Human Facial Expression Recognition Using TensorFlow And OpenCV

Thesis · August 2020

DOI: 10.13140/RG.2.2.19218.89288

1 author:

Saransh Srivastava
VIT University
3 PUBLICATIONS **0** CITATIONS

SEE PROFILE

Human Facial Expression Recognition Using TensorFlow And OpenCV

Saransh and Dr. Muthamil Selvan T
School of Information Technology and Engineering, VIT University, Vellore

saransh.2018@vitstudent.ac.in, and tmuthamilselvan@vit.ac.in

Abstract:

This project is a real time recognition system that traces the very mood of the human. Human expresses their mood and sometimes what they need through their expression. It can be a smiling face, or it can be the face full of anger. Sometimes words are not that powerful as our expressions. This project consists of models made through various algorithms of machine as well as deep learning. It also uses some of the very powerful packages in python to create an application software that recognizes the expression of human in real time. Some of the libraries are: TensorFlow, Keras, OpenCV, Matplotlib. This implementation can be used at various places and platforms. The very first example can be feedback through moods at any restaurants and hotels about their services and foods. It can be much impactful in the field of military. Its very usage can be helpful for recognizing the people's behaviour at the border areas to find out the suspects between them. This project is just an implementation of the environment and not the actual software that can be useful in the real time environment. This project consists of two modules: (i)Processing and generating the model for the application using different algorithms and (ii) Application for using the model using OpenCV to recognize the human facial expression.

I. Introduction:

This project uses dataset from Kaggle website which consists of 48x48-pixel grayscale images of face. This project focuses more on improving the previous models which have less accuracy. On the grayscale images of face pixels of forehead can be included in few emotions. This project proposed a model which is more accurate and faster than the previous models. Python has a very strong module for deep learning i.e. TensorFlow which runs deep learning neural Convolutional networks for digits (which are handwritten) classification, image pre-processing and recognition, sequential models for translation, natural language processing (NLP), and partial differential equation (PDE) based tasks. For real time image processing, this model uses a library called OpenCV. Various other libraries e.g. NumPy, Pandas, Matplotlib has also been used for loading, visualizing and analysing the data. This model also uses Keras which is one of the libraries which is used to code deep learning models. In its back end it uses TensorFlow.

This project substantially includes the attributes required for a thorough facial expression recognition. Though there seven of them, and two with least no of tuples which has been removed for better results. This project takes the idea from day to day lifestyle that people often think as of least importance. Think of a restaurant which has many customers and they are regular. It would be a bad day for the owner of the restaurant if a single person leaves to go to another restaurant. Many customers either hesitate or does not bother to give feedback for the service they received at a restaurant or hotel. And the person who gives them service often missed or does not understand what the customers feel. What if one can be able to install a system that tracks the emotions of people from their face at the bill counter? The feedback recorded in it can be used for further improvement of their service. Now think of a security risk that a mob poses at the time of protest. It can be peaceful, but nobody knows what is in their mind. What if we have an efficient camera installed at each security personnel vest that doesn't only recognises faces but also emotions. It is very hard for a person to hide his/her intentions if he/she is not professional. A violent mob can be easily be brought back to normalcy if the security personnel have enough equipment including that facial emotion recognition camera.

Computer Vision:

Leading researches in computer vision encompasses the very ability to provide any system to have a high level of understanding of images and videos as humans have. Computer Vision is a next step in field of Artificial Intelligence. It enables the system to get even a higher level of information which sometimes a human can't. An example of it can be seen as the medical imagining of bones and muscles in different parts of the body which is so sophisticate that even a doctor sometimes can't understand that. This area revolves around image analysing, processing and get an understanding in form of machine level code. Computer vision can help in several methods including acquiring, processing and recognising digital images, and taking multi-dimensional data from the real time in order to obtain minute information which can be either numerical or symbolic. The research or breakthrough made in this field is concerned with the problems in extracting information about digital images artificially. There are no neurons that works in a computer as it does in a human brain that increases the capacity to understand images. These researches are generally funded by industries and organisations to find out solutions that a human mind can't. Computer vision has a vast field of study and research and researchers have categorised the computer domains in scene reconstruction, video tracking, object recognition and detection, motion estimation, and image restoration. Everybody dance now is one of the most notable and most motivated work has done so far in this field. It includes motion translation. The transition made between two frames and that also with another person. In this highly motivated project one can be seen dancing perfectly in a video by replacing its frame motion with another frame. A lot of research is still needed in this. This has also been made to public domain. Another impressive milestone that were achieved in this field has been made in the area of neurobiology, signal processing, information engineering and artificial intelligence. All of them

helping humans to enhance efficiency and get better results day by day. It is always used as a tool in assisting humans for identifying tasks at grass root level.

II. Related Work:

In paper [1], The authors have shown a Deep Learning alignment work, which is a vigorous face calibration procedure that is based on Convolutional NN. They have proposed Deep Alignment Network performs the face calibrations mostly depends on the whole face images in contrast to what recently face alignments techniques perform, which make it very accurate to immense fluctuations in both initializations and forehead poses. Using heatmaps which has landmark, and which transmits the detail of the locations of landmarks among DAN phases, it helped them to use face images instead of locally available marks which is extracted around landmarks. Extensive performance evaluation improves the ultra-modern failure rate by a relatable limit more than 70% which were performed on two different challenges.

In paper [2], The authors have described the system "Affective Computing" as to develop systems, devices and mechanisms those of which are recognizable, interpretable, and which imitates a person affects through various attributes such as how he/she looks, the depth and modulation in his/her voice, and biological signals he/she may have. They have discussed about several network architecture driven models in their literature to shed lights on emotive facial expressions: 1) explicit, where the emotion is fetched from an emotive-related category such as FER datasets which have six basic human emotions in it. 2) extent, where a numerical value is taken from a simultaneous face expression scale in images which are valence and arousal.

