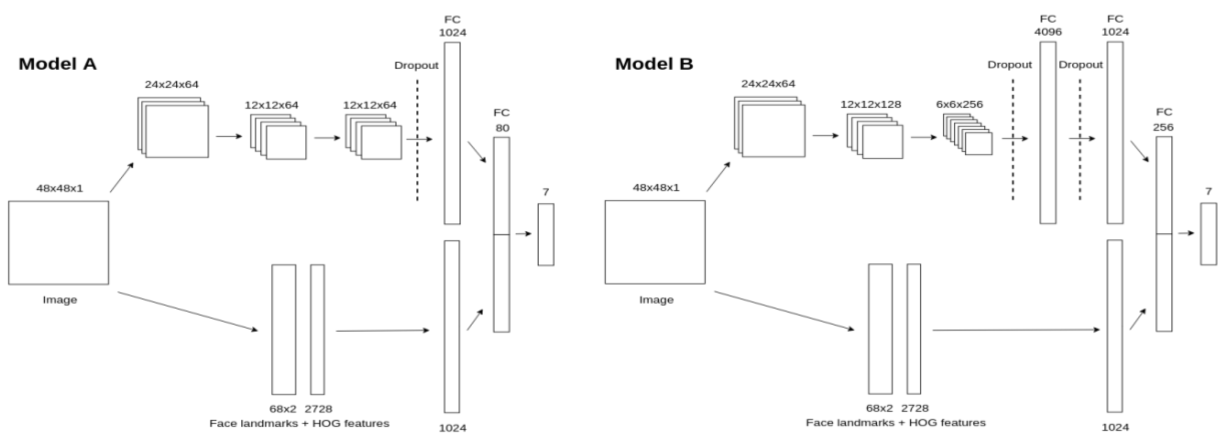
An instance of the dissimilarity among Nearest Neighbor with a 5 Nearest Neighbor classifier by means of 2D points and 3 classes (red, blue and green). The colored regions show the decision boundaries induced by the classifier with an L2 distance. The white section show points which are randomly classified. Also, in the case of a NN classifier, outlier data points (green point in the center of a cloud of blue points) generate little groups of probable incorrect calculations, as the 5-NN classifier is level in excess of these irregularities expected to lead to improved simplification on the experimental data set. Gray areas in the 5-NN picture are sourced through ties in the votes amongst the nearest neighbors.

Execution of KNN, have utilized L2 distance Norm which have the arithmetical explanation of computing the distance amid two vectors. The distance formula used is of form:

$$d_2(I_1, I_2) = \sqrt{\sum_p \left(I_1^p - I_2^p\right)^2}$$

4. CNN Based

Histogram of Oriented Gradients: A HOG relies on the property of objects within an image to possess the distribution of intensity gradients or edge directions. Gradients are calculated within an image per block. A block is considered as a pixel grid in which gradients are constituted from the magnitude and direction of change in the intensities of the pixel within the block.



We have worked on CNNs among the changeable depths in the direction of assessing the results of these models for facial expression recognition. Network architecture considered is:
[Convolution-(SBN)-ReLU-(Dropout)-(Max-pool)]M - [Affine-(BN)-ReLU-(Dropout)] N - Affine - Softmax.

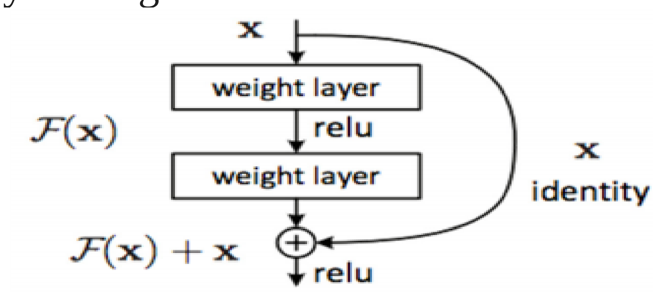
The 1st element of the system submits to M convolution layers with the intention of having spatial batch normalization (SBN), max-pooling and dropout along with addition toward the convolution layer. Furthermore, M convolution layer, the system is directed to N completely associated layers with the intention of at all times having Affine procedure and ReLU nonlinearity, it may comprise group normalization and dropout.

