Emojify-Sentiment Analysis Using Webcam

Darsh Vaghasia, Varun Shah Computer Science & Engineering, DEPSTAR, CHARUSAT, Changa, Nadiad (Gujarat), India

Abstract - Emotion has been a great role in one's life. Emotions are the way to understand people out there so that they can get immediate help need. Besides emotions are the way through we express our feelings to the world outside. Today Mental Health is one of the most important factor along with the physical health. Mental Health refers to cognitive, behavioral and emotional well-being. It can affect daily lives, relationships, and physical heath. If the Mental Health is not given proper attention in time, then it can cause tragedy and make one's life a living hell. Depression, Stress, Anxiety, Mental Lockdown all these terms are not just terms but it needs to be given attention and with a try towards early diagnosis until it's too late to take a necessary step which must have taken earlier. With this we have tried to implement using the current trend which is booming with great inclination i.e. Artificial Neural Intelligence or Deep Learning and Computer Vision.

So, basically Deep Learning is the subfield of machine learning which is again a subset of Artificial Intelligence. It is inspired by the Structure and Function of Human Brain called Artificial Neural Networks. Spreading light on Emojis and Avatars. They are the ways to indicate non-verbal cues which is conspicuous and understandable by people easily. These Cues plays one of the most vital role in online chatting, Social Media, product review and many more. It has certainly lead to boom data science research towards emoji-driven storytelling. Today it is possible to detect emotion or analyze emotion with

Human Emotions. So with this we put forward the definition of the research project i.e. Classification of Human Facial Expression and Detection of particular emotion and map corresponding emoji's or avatars. The main proposed idea of this paper is to use the Neural network approach towards Emoji Detection and map to parallel Emoji Screen. Convolution Neural Network (CNN) is the center of all deep learning strategies.

In this paper for performance evaluation of CNN we have used FER2013 dataset (facial expression recognition) which consists of 48*48-pixel grayscale face images. An achieved accuracy of 95.45% accuracy for detection of emotion. In here we have trained 80 % of dataset and left 10% of dataset for test purpose. We have Used Deep Neural Network. The motive of this paper is enhance emotion detection and a miniscule step towards treating the mental health disorder. The idea of this project is implemented in a safari of startup definition for us. Here little try from us and we have integrated it with web application and working on enhancing it and even looking forward to develop an application.

Key words: Emoji Detection, Convolution Neural Networks (CNN), FER2013 dataset, TensorFlow, Keras, OpenCV, Deep Learning.

INTRODUCTION

Detection of human detection is generally implemented in places that require many details of a human being or in the areas that require more security. The first task in emotion detection is face detection, which can be considered as first layer for the security, in a two-stepped security model. The second stage can be considered as emotion detection in a human being. The emotion detection system helps us in identifying that the person standing in front of camera is a genuine live person instead of a photographic image.

The other important field in today's world where this system can be useful is the marketing of products along with promotions of business. This can be combined with artificial intelligence where the system will be able to classify if the user likes the product or not based on the emotions that are getting reflected by the user's face.

We already know that, today in the field of security, many techniques such as voice detection, fingerprint matching, etc. are being used. We are aware that nowadays accident count is increasing rapidly so this system can be used to detect the mindset of the driver. The drowsiness factor can give us an idea if the driver is feeling sleepy or not. So this may lead in reduction of road accidents happening every day.

Even in today's world identifying human emotions is a difficult task as a slight change in face muscles gives totally a different set of emotion. Similarly, it also

depends on person to person how their face gives reactions, which makes it more challenging to predict the emotion correctly and accurately on every instance. But nowadays machine learning and deep learning are proving to be very successful in accurately predicting the emotions.

The basic steps that can be considered for face detection are:

I. Preprocessing the data
II. Detection of face
III. Selecting the features
IV. Classification on the basis of trained model

In this article, we will be presenting the techniques and models that can be used to train and also the difference between different models' accuracy levels. Also, we will be having a brief idea of different tools and technologies that have been used to develop this project. The results and conclusions will also be discussed along with the future scope of this project and the enhancements that can be brought in this project.