In paper [3], The authors have shown the facial expression recognition system which is a real-world application and solves the phases occurred post changes made. The authors have generated the several new tests over FER datasets on these phases and proposed a new "Region Attention Network (RAN)" which itself depicts the importance of the facial landmarks. They further shown the implementation of a "Region Biased loss (RB-Loss)" function that is used to strengthen the high attention weight for regions which are the most salient. The authors also evaluated their method on the collection of their datasets and made the extensive studies on FER-Plus and Affect-Net. The work proposed the method which achieves the ultra-modern results on different datasets which includes FER+, RAF-DB, SFEW, and Affect-Net.

In paper [4], The authors have made their outlook on a effort in progress technique for the facial expression recognition which enables the system to get much from the facial landmarks. The findings that are figured on the JAFFE-dataset which suggested some signs for a place for the development and more precision. The authors have made their overview saying that the proposed method has strong potentials that can outperform the currently proposed methods.

In paper [5], The authors propose a Convolutional Neural Network technique which is a 3-Dimensional for FER in frames of videos. This model develops an 3D Inception-ResNetlayers followed by a unit called LSTM that simultaneously grasps the relations of spatial within images of faces and the temporal instances among different frames of the video. Facial curve dots are also used as samples to their network design which focus on the instances of facial landmarks

rather than some noted facial patches that won't be beneficial and may not be able to generate facial expressions significantly.

In paper [6], There is a research conducted by the author to categorise the facial emotions over the static facial pictures with the help of deep learning techniques. The results that were achieved were non-futuristic, and slightly better than other methods including the characteristics engineering. It means that eventually Deep Learning systems will be able to remove this problem given an ample amount of the labelled tuples. Characteristics engineering is not that essential, image pre-processing reduces the inconsistencies of the classification. That's why it increases the visibility and the quality on the input image. In today facial emotion detection software includes the use characteristics engineering. A finding that is totally dependent on the characteristic learning that does not seem near yet because of the major restraint and that shows the absence of a wide-ranging dataset of reactions. With the presence of a bigger dataset, systems that have a larger ability which is used to learn structures that could be applied. Thus, emotion classification could be attained with the help of deep learning approaches.

In paper [7], the authors have proposed an architecture where convolutional neural network (CNN) are trained to classify facial emotions/expressions. The authors have used Japanese Female Facial Expression (JAFFE) dataset of facial emotion images for training CNN in order to achieve good accuracy during training phase. Concept of Hybrid Vehicle Employing of CNN has been used for detecting drowsiness or alertness of the drivers in real time.

In paper [8], the author has proposed a system of programmed facial Expression Recognition to perform detection and location of faces landmarks in a muddled scene, set of facial movements extraction and facial emotions classification. This model is developed using Convolutional NN which is totally dependent based on a network design called "Le-Net", Kaggle facial expression (FER2013) dataset with seven facial expression class labels which includes happy, sad, surprise, disgust, fear, anger & neutral.

In paper [9], the authors have worked on the channel link and in which have proposed a fresh design unit, termed as "Squeeze and Excitation (SE)" block which tries to set right features channel wise by manipulating channel since they are independent. This paper has showed that chunks of patches can be loaded together to form SE-Net architecture to generalize extremely effectively across different datasets. "Squeeze and Excitation" Networks has formed the foundation of ILSRVC classification submission.

In paper [10], the authors have provided a complete survey on a design which is deep "Facial Expression Recognition (FER)" which includes databases and algorithms that features selection of data acceptance and evolution designs for these sets of data. The authors have reviewed some already constructed Deep Neural Network Models and related training modules designed for "Facial Expression Recognition 2013" based on sequential images which are static and dynamic as well.

Overview of FER:

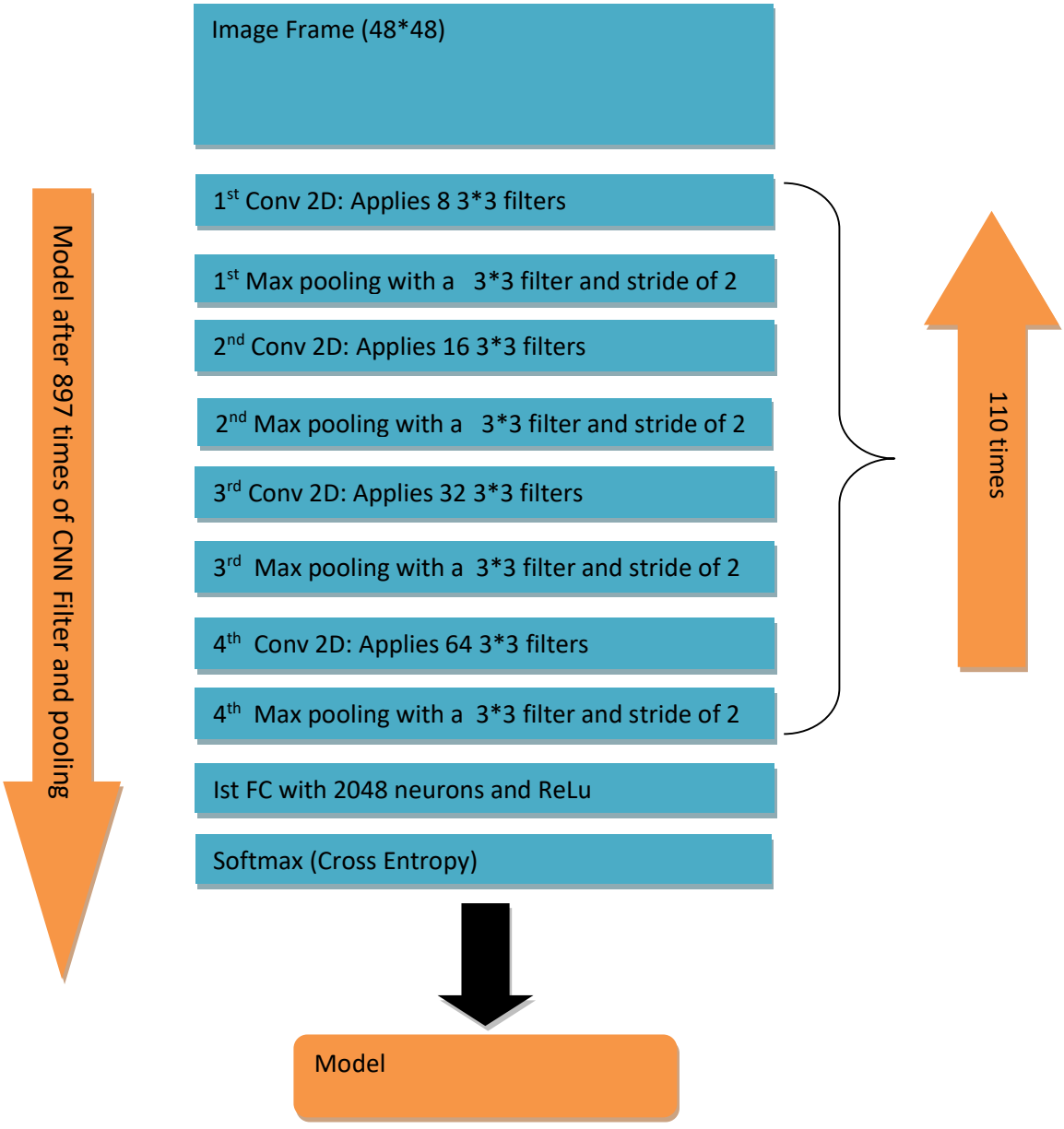
The data consists images of faces of 48x48 pixel grayscale. These images have been registered so that the face covers large parts in the centre of the image and fits in same amount of space in each image. The Kaggle in 2013 hosted a competition to categorize this dataset in one of the following moods (0=Angry, 1=Disgust, 2=Fear, 3=Happy, 4=Sad, 5=Surprise, 6=Neutral). The fer2013 dataset used in this project contains two columns, "emotion" and "pixels". The "emotion" column contains a single digit ranging from 0 to 6 (inclusive) as described above for the emotion that is present in the image. The second column contains a string for each image. These strings are space-separated pixel values in row order. The dataset contains only the "pixels" column and the task was to categorize the emotion column. The training set consists of 28,709 examples. From their research, Pierre-Luc Carrier and Aaron Courville have provided this dataset in public domain. Various implementations and comparisons have been made through models generated through various algorithms using different technologies. This project uses algorithms which shows how this dataset can be improved furthermore in order to give more emphasis on the important attributes of the face and the alignment. Accuracy percentage predicted through the models will give a better idea for the improvements that can be made through several iterations of pre-processing. The following table consists the number of tuples available with each attribute which has been further categorised in training and testing tuples.

III. Existing Methods:

Automation has always been a centre of focus in this 21st century. Though it cuts jobs but at the same time it provides the better and standard way of living in the society. In modern era, traditional businesses are on the verge of almost leaving the race since new market leaders have come up with ideas of intelligent businesses. A business that consists of highly advanced systems for providing and taking decisions for the organizations. What would be the case if we would be able to implement a real-world application that can collect feedbacks from customer directly through their expression. What if we would be able to develop a system that can detect any anti-social activities before it happens through the mood of a mob. Human facial Expression recognition has applications in various fields. From customer feedback to criminal confession and sometimes finding the anti-social elements in the crowd. This project mainly focuses on the grayscale image conversion. The data is mainly in the form of grayscale which is matrix. It also uses some of the most used open source libraries which comes ready made with n numbers of algorithms. It contains codes with implementations. The very use of these libraries helps the later generated model to achieve the maximum accuracy. Furthermore, this project has most of the works been done on sublime text which is an efficient editor for Python3 interpreter. This should be noted that since the libraries are open source, it can show some problems when the environment variables won't match with current versions of the libraries. The dependencies of libraries also rely on the versions that one uses. This project uses the very current stable version of Python3 i.e. Python3.5. The libraries have been downloaded from PYPI i.e. Python Package

Index which a software repository where many works are published. Currently it has 113,000 libraries including more than 10,000 for data science. Even though CPU has been for the epochs of the Convolutional Neural Network, this project doesn't oversee the system requirements since it can't work properly on less than a GPU.

The first Module is the consists of several CNN filters and Max Pooling which includes pre-processing steps. The data here is pre-processed and after the data is filtered and pooled through various iteration, it generates the model that is used in the next module. Below is the diagram for the CNN filtration and Max Pooling.



The images are taken in real time from the video capture in OpenCVin the second module. The images are then converted in the grayscale images 48x48. The grayscale image then is matched with model we have generated in the first module. This module uses HaarCascade classifier to detect faces of the user. Faces need to be detected before it can be converted to a grayscale image.

