Emotion Recognition using Convolutional Neural Network

Bharath Roshan Bhombore
Computer and Information Science
University of Strathclyde
Glasgow, United Kingdom
bharath.bhombore.2021@uni.strath.ac.uk

***Abstract* – Facial expression of human beings are raw, significant data that be made useful by efficient extraction. Emotion recognition has become an important field of research with the increased use of IOT devices. With proper implementation and training AI can detect the emotions of a person through their facial expressions. Reading facial expression for emotion detection has always been an easy task for humans, achieving the same task with a computer algorithm generates powerful data which is why this field has attracted interest of many problem solvers in the field of artificial intelligence. With the increase of camera devices, we are generating huge video amount of data every day. This data can be made useful if we can extract the important information from it. This information is equally valuable for industry companies and researchers. This paper consists of CNN and Computer Vision approach to detect the emotion of a person automatically in real time from videos. The dataset used in this paper is FER – 2013 which consists of 28,709 training images with 7 different classes and 7,178 images used for testing. We will be also using the Haarcascade Frontal Face to Identify the face in Real Time using OpenCV.**

Keywords – CNN, Open CV, Emotion Recognition

I. INTRODUCTION

Emotion Recognition is the process of detecting human emotions from facial expression. Our human brain does it in fraction of second automatically. With the use of Artificial Intelligence, we can train machines to do the same. Our facial expression gives of much important information that can be useful in many ways for different purpose.

With IOT, smart environment like smart home etc. the demand of real time facial recognition is increasing. We can make use of powerful deep learning algorithms such as, CNN, RNN, LSTM, Transfer learning etc. to train the machines to achieve very high accuracy for emotion recognition. Machines as such can generate extremely valuable information that can be used in variety of fields which is why Emotion recognition is demanding field of Artificial Intelligence. Billions of dollars are being invested each year to build machines that can outperform its predecessors. As this is area of AI is quite resource heavy to achieve high accuracy.

The Research in this field has been going on from a long time when we look at the paper from Paul Ekman from 1971, We can see that they were able to describe 6 classes of image i.e., Happy, Anger, Sad, Disgust, Surprised. Even today researchers are working on trying to identify different emotions in this area.

The major challenge that one gets in this field is the correctness of the data which is collected. Many of the datasets have different Human poses and the photos need to be taken in a stable environment to help train the model.

As this field is growing day by day, there are various work which have been taking place in the field of facial emotion recognition. There are different approaches that have been used here like Geometric and appearance.

1. Geometric - Geometric changes in the face like the mouth, nose, eyebrows in the form of facial expression with transient features and the different formation in the expression.
2. Appearance – Transient Changes in the appearance of the person’s face.

II. METHODOLOGY

A. DATASET

The training for a neural network requires a large amount of data. The choice of the dataset can be different for different approach like it can be Image dataset, text data, speech data. For Emotion Recognition there are large number of available Datasets which have different resolution of images, some can be of high resolution, and some can be of medium or lower resolution images. Training High Resolution Images and a large dataset requires us to have good computation power and GPU power is required.

In this paper we will be using FER – 2013 Open Database for Facial Expression from Kaggle.

The dataset contains 48x48 pixel grayscale images of human face. For the simplicity of training the faces have been automatically registered to make sure that the face is centered and takes about the same amount of space in every image. The Data in the dataset are categorized into 7 different classes i.e., Happy, Angry, Sad, Neutral, Surprised, Disgusted Task is to train the model to categorize each image according to the emotion displayed on the face in each image.

B. NETWORK

The architecture used in this include 4 Convolutional Layer and 2 Fully Connected Layer has shown in Fig [1]. The image which is being given has the input to the model is resized into 32x32 using the TensorFlow ImageDataGenerator. The Input Layer and the convolutional layers and 1 fully connected layer uses the “ReLU” activation function, and the output layer uses the “Softmax” activation function. To the input layer we are feeding of an image shape of 32 x 32 x 3 and there are 32 neurons present in this layer. The pre-processing done in the Convolutional layer is very less when compared to other classification algorithms.

The output of first layer is then fed has the input to the second convolutional layer which has a kernel size of 3,3 with the activation has “ReLU” and has 64 neurons.

We are then downsampling the feature maps obtained from the output of the previous layer. The MaxPooling function takes in the pool size and downsamples the output. The Dropout of 0.5 has been added to avoid overfitting. The next 2 convolutional layer each consists of 128 neurons and uses the “ReLU” activation function. The output of this layer fed onto the fully connected dense layer which contains 1024 neurons, and here we are using L1 Regularization.

The output of this layer is given has input to the output layer which contains 7 neurons, this is because we have 7 different class in our Dataset and the activation used here is “SoftMax”.

The loss function we are using in this paper is Categorical Cross Entropy and the optimizer we are using is Adam with a learning rate of 0.0001.

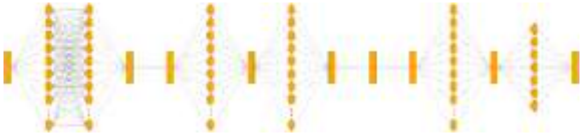

Fig [1] - Network

C. ReLU

The Rectified Linear Unit is mainly used activation function at the Internal Layer a network and it provides faster convergence when compared to that of Sigmoid and Tanh functions. This function also facilitates to faster learning.

$$f(x) = max(0, x)$$

D. SOFTMAX

The Softmax is mainly used for Multi Class Classification and is mostly used in the output layer.

$$f(x_i) = \frac{e^{x_i}}{\sum_i e^{x_i}}$$

E. LOSS FUNCTION

The Categorical Cross Entropy is well suited to work on Classification Problems, and it is used for multi - class classification. It is normally used to quantify the difference between probability.

$$Loss = -\sum_{i=1}^{size} y_i \log \hat{y}_i$$

F. RECOGNITION

The Recognition part is using OpenCV, where we are passing on the saved model and using the HaarCascade Frontal Face to recognize the face in real time. And the Web Camera Detects the Different Emotion of the person in front of the camera as shown diagrammatically in Fig [2].