At last the network is pursued through the affine layer with the intention of computing soft-max loss function. So the completed production furnishes the customer the liberty to make a decision on the subject of the quantity of convolution and entirely associated layers, as well as the existence of normalization, dropout and max-pooling layers.

Besides, the quantity of filters, strides, and zero-padding is able to be specified by user. The default (pre-fed) data is well thought-out if they are not set. The planned idea of uniting HOG features with the ones extracted by convolution layers through mean of unprocessed pixel data. We made use of the similar structural design explained, other than by means of this dissimilarity that added the HOG features to those which ways out the last convolution layer. The hybrid feature then lays down cross the threshold the completely related layers for score and loss estimation.
We put into practice the aforesaid representation in Torch and took benefit of GPU increased speed deep learning features to create the model training course quicker.

5. RES-NET Based

As CNNs grow deeper, vanishing gradient tend to occur which negatively impact network performance. Vanishing gradient problem occurs when the gradient is back -propagated to earlier layers which results in a very small gradient.

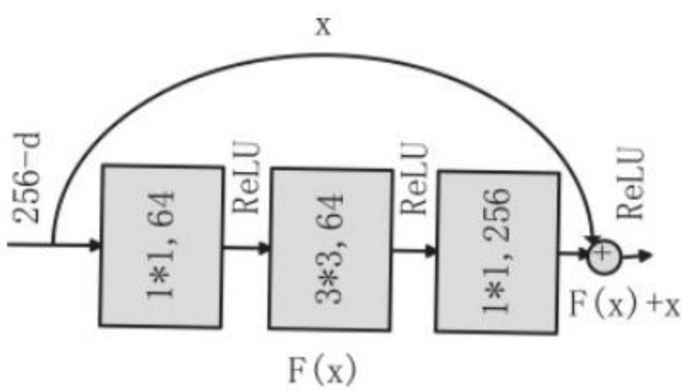


The profundity of the system is vital to the presentation of the model. As soon as the number of network layer are augmented, the network becomes capable of moving out extra compound feature outline extraction, consequently enhanced outcomes be able to be find supposedly when the model is deeper. On the other hand, the trial originates that the deep network was degenerating.

In the midst of the boost of network depth, the correctness of the network is inclined in the direction to be saturated or still reduced. Presently there is a reduction in the accurateness of the training set. We are able to settle on that, this is not on the basis of over-fitting. Since the correctness of the training set subsists, supposed to be elevated in the situation of over-fitting. The remaining network in ResNet is intended to explain this trouble, and subsequent to solving this difficulty, the profundity of the network climbs by more than a few orders of scale.

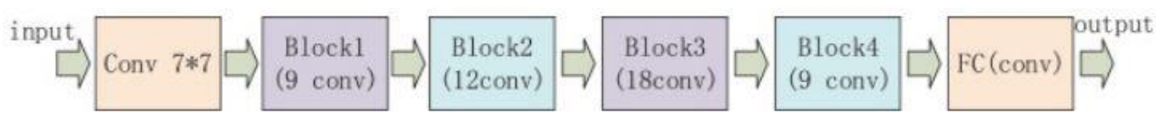
Res-Net planned 2 class of mapping:

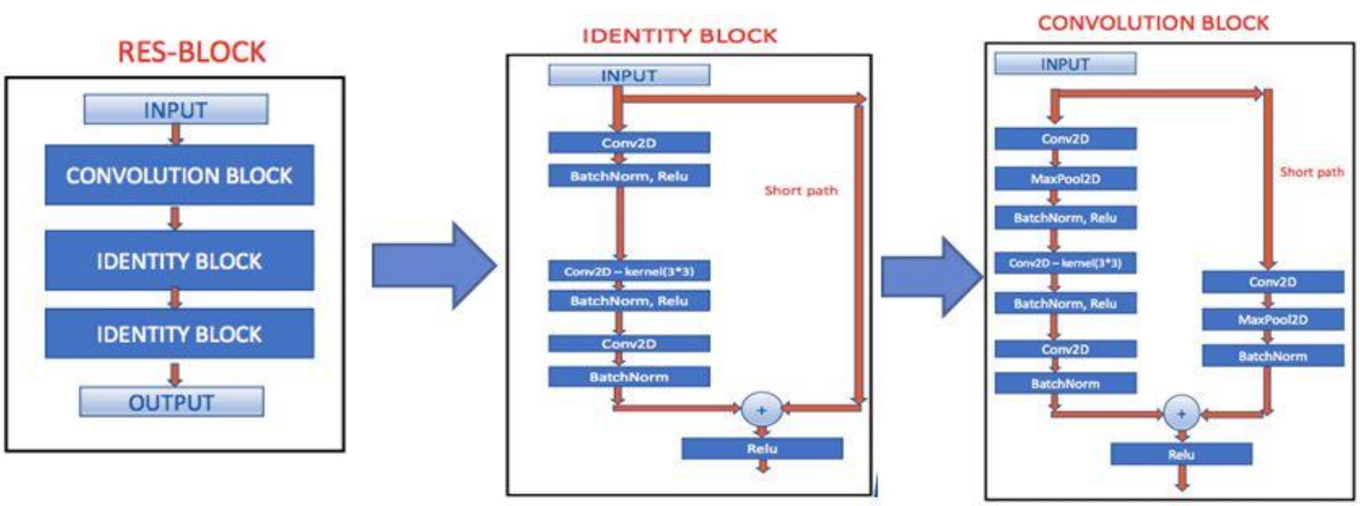
- First is individuality mapping, corresponding to the "curved curve" in figure
- The second is residual mapping, which means in the direction of the element apart from the "curved curve", as a result the final output is $y = F(\mathrm{x})+x$.



Identity mapping, while the given name entails, correspond to itself, which is x in the formula, at the same time since residual mapping corresponds to "dissimilarity", which is, $y - x$, so remaining refers to F(x). Residual Neural Network comprises of “skip connection” feature which allows training of many layers devoid of disappearance gradient issues. The mechanism works by adding up “identity mappings” on summit of CNN.

In beginning, ResNet-50 carried out convolution process on the input, pursued by four residual blocks, and in the end it established complete association process to attain categorization work. The system arrangement of ResNet-50 is shown in figure, which has fifty Conv2D operations.

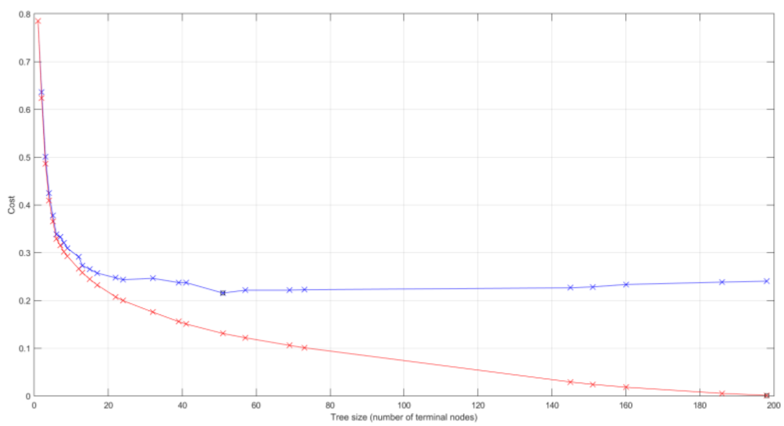




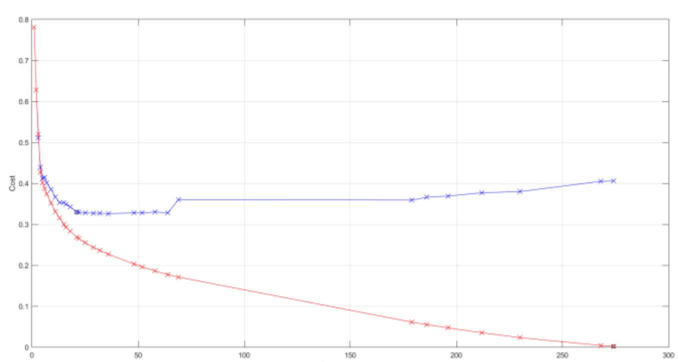
Result and Analysis

Classifier	Test Accuracy
Decision Tree	69.4%
Linear SVM	38.4%
KNN Based	60.5%
CNN Based	65.9%
RES-NET Based	78.1%