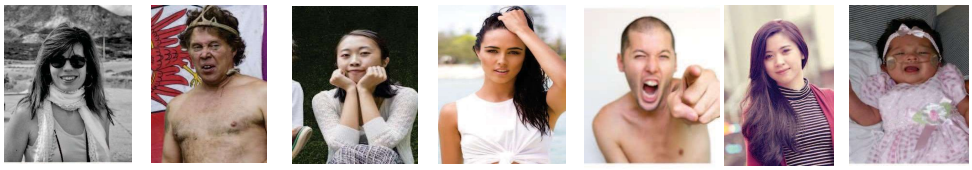
Related Dataset:

JAFFE

CK+

FERplus

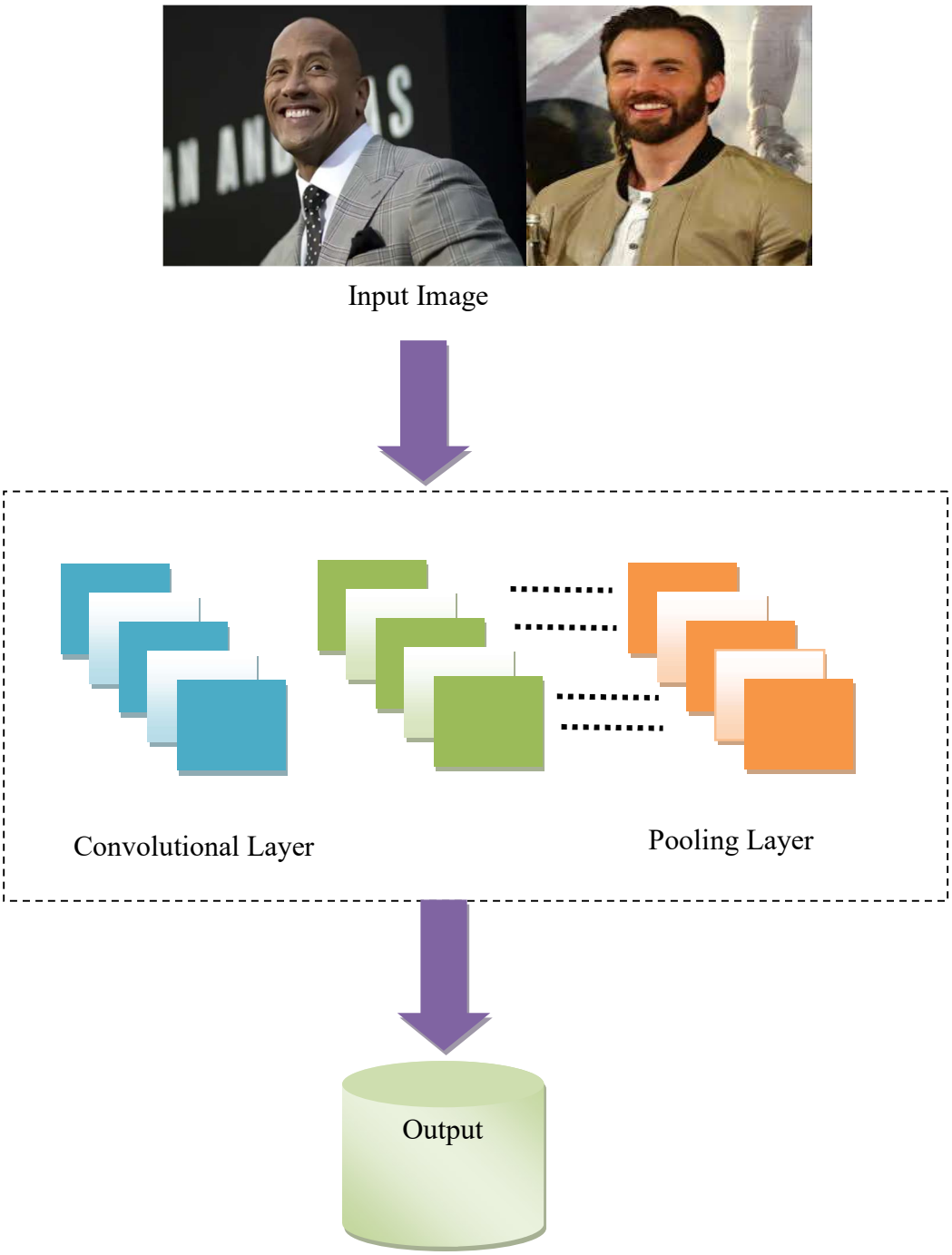
RAF-DB



Convolutional Neural Network:

Convolutional Neural Network or CNN/Conv-Net is an algorithm of Deep learning. An input image is feed for the algorithm to assign learnable weights and biases and try to find importance to various characteristics in the picture provided, these networks helps to differentiate each characteristic from one another. The important feature of CNN is that pre-processing needed in this is much lower when compared to another algorithms (classification). The network neurons architecture in Convolutional Neural Network is somewhat similar to patterns that human brain cell has while connecting to each other. The Receptive Field which is the visual field of the restricted region where single neurons respond to stimuli. The whole area (visual) is covered with

a collection of such fields which overlaps. The below example shows how an input image of a handwritten digit which is being feed to the Conv-Nets which goes through pooling layers.

