CSE 676LEC: Deep Learning

Team Members
1) Saikrishna Dirisala - sdirisal@buffalo.edu - 50411928
2) Suradhya Gadde - suradhya@buffalo.edu - 50442661

1 Project Report

Title : FACE RECOGNITION

1.1 Abstract

Face recognition is considered as one of the most crucial aspects of security. The process of face recognition is done through identifying the movement of muscles within the face. Our project model tries to analyze the facial features of the people and determines the emotional expressions from the recognized faces. To implement this project successfully, we used the CNN algorithms to train, and we have gathered the image data from ten people in real time and tried to train the model to predict the emotions.

1.2 INTRODUCTION

In the modern world, the necessity to maintain security for information and property has become important. Nowadays we hear a lot of credit card fraud, computer break-ins by hackers, and security breaches in a company or government sites. In the year 1998, sophisticated cyber crooks caused well over US 100 million dollar in losses. In March 2020, 538 million of customer details on Sina Weibo have been sold out for 250 dollars to the dark web.

All the data breaches in e-commerce sites or governmental sites are because of the flaws in the conventional access control systems: the systems don't authenticate the person based on the person's physiological characteristics. Users are authenticated or identified based on the identifications provided by government authorities like passport, SSN, and Aadhar when a person must be identified in any government agencies. The Passwords that are being used in social media sites or e-commerce websites are passwords where people are more used to using their date of birth or their own government ids or a few specific dates that are very important in their life. Advancements in technology have led us to use biometrics. In this field of authentication, we use physiological and personal behavioral patterns as a way of authenticating. Authentication using fingerprints or facial features comes under the Physiological characteristics of biometrics and authentication done using the person's writing, signature, or amount of pressure applied on the keyboards is something that can be used as behavioral biometrics. This type of authentication done using a person’s biological characteristics are difficult to forge.

When it comes to biometric authentication methods few stable methods are physiological methods which are fingerprints, faces, and DNA than the other behavioral gestures like the signature, pressure applied on keys of the keyboard, and voice. Physiological characteristics change only when there is a severe injury else, they are very stable. Behavioral patterns like signing, pressing, and voice change based on the mood of the person. Voice will have a low pitch when they are in a good mood and a high pitch when they are in a dull mood. Writing or signing based on the person’s anxiety, and stress. But people are more likely to use the voice, and pins compared with scanning documents.

An authentication system that has both high accuracy and low intrusiveness is Face Recognition without being intrusive. Which has attracted researchers in the field of security, psychology, and image processing to computer vision.

Facial expressions are one of the most flexible and efficient ways to share our feelings and thoughts. As humans, we can easily identify a human expression and determine the person’s emotion. However, it is not as trivial as it sounds to be, as even the slightest change in the expression gives us a signal of a different emotion. Identifying emotions has a wide variety of advantages, for instance, governmental organizations can easily identify the criminal’s emotions like guilt, anxiety, and uncertainty. Another advantage is to identify the drowsiness of the driver and send alert messages accordingly. Spatial information from a single image is used in static images, whereas in sequence-based methods we use images from adjacent frames. There have been significant models developed in recent years on Facial Emotion Recognition and classification, however, designing an artificial intelligence system that satisfies this requirement is difficult.

we will use Convolutional Neural Networks (CNN) for recognizing facial emotions. Our model receives images as input and then using CNN it predicts the facial emotion label, as one of the following: anger, happiness, fear, sorrow, disgust, or neutral. The overall project aim is to implement an automatic Face Emotion Detection model, where the inputs are images of facial emotions and the output is the recognition and classification of a specific input image’s expression into seven different categories, such as happy, sad, fear, anger, disgust, surprise, and neutral. Several research has been conducted in this domain in the past, however, this model outperforms previously proposed models in terms of accuracy.

1.3 RELATED WORK

To keep ourselves abreast of the ongoing research and advancements in the field of facial recognition and carry out our work after we referred through several research publications on the field. Following are some of the papers that are identical to our approach and describes the similarities and differences between those papers and our approaches.

Kewen Yan , Shaohui Huang. Face Recognition Based on Convolution Neural Network; In this paper, the author tries implementing a CNN model of nine layers which includes three convolution layers, two pooling layers, two full-connected layers and one SoftMax regression layer. Implementation of this model has given them a better prediction of 99.82 percentage and 99.78 in recognizing faces, whereas when compared to that of our model it consists of four layers, with output as one layer, and few dropout layers to escalate the model overfitting with a prediction of 97 percentgae.