LITERATURE REVIEW

To make such a system successfully work, we need to make sure that the system that we design functions very similar to how our brain does. Our brain recognizes the emotions perfectly based on seeing the opposite persons' face. Similarly, if we want to predict the emotion precisely and correctly we need to use and incorporate the neural networks for a higher accuracy. In today's world, neural network is playing an important role in machine learning and deep learning. CNN is used for the better accuracy that can be achieved using the system. It acts very similar to how a human/animal brain functions. Therefore, to achieve accurate emotion detection it is the best possible approach in today's world. When we look at the other systems that are being used, they are not focusing on face detection and instead they tend to use the facial features of a person such as eyes, mouth, etc. Here we have tried to work it out differently by using haar-cascade classifier for detection of the face and try to predict the emotion based on the trained model using CNN. Rule based systems and KNN(K-Nearest Neighbor) are also used by many of the systems to predict the emotion of the person. Using KNN, Busso achieved upto 80% of accuracy for a similar system. Similarly, Yacoob presented a system which was accurate upto 88% with a rule based approach for the six basic emotions.

METHODOLOGY AND CLASSIFICATION

We have developed a Deep Learning Model; there we have implemented Deep Convolution Neural Network. The working of CNN is inspired by the working of the Human Brain. Similarly there hundreds and thousands hidden layers which help improvising the feeding and forwarding of the data which certainly improves the accuracy. The main advantage of CNN is that it is capable of detecting the relevant features without any human supervision.

FER2013 dataset consists of 48*48 pixels' grayscale face images. All the images are properly centered and they occupy equal amount of space. This dataset consists following facial emotion categories:

1> 0: Angry
2> 1: disgust
3> 2: feat
4> 3: happy
5> 4: sad
6> 5: surprise

7> 6: natural

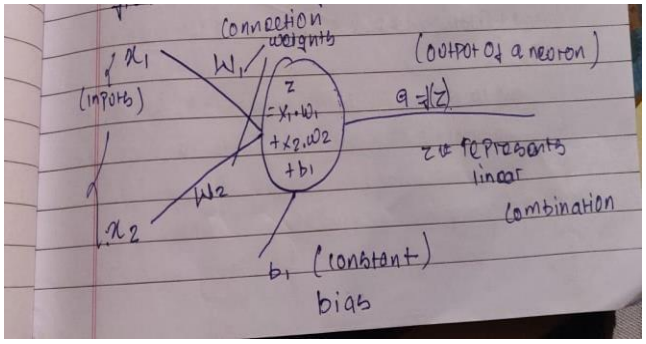
Here, The Training Set Images and Testing Set Images are already into classified folder which consists of 28,709 examples and 3,589 examples respectively.

THE ARCHITECTURE AND THE PREPROCESSING

CNN is formed of Various Layers but mainly divided into three types of layers i.e. Input Layer, Hidden Layer & Output Layer. There is only one Input & Output Layers and there are Hundreds and Thousands of Hidden Layers. Input and Output Layer have only one node and Hidden Layers can have Hundreds of Nodes all nodes have certain weight and their Connecting link i.e. what we call in brain is axon have certain bias. The Input Parameters can be transferred through various methods i.e.

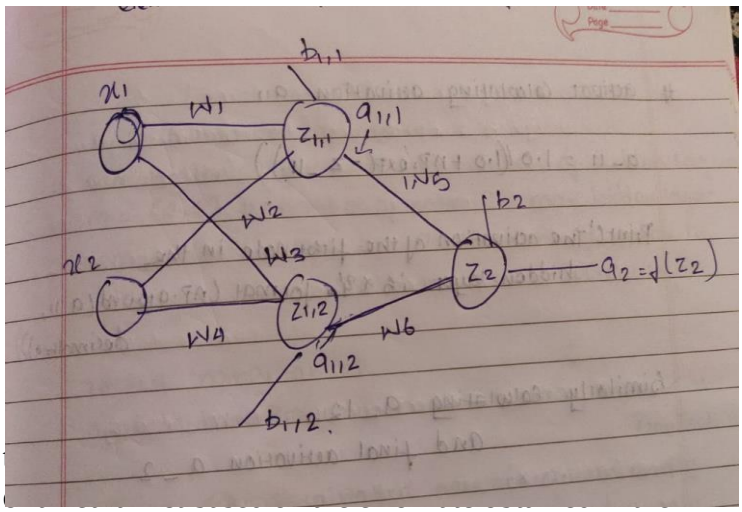
1) Forward Propagation

It is the process through which data passes through layers of neurons in a neural network from the input layer all the way to output layer.