- Decision Tree:

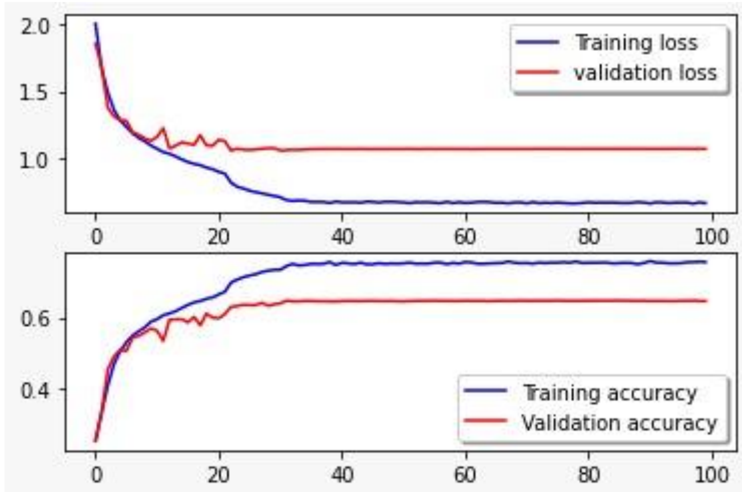


(pruning_example function on clean data)

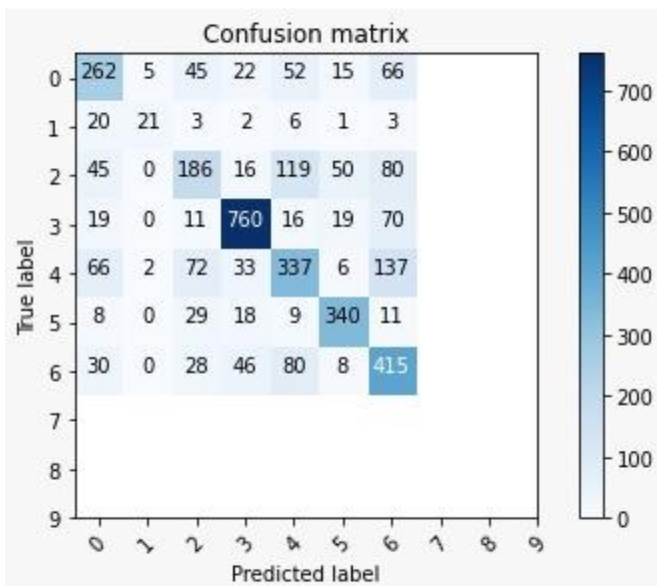


(pruning_example function on noisy data)