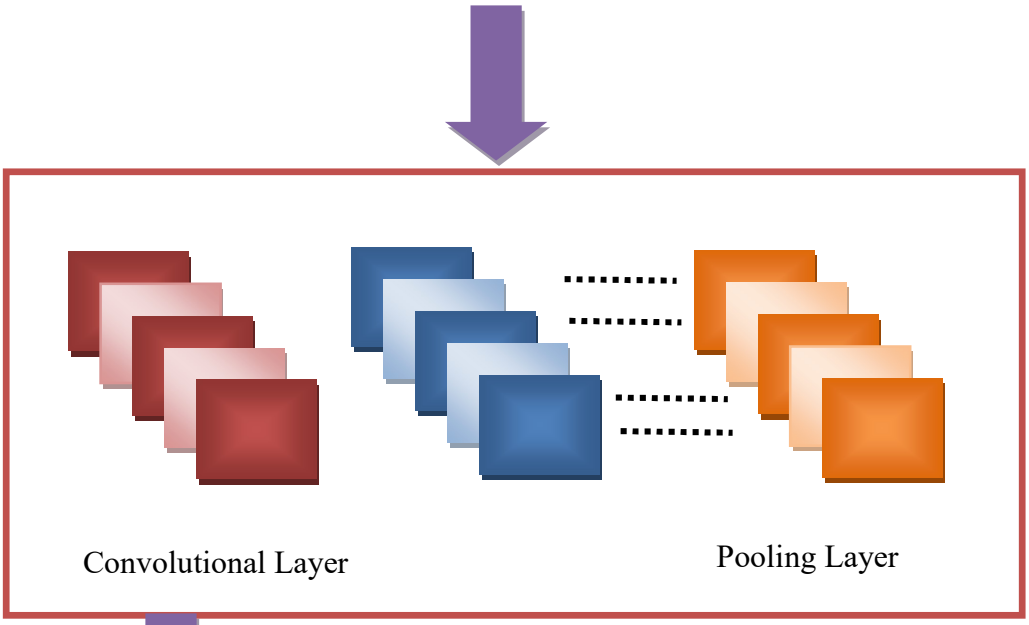


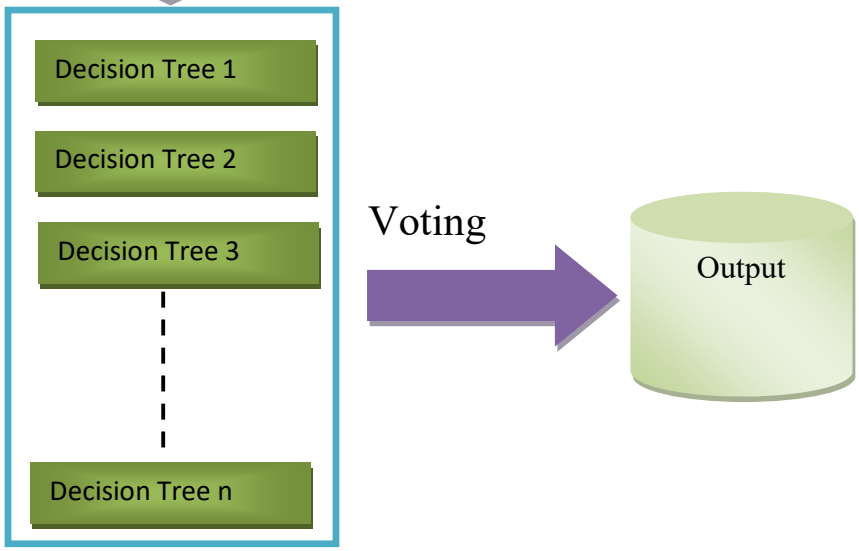
IV. Proposed Model:

The proposed model has been categorised into two major portions: first part is regarding the removal of the tuples whose attributes are of least importance and another part deals with the adding of the other algorithm for improving the model. Disgust and fear are the attributes with least number of tuples both for testing and training. By weeding out these attributes, the performance improved for the model. The propose model uses random forest that is applied to models which is generated from the Convolutional Neural Network. The decision trees that are generated by applying any algorithm of decision tree are voted for the best among the groups of several decision trees. The random forest works on a principle of Divide and Conquer which is why it is called an ensemble machine learning technique. The model is already an improved version from previous models after weeding out least important tuples from the FER dataset. This algorithm uses the combination of several trees. This project uses C4.5 algorithm for generating Decision tree. Since it was needed to just compare the algorithms and get the best accuracy, only algorithms have used to generate the model in this module and not for the application driven systems.

Ross Quinlan developed C4.5 to generate decision tree which has preceded ID3. Classification using generated decision tree through C4.5 algorithm is inferred as statistical classifier. The Weka provides J48 classifier for generating trees using C4.5 algorithm. The generated trees will be voted to get the best tree consisting less inconsistency and high accuracy. C4.5 algorithm has already been described “a landmark decision tree” by the creators of The Weka Machine Learning Software and stated most widely used in practice till date. The number of correctly classified instances are 77% and number of incorrectly classified instances are 23%. The classifier accuracy reaches the accuracy of 66% which almost 10% more then the previous model. Since the model used is the one generated from the last one that is why the number of labels is same as the previous one. The mean absolute error was found to be around 0.4983 which is also lesser than the previous model. The model has an accuracy of 98% for the training cases and 62% for the testing cases which is more than the previous model. The below picture is the proposed model which succeed the earlier one only with the change of Random forest being used for classification before getting the output.