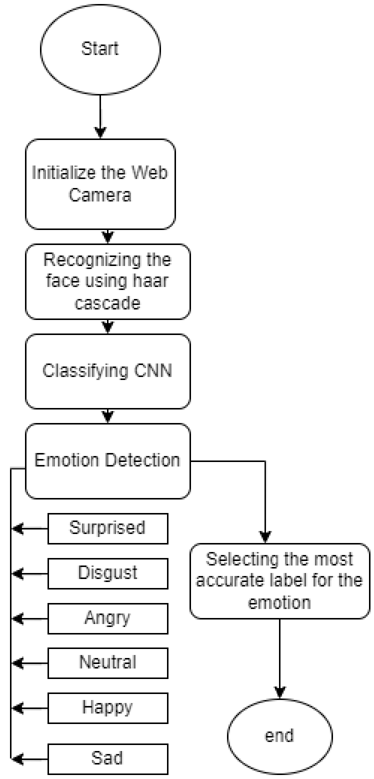

Fig [2] – Illustration of Recognition using OpenCV

III. RESULTS

This Network was trained with a learning rate of 0.0001 and with 300 epochs and the accuracy obtained after the completion of the 300th epoch was 69.02%. But the best accuracy was observed to be at the 298th epoch which was 69.80%. The dataset was trained with 28709 images in the training set and with an image size of 32 x 32. The Fig [3] shows us the graphical trends in Accuracy vs Epochs.

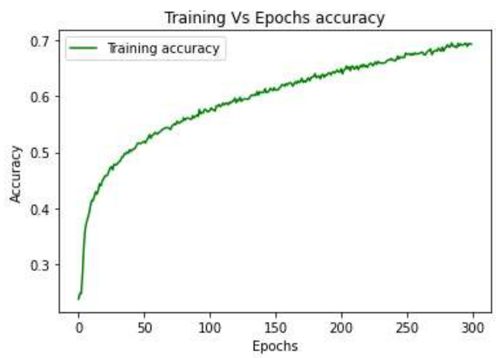

Fig [3] – Training Accuracy vs Epochs

At the 300th epoch we got a loss of 0.9232 and loss was minimum at the 298th epoch with a loss of 0.9071. The Fig [4] shows us the graphical trends in Loss vs Epochs.

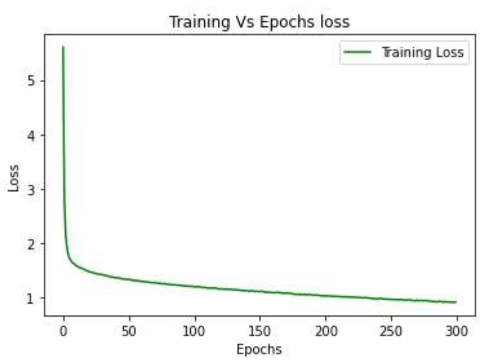

Fig [4] – Training Loss vs Epochs

This model could have performed better when we could have increased the number of epochs. Due to limited Computation Power and GPU the maximum number of epochs used to train was 300.

IV. COMPARATIVE STUDY

This model could have had a better accuracy when the optimizer used was Stochastic Gradient Descent. But when compared to the reference [3], my model achieved an accuracy of 69.80%. Whereas the accuracy they had achieved was less compared to this. They had a learning rate of 0.001 and the regularization was 1e-6 and the number of hidden neurons was 512.

They had batch size of 132 and the shallow model gave an accuracy of 55% on the validation set and 54% accuracy on the test set. I achieved the accuracy of 69.80 and the validation accuracy of 59.40.

V. DISCUSSION AND FUTURE WORK

In this paper, I had used 300 epochs to train this model from scratch . In the future, I plan to work on this using the Transfer Learning approach using different pre-trained model available such as VGG-19, ResNet, Inception V3. And improve the accuracy of this model. There are lot of research going on this area of study and lot more can be implemented in the Future with Speech Emotion Recognition etc. And this area is expanding good investment in more advanced computation power to train the model.

VI. REFERENCES

[1] I. Lasri, A. R. Solh and M. E. Belkacemi, "Facial Emotion Recognition of Students using Convolutional Neural Network," 2019 Third International Conference on Intelligent Computing in Data Sciences (ICDS), 2019, pp. 1-6, doi: 10.1109/ICDS47004.2019.8942386.

[2] R. GUETARI, A. CHETOUANI, H. TABIA and N. KHLIFA, "Real time emotion recognition in video stream, using B-CNN and F-CNN," 2020 5th International Conference on Advanced Technologies for Signal and Image Processing (ATSIP), 2020, pp. 1-6, doi: 10.1109/ATSIP49331.2020.9231902.

[3] Convolutional Neural Networks for Facial Expression Recognition Shima Alizade and Azar Fazel

[4] https://www.kaggle.com/msambare/fer2013

[5] Emotion Recognition Using Convolutional Neural Network (CNN) Nur Alia Syahirah Badrulhisham and Nur Nabilah Abu Mangshor 2021 J. Phys.: Conf. Ser. 1962 012040

[6] Facial Expression Classification Using Deep Convolutional Neural Network J Electr Eng Technol.2018; 13(1): 485-492 http://doi.org/10.5370/JEET.2018.13.1.485