2) Back Propagation

are the three major types of layers used to build architecture. Input - Convolution - ReLU - Pooling - Fully Connected make up the architecture.



previous epoch. It helps to calculate the gradient of a loss function in proportion of all the weights in the network.

The Dataset has total 32,298 images of all sorts of emotions which are already classifies into two sets i.e. Train Test and Test set. The Images are already Centered and occupy equal amount of space. But after loading the data we normalize the data in this case since each image is made up of pixels and the maximum possible value of pixel is 255 so we divide the value of each value by 255 so that we get the values of all the images in range of 0 to 1. All the images are of 48*48 pixels. For the Training and Testing Part we divide the dataset in a batch of 64 Images with a class mode = ”categorical” Briefed about the implementation of the CNN model in the later part of the paper.

CNN LAYERS: OVERVIEW

A CNN is made up of many layers. Convolutional, Pooling, and Fullyconnected layers

Input layer:

The unprocessed pixel estimations of picture of thickness 32, altitude 32, and three shade channels of Red, Green, and Blue will be stored in the 32 x 32 x3 format.

Convolutional layer:

This will work out the production of neurons related through the input limited region, all computing a dot multiplication of their heaviness and an enter volume. This could result in a volume of 32 * 32 * 12 for 12 filters, for example. It creates answer maps by combining input frames with linear sliding filters.

$X_i = \sum_{=0} \square_{,}\square_{,} + \square$

Xi denotes the input function map, wi = [Wi1, Wi2,. ...., Wik] denotes the filter, xi denotes the enter filter, and bi denotes the bias.

ReLU layer:

Rectified Linear Units (ReLU) is an contraction for Rectified Linear Units. The non saturating activation mechanism is used in this sheet. f(x) = max is the ReLU function (0; x). The volume's size remains 32 x 32 x 12 after this layer is applied.

Pooling layer :

Pooling uses a down sampling process, as well as width and height, to produce a volume of 16 x 16 x 12. The outputs of neuron clusters at one layer are combined into a single neuron in the next layer of the pooling layer. The maximum value from each cluster of neurons at the previous layer is used in peak pooling. The average value from each cluster of neurons at the previous layer is used in average pooling.

Fully-Connected layer:

The class scores will be computed by the completely connected sheet. Completely linked layers in the neural network perform high-level reasoning. To change answer maps as close up to the land fact as possible, Layer needs a fixed number of inputs and outputs.

OUR FLOW OF THE SYSTEM

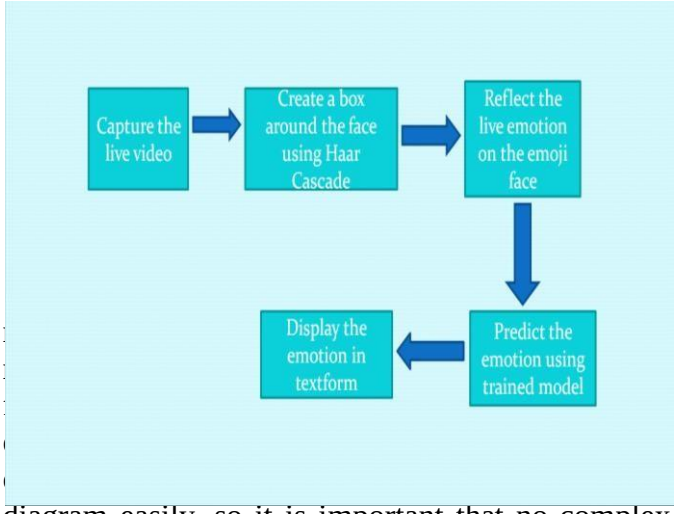
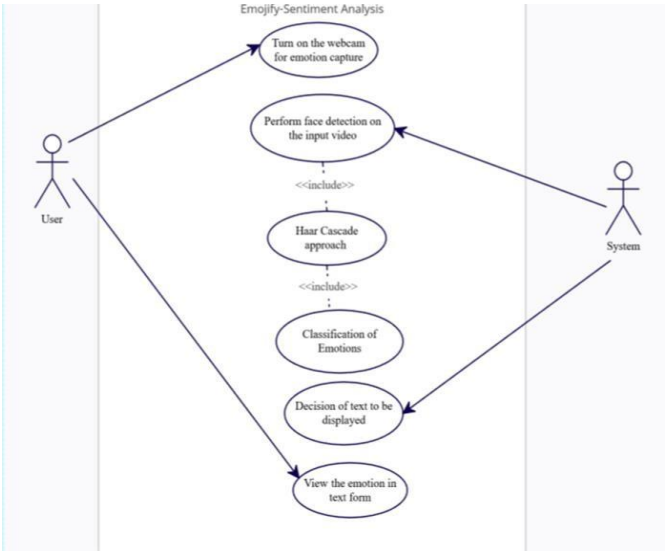