- Convolutional Neural Network

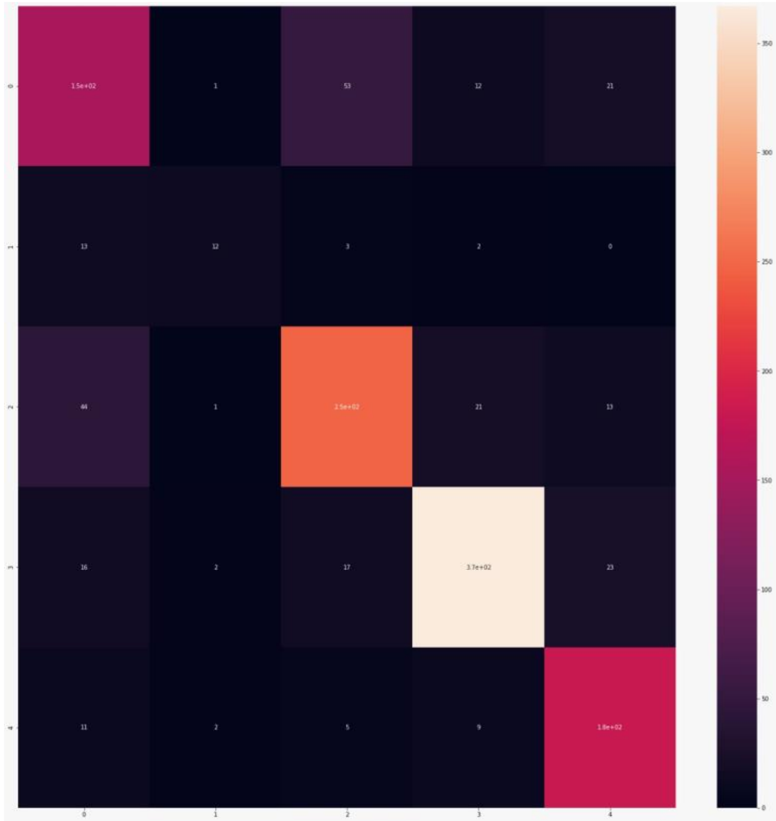


(Loss and Accuracy)

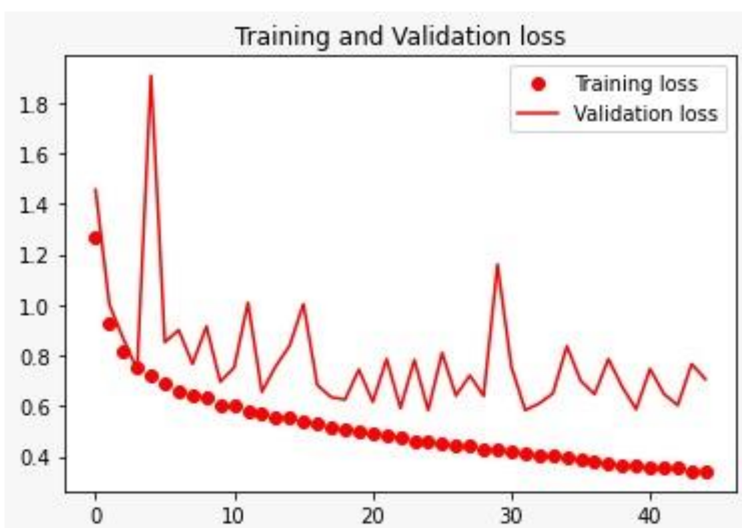


(Confusion Matrix)