Input Image


Convolutional Layer
Pooling Layer


Decision Tree 1
Decision Tree 2
Decision Tree 3
Decision Tree n
Voting
Output

V. I/O Screenshots:

This project has been divided into 2 modules i.e (i) Model Generation Modules (ii) Live video Capture Module. Fig1 is the flowchart of the Convolutional Filters and Max Pooling in which the data are the coordinates in the 48x48 pixel grayscale. It goes through several iteration. Minimum iteration needed is 255 for CNN filtration and max pooling. The model generated through this module uses keras for pre-processing and TensorFlow uses deep neural networks algorithm to enhance the model. After several iteration the model reaches an accuracy for 95% for training tuples and approx. 56% for testing tuples. Below is the snippet of the model generation module. The model is going through several Iterations or epochs in order to train and test the data.

```
25/256 [=>............................] - ETA: 2:50 - loss: 1.0105 - acc: 0.6216
26/256 [==>...........................] - ETA: 2:49 - loss: 1.0076 - acc: 0.6230
27/256 [==>...........................] - ETA: 2:48 - loss: 1.0091 - acc: 0.6227
28/256 [==>...........................] - ETA: 2:50 - loss: 1.0085 - acc: 0.6218
29/256 [==>...........................] - ETA: 2:52 - loss: 1.0102 - acc: 0.6203
30/256 [==>...........................] - ETA: 2:53 - loss: 1.0092 - acc: 0.6214
31/256 [==>...........................] - ETA: 2:51 - loss: 1.0085 - acc: 0.6224
32/256 [==>...........................] - ETA: 2:50 - loss: 1.0059 - acc: 0.6239
33/256 [==>...........................] - ETA: 2:49 - loss: 1.0052 - acc: 0.6245
34/256 [==>...........................] - ETA: 2:47 - loss: 1.0040 - acc: 0.6242
35/256 [==>...........................] - ETA: 2:46 - loss: 1.0006 - acc: 0.6249
36/256 [==>...........................] - ETA: 2:45 - loss: 1.0007 - acc: 0.6247
37/256 [==>...........................] - ETA: 2:44 - loss: 0.9964 - acc: 0.6270
38/256 [==>...........................] - ETA: 2:43 - loss: 0.9972 - acc: 0.6271
39/256 [==>...........................] - ETA: 2:42 - loss: 0.9947 - acc: 0.6279
40/256 [==>...........................] - ETA: 2:41 - loss: 0.9932 - acc: 0.6278
41/256 [==>...........................] - ETA: 2:40 - loss: 0.9923 - acc: 0.6286
42/256 [==>...........................] - ETA: 2:39 - loss: 0.9926 - acc: 0.6282
43/256 [===>..........................] - ETA: 2:39 - loss: 0.9925 - acc: 0.6285
44/256 [===>..........................] - ETA: 2:40 - loss: 0.9910 - acc: 0.6294
45/256 [===>..........................] - ETA: 2:41 - loss: 0.9909 - acc: 0.6299
46/256 [===>..........................] - ETA: 2:40 - loss: 0.9927 - acc: 0.6291
47/256 [===>..........................] - ETA: 2:39 - loss: 0.9903 - acc: 0.6301
48/256 [===>..........................] - ETA: 2:38 - loss: 0.9889 - acc: 0.6308
49/256 [===>..........................] - ETA: 2:36 - loss: 0.9895 - acc: 0.6307
```

Fig. epochs in progress

```
Train loss: 0.1440010733440248
Train accuracy: 94.93538611605979
Test loss: 2.941766869360838
Test accuracy: 56.42240178364172
```

Fig. Results and Accuracy for the model

```
Time taken to build model: 0.01 seconds

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances          77               77      %
Incorrectly Classified Instances        23               23      %
Kappa statistic                          0.4983
Mean absolute error                      0.2641
Root mean squared error                  0.4414
```

```
=== Detailed Accuracy By Class ===

               TP Rate   FP Rate   Precision   Recall  F-Measure  MCC      ROC Area  PRC Area  Class
               0.803     0.294     0.841       0.803   0.822      0.499    0.753     0.788     c0
               0.706     0.197     0.649       0.706   0.676      0.499    0.753     0.615     c1
Weighted Avg.  0.770     0.261     0.776       0.770   0.772      0.499    0.753     0.729

=== Confusion Matrix ===

  a  b   <-- classified as
 53 13 |  a = c0
 10 24 |  b = c1
```

Fig. J48 accuracy, error and confusion matrix

Followings are the snippets of the second module i.e. OpenCV Module that is used for Human Facial Expression Recognition.

Step1: The face is detected in the real time through the webcam using HaarCascade Classifier.

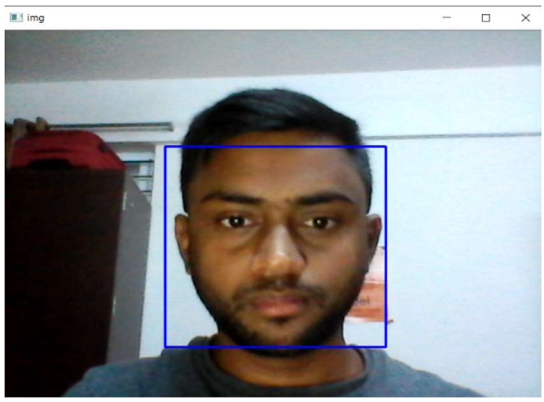


Fig. Face Detection

Step2: The captured face is then converted into a grayscale in 48x48 pixels frame.

Fig. Conversion into Grayscale

Step3: The model then predicts the expressions. Model is saved in a json file and then it will be imported the OpenCV module to recognize the real time images in the videos taken from the user.