diagram easily, so it is important that no complex things are added in it. In case of a huge system, it is preferable to depict your system in terms of a proper flowchart.

Our system basically focuses on 5 main steps. The first step is to get the input live feed of the user in form of a video through the users' webcam. Once the feed is received, the system will try to form a bounded box around the face using Haar-Cascade classifier, in order to find the face correctly. Now, the system will try to reflect the emotion change that is being shown by the user face on the dummy emoji-face that we have created simultaneously. As the change will come into effect immediately, the

system will try to classify the emotion that is being shown on the face. On completing this procedure, the resulting output will be displayed on the screen in the form of text. In this way, the system will be able to identify which emotion the person is exhibiting at every moment through its live feed.

USE-CASE DIAGRAM OF SYSTEM



Use-case diagram basically is one of the requirement gathering techniques that is required before start of any basic system/project. Use-case diagram helps us to feature the functionalities that will be available to the user when the system becomes ready and also provides us with an idea of functionalities that will be provided to the user by the system. The another feature of the use-case diagram it gives us idea about the relationship that is between different features of our system. Generally, if a big system is being created then there are multiple use-case diagrams that are need to be prepared to depict the whole system. Each use-case diagram will represent a sub-system of the whole system.

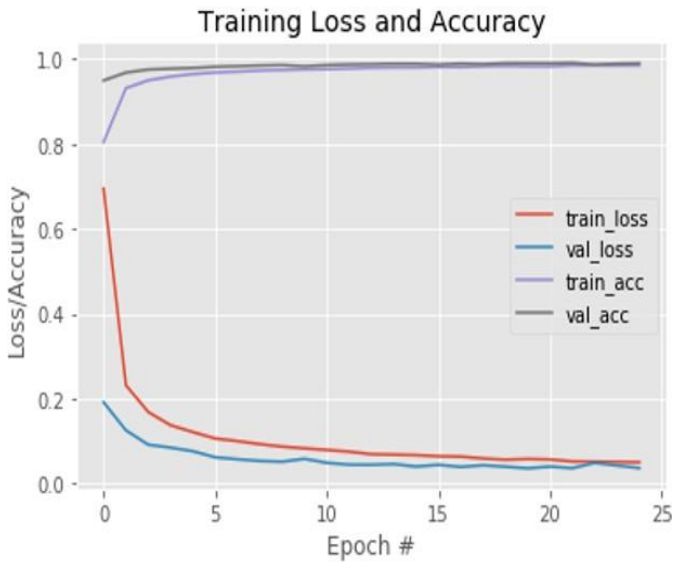
In our project, the user gets to use two features. First is to provide the live input video to the system and another one is to get the predicted emotion by the system in text form. The system will perform face detection based on the live input that has been received through the user. The system will be using the Haar-Cascade approach for face detection which has been described using the include relationship in the diagram. Further, post detecting the face, it will try to classify the emotion in a particular category using the trained model, which is again shown using the include relationship. After classification of emotion, it will generate a text-form of that emotion and print it to the user, which will be final predicted emotion, visible to the user's screen.