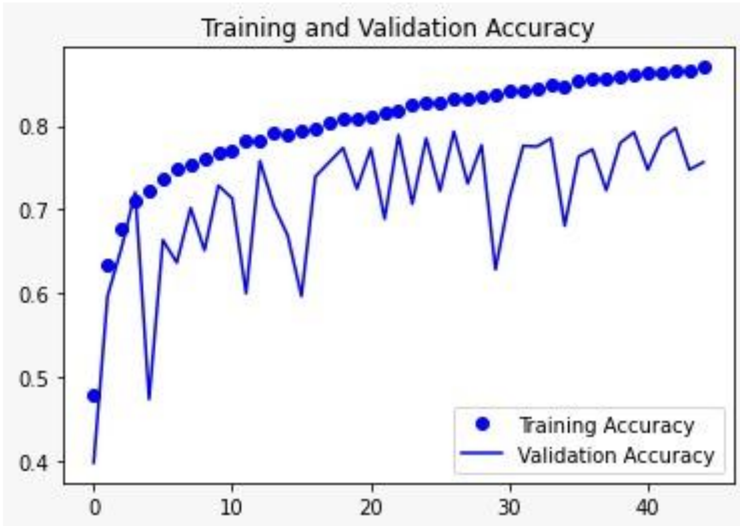
- RES-NET Based



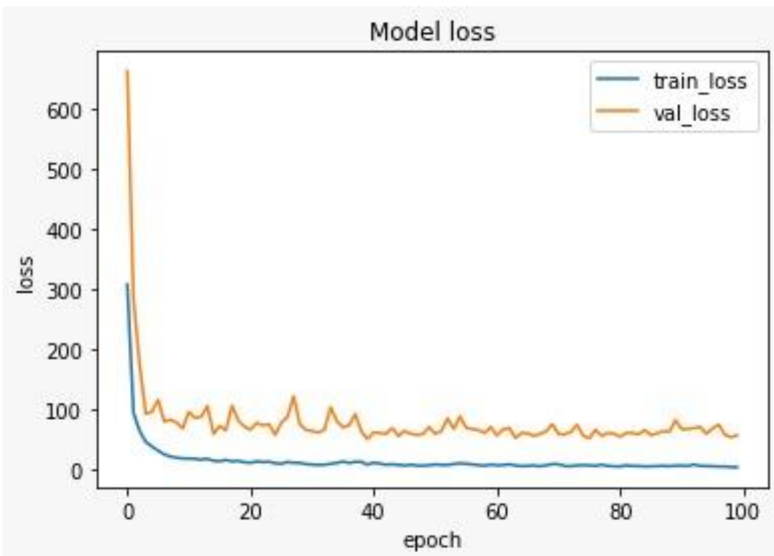
(heat map)



(Training and Validation Loss)



(Training and Validation Accuracy)



(Model Loss)

Conclusion

This learning inspected the presentation of a variety of algorithms on classifying facial expression data. In general, we were able to classify emotions based on the facial expression using all the classifiers.

With most average accuracies being in the 85% range all the models performed fairly well. However the classifier with maximum accuracy was

Future Scope

Even though this learning gave imminent knowledge into learning algorithms, some had better accuracy than others for facial expression recognition; few issues need to be considered in future before making a comprehensive record of algorithms or generalizing the algorithms' results to most facial expression recognition.

First, this work did not experiment a range of factors for the algorithm functions, also it does not explore characteristic selection. Some previous works have proven that feature selection by means of Ada-Boost and after that using SVM showed improved accurateness, insred of running them alone.

One more restraint was the likelihood to over fitting. Consequently, in regulation to create a universal statement of classifying emotion, the algorithms should also be tested on different AAM, AU, and emotion-labeled datasets.

Also, if this study has to be utilized towards building real-time emotion recognition tool, calculation rate and capacity to dissimilar environments like origin of person and lighting in the room has also to be considered.

References

[1] Mase K. Recognition of facial expression from optical flow. IEICE Transc., E. 74(10):3474–3483, 0ctober 1991.

[2] A.Mehrabian, ”Communication without Words” Psychology Today, Vol.2, no.4, pp 53- 56, 1968

[3] Black, M. J. and Yacoob, Y. Tracking and recognizing rigid and non-rigid facial motions using local parametric model of image motion. In Proceedings of the International Conference on Computer Vision, pages 374–381. IEEE Computer Society, Cambridge, MA, 1995.

[4] Essa, Pentland, A. P. Coding, analysis, interpretation, and recognition of facial expressions. IEEE Transc. On Pattern Analysis and Machine Intelligence, 19(7):757–763, JULY 1997.

[5] Mase K. Recognition of facial expression from optical flow. IEICE Transc., E. 74(10):3474–3483, 0ctober 1991.

[6] Tian, Ying-li, Kanade, T. and Cohn, J. Recognizing Lower Face Action Units for Facial Expression Analysis. Proceedings of the 4th IEEE International Conference on Automatic Face and Gesture Recognition (FG'00), March, 2000, pp. 484 – 490.

[7] Yacoob, Y., Davis, L. Computing spatio-temporal representations of human faces. Computer Vision and Pattern Recognition, 1994. Proceedings CVPR '94., 1994 IEEE Computer Society Conference on , 21-23 June 1994 Page(s): 70 –75.

[8] Salman, Madaini, & Kissi (2016) " Facial Expression Recognition using Decision Trees". IEEE International Conference on Computer Graphics, Imaging and Visualization (CGiV), 120-130.

[9] Chu, Chen, & Hsihe (2015) " Low-cost Facial Expression on Mobile Platform", IEEE International Conference on Machine Learning and Cybernetics (ICMLC), pp 586-590,.