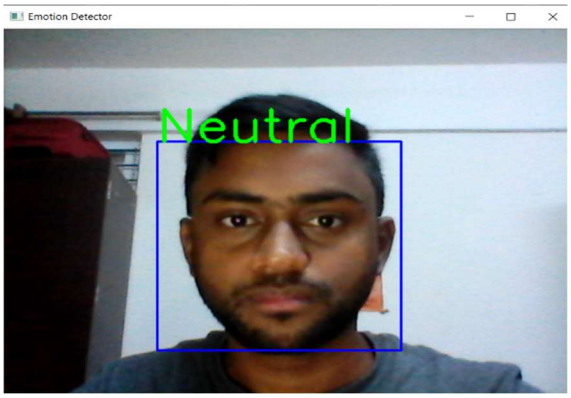
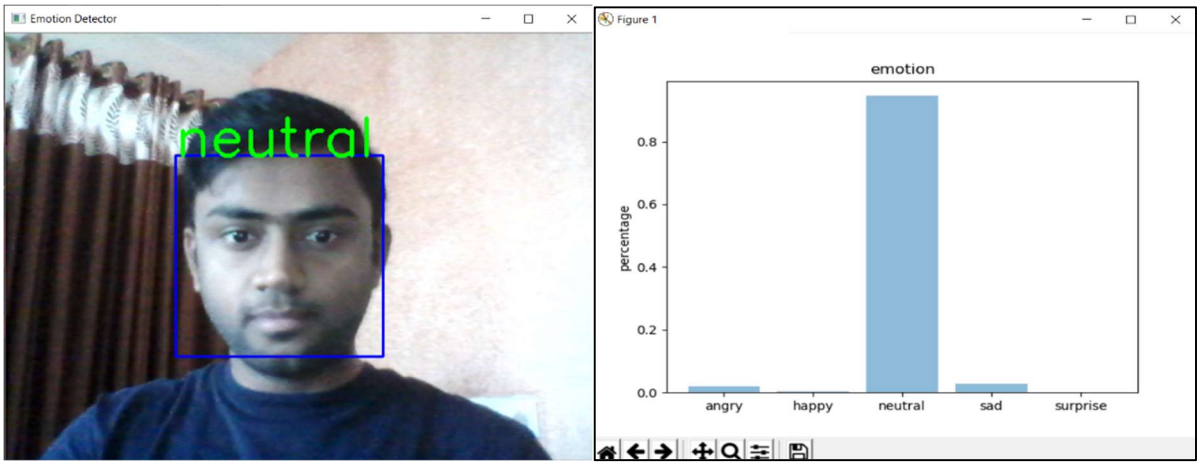


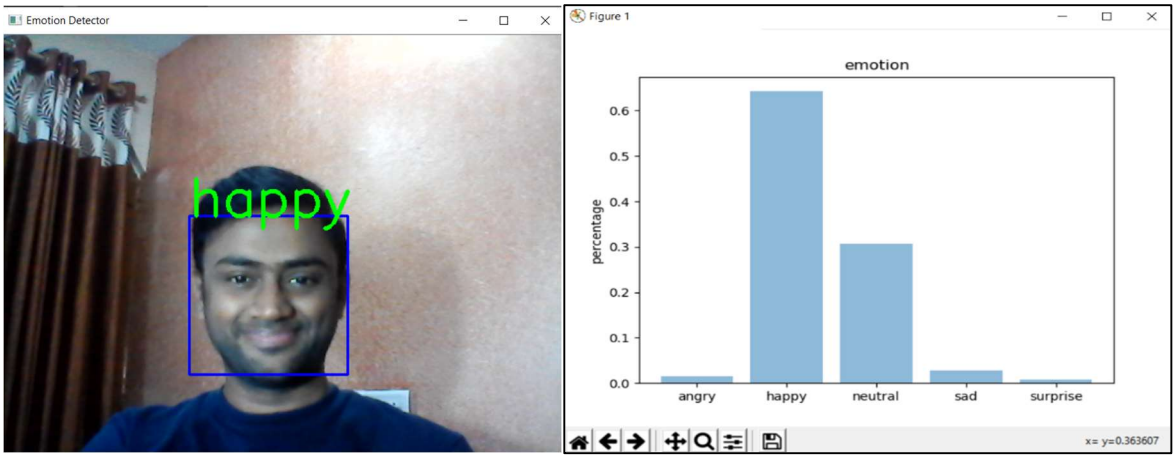
Fig. Facial Expression Recognition

Results:

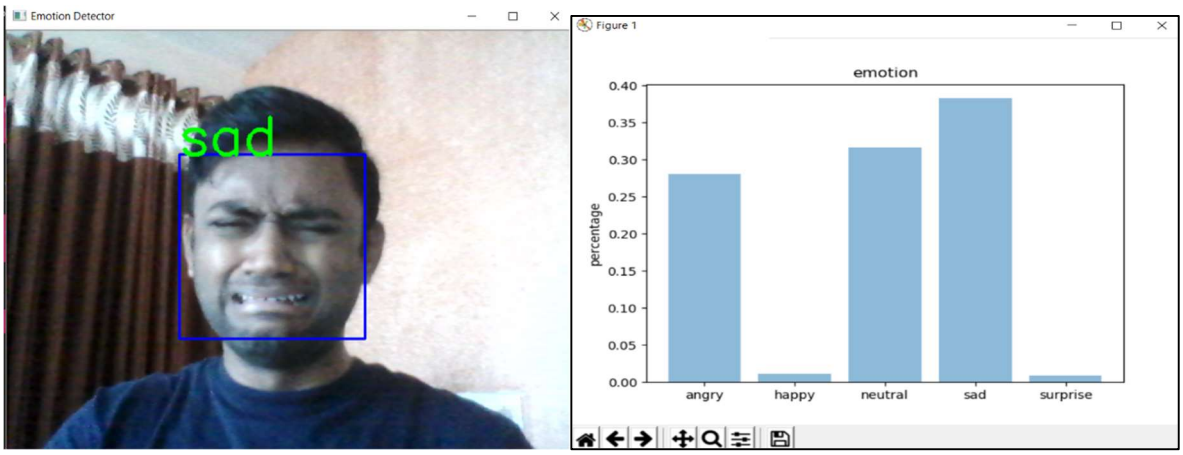
i. Neutral



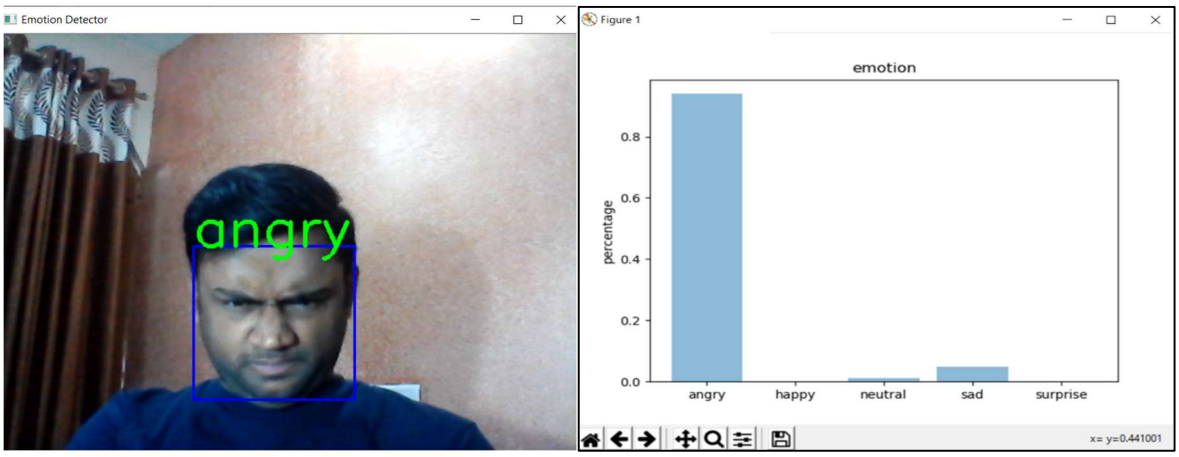
ii. Happy



iii. Sad



iv. Angry



v. Surprised

VI. Results and Discussions:

The two proposed model which has been discussed and one of which is implemented has seen differences in accuracies and errors. The first model is the simple implementation of CNN on FER dataset and includes the removing of the two classes with their labels i.e. Fear and Disgust since they consist of least number of tuples. The codes of the first proposed model are written in Python3. Later, the model has been used for generating outputs for input images to find the emotion of the person. This model has used 75% of the tuples for training and 25% for testing. The following details shows how good the first model performed:

There was an accuracy of 49.3% after half of the epochs were gone through pooling and filtration through several layers. And to the end it was around 64%. The overall conclusion drawn from the first model was that the train accuracy ended up with 94.93% and test accuracy was 56.42%. The second model is not an implementation was done on Weka on the model that is already generated through the Convolutional Neural Network. The algorithm which has been used for generating decision trees for random forest is J48 which is C4.5 algorithm. The number of tuples that were put for cross-validation were 10 and the percentage of the splits between train test cases were 75% and 25% like earlier model. Following conclusions have been drawn from the overall model.

VII. Conclusion:

This work can be further studied and researched to find more accurate models using different algorithms and image processing techniques. With people coming more in this research field, there are chances that a fully automated facial expression recognition system can be brought to the markets with 100 % of accuracy. Those models will be able to help researchers to build an efficient Artificial Intelligence. One can't think of a humanoid without the ability of knowing what a person feels in order to help or give service to him/her. With the ability to feed the

pictures automatically in dataset after converting it into grayscale, it will increase the chances of making a whole new dataset and generate models. It can also feed into any microcontroller to make it a live project or an IOT project. The best microcontroller can be Raspberry pi since it works as an operating system which will reduces the efforts for writing code for microcontroller, one just need to dump these codes in it.

This project code is neat and maintained well with latest versions and each dependency being able to meet the environment variables. Thus, there won't be any problem regarding the code usability and maintenance even if the versions are updated. In this era of cutting-edge technologies, one can never imagine it without automation. The very first thing someone imagines: how any computer visualizes things near them apart from logical reasoning, this answer that imagination. The various applications of this project can be the security threats posses by normal public, receiving feedback from customer at hotels, restaurants, and other profitable businesses. It can also be used in improving the algorithms. This will also allow the researchers to find some other alternatives to these for generating models since this code is dataset independent. It only needs images as input that will be converted to the grayscale.

VIII. References:

[1] Marek Kowalski, Jacek Naruniec, Tomasz Trzcinski, Deep Alignment Network: A convolutional neural network for robust face alignment"

[2] Ali Mollahosseini, Behzad Hasani, and Mohammad H Mahoor. Affectnet: A database for facial expression, valence, and arousal computing in the wild. Transactions on Affective Computing, 2017.

[3] Kai Wang, Xiaojiang Peng, Jianfei Yang, Debin Meng, "Region Attention Networks for Pose and Occlusion Robust Facial Expression Recognition"

[4] Ivona Tautke, Tomasz Trzcinski, Adam Bielski, "I Know how You Feel: Emotion Recognition with Facial Landmarks"

[5] B. Hasani and M. H. Mahoor, "Facial expression recognition using enhanced deep 3d convolutional neural networks," in Proceedings of CVPRW. IEEE, 2017.

[6] Daniel Llatas Spiers, "Facial Emotion Detection Using Deep Learning"

[7] Sivo Prasad Raju, Saumya A and Dr. Romi Murthy, "Facial Expression Detection using Different CNN Architecture Hybrid Vehicle Driving", Centre for Communications, International Institute of Information Technology.

[8] Deepesh Lekhak, "Facial Expression Recognition System using Convolutional Neural Network", Tribhuwan University Institute of Engineering.

[9] Jie Hu, Li Shen, and Gang Sun "Squeeze-and-excitation networks", in Proceedings of the IEEE conference on computer vision and pattern recognition, pages 7132–7141, 2018.

[10] Shan Li and Weilong Deng, "Deep Facial Expression Recognition: A survey", arXiv:1804.08348v2 [cs.CV] 22 Oct 2018.