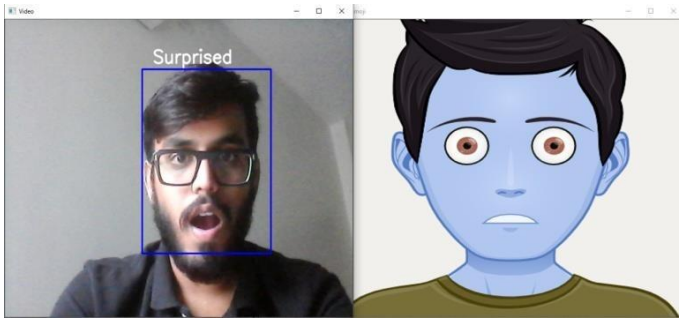
RESULTS AND DISCUSSIONS

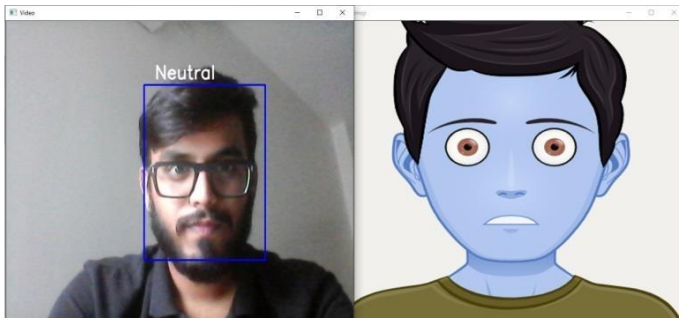
As our model is ready and compiled, we have to fit the data in it. As mentioned previously, we have used FER2013 dataset to train our model. But before feeding data to the model we have to perform some preprocessing task on the dataset for faster learning and accurate results. And also, we have to split the dataset into training and testing sets.

First, we categorized the dataset into 7 categorical classes for 7 digits. After that we normalized the images in datasets by dividing 255 (because max. value of any pixel will be 255 in an image). We have split the dataset into 9:1 ratio (90% for training and 10% for testing).

After training the model, training accuracy and training loss is calculated:





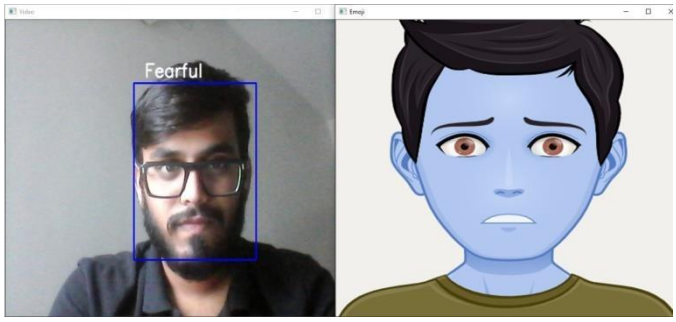


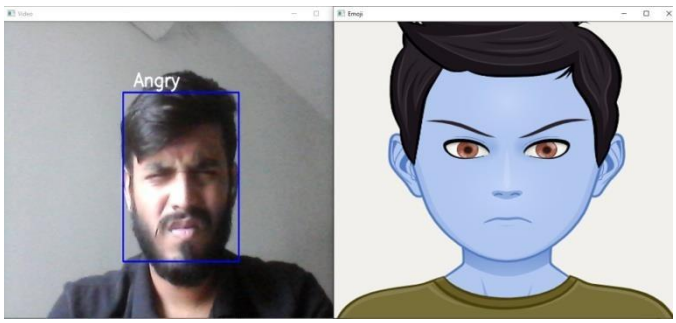
Classification report:

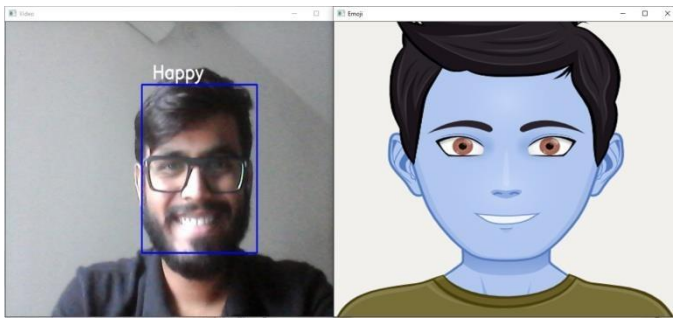
The Figure 12 shows the classification report of the model.

	precision	recall	f1-score	support
0	0.99	1.00	0.99	980
1	1.00	0.99	1.00	1135
2	0.99	0.99	0.99	1032
3	0.99	0.99	0.99	1010
4	0.98	1.00	0.99	982
5	0.98	0.99	0.99	892
6	1.00	0.97	0.99	958
7	0.98	0.99	0.99	1028
8	0.98	0.99	0.99	974
9	0.99	0.98	0.98	1009
accuracy			0.99	10000
macro avg	0.99	0.99	0.99	10000
weighted avg	0.99	0.99	0.99	10000

Classification report of model







FINAL OUTPUT

CONCLUSION

This project has helped us to learn many new concepts related to machine learning and deep learning along with new tools such as flask. It has also been helpful to us in understanding about team work and how to coordinate well as a team. We would like to enhance this project as much as we can even in the future and achieve our future goals. Making progress in this field as much as we can is our next target so that we can help to solve the problems that are being faced by the society. This is the small step that we have put forward in this huge world of machine learning and we would surely like to carry it forward with all our efforts in future as well.

REFERENCES

[1] W. Swinkels, L. Claesen, F. Xiao and H. Shen, "SVM point-based real-time emotion detection," 2017 IEEE Conference on Dependable and Secure Computing, Taipei, 2017.

[2] Neerja and E. Walia, "Face Recognition Using Improved Fast PCA Algorithm," 2008 Congress on Image and Signal Processing, Sanya, Hainan, 2008

[3] Open CV Python Tutorials. [Online] Available: http://opencv-python-tutroals.readthedocs.io/en/latest/py_tutorials/py_feature2d/py_table_of_contents_feature2d/py_table_of_contents_feature2d.html

[4] R. Raja, ―Face detection using OpenCV and Python: A beginner‘s guide,‖ 2017.[Online] Available: https://www.superdatascience.com/opencv-face-detection/

[5] J. M. Saragih, S. Lucey and J. F. Cohn, "Real-time avatar animation from a single image," Face and Gesture 2011, Santa Barbara, CA, USA, 2011

[6] Raut, Nitisha, "Facial Emotion Recognition Using Machine Learning" (2018). Master's Projects. 632. DOI: https://doi.org/10.31979/etd.w5fs-s8wd https://scholarworks.sjsu.edu/etd_projects/632

[7] Burges, C. A tutorial on support vector machines for pattern recognition. Dat Mining and Know. Disc., vol. 2(2), pp. 1– 47, 1998.

[8] C. Darwin and P. Barrett, A concordance to Darwin's The expression of the emotions in man and animals. Ithaca: Cornell University Press, 1986.

[9] M. Suwa, N. Sugie, and K. Fujimora, “A preliminary note on pattern recognition of human emotional expression,” in Proceedings of the 4th International Joint Conference on Pattern Recognition, pp. 408–410, Kyoto, Japan, November 1978.

[10] P. Ekman and W. V. Friesen, Facial Action Coding System, Consulting Psychologists Press, Stanford University, Palo Alto, Calif, USA, 1977.

[11] J. C. Hagerand, P. Ekman, and W. V. Friesen, Facial Action Coding System, A Human Face, Salt Lake City, Utah, USA, 2002

[12] Davis E. King. Dlib-ml: “A Machine Learning Toolkit.” Journal of Machine Learning Research 10, pp. 1755-1758, 2009

[13] T. Kundu and C. Saravanan, "Advancements and recent trends in emotion recognition using facial image analysis and machine learning models," 2017 International Conference on Electrical, Electronics, Communication, Computer, and Optimization Techniques (ICEECCOT), Mysuru, 2017, pp. 1-6.