Shrey Modi; Mohammed Husain Bohara Facial Emotion Recognition using Convolution Neural Network; In this paper, the authors implement the CNN model by using the following layers as in their architectural model for feature extraction-ReLU, Leaky ReLU, LeakyReLU. With these techniques they were successfully able to predict three emotions with prediction rate of 91 percentage. However, our model is better than compared to that of this one, it is because, we were successfully able to predict seven emotions with an accuracy of 97 percentage using four layers.

Arushi Raghuvanshi, Vivek Choksi Facial Expression Recognition with Convolutional Neural Networks; in this paper, the authors try to implement the similar techniques to that of our model, but our model has better prediction than compared to that of their model. The architecture of their model can be improved by implementing better methods than they have picked.

Akriti Jaiswal, A. Krishnama Raju, Suman Deb , Facial Emotion Detection Using Deep Learning; In this paper, the author tries implementing a convolutional neural networks (CNN) based deep learning architecture for emotion detection from images. They used two different types of datasets Facial emotion recognition challenge (FERC-2013) and Japanese female facial emotion (JAFFE) and achieved an accuracy of 70.14 and 98.65 respectively, where as when compared to that of our model it consists of four layers, with output as one layer, and few dropout layers to escalate the model overfitting with a prediction of 97 percentage.

Ma Xiaoxi, Lin Weisi , Facial Emotion Recognition ; In this paper, the author tries implementing Support Vector Machine (SVM) and Deep Boltzmann Machine (DBM) along with the fusion of both for facial emotion recognition with an accuracy of 85.7, 89.3 and 91.0 respectively. whereas when compared to that of our model it consists of four layers, with output as one layer, and few dropout layers to escalate the model overfitting with a prediction of 97 percentage. Zeynab Rzayeva, Emin Alasgarov, Facial Emotion Recognition using Convolutional Neural Networks; In this paper, the author tries implementing a convolutional neural networks (CNN) with two datasets , Cohn-Kanade, RAVDESS and the two combined with an accuracy 88 percentage , 92 percentage and 92 percentage respectively, where as compared to that of our model it consists of four layers, with output as one layer, and few dropout layers to escalate model overfitting with prediction of 97 percentage.

Kahina Amara, Naeem Ramzan, Nouara Achour, Mahmoud Belhocine, Cherif Larbas, Nadia Zenati , Emotion Recognition via Facial Expressions ; In this paper, the author tries implementing The K nearest neighbors (k-NN) with an accuracy of 97.09 , where as when compared to that of our model it consists of four layers, with output as one layer, and few dropout layers to escalate the model overfitting with a prediction of 97 percentage

1.4 CONVOLUTION NEURAL NETWORK

Convolution neural network is the updated version of the Feed Forward Neural Network, which helps the neural network model to know the difference in the images. For example, difference between the DOG and CAT photos. The CNN has convolution layer which helps into reduce the number of the computations by reducing the number of values. The services, of the convolution network, pooling and stride helps in shrinking the number of columns and rows in the image. The mix of convolution and neural network helps to build the model which solve the image classification problem.

In the project we have developed the model which helps to classify the images of our teammates. There are three convolution layers followed by feed forward network of four layers, with output as one layer, and few dropout layers to escalate the model overfitting.

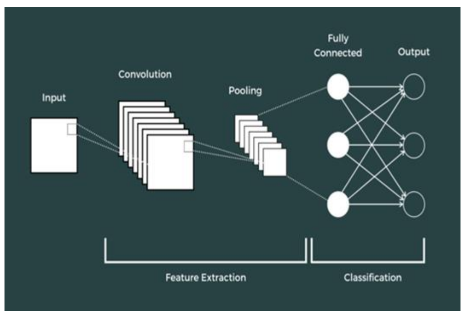


Figure 1: General Architecture of CNN

Layer (type)	Output Shape	Param #
conv2d_375 (Conv2D)	(None, 100, 100, 64)	1792
max_pooling2d_375 (MaxPooling2D)	(None, 50, 50, 64)	0
dropout_750 (Dropout)	(None, 50, 50, 64)	0
conv2d_376 (Conv2D)	(None, 48, 48, 32)	18464
max_pooling2d_376 (MaxPooling2D)	(None, 24, 24, 32)	0
dropout_751 (Dropout)	(None, 24, 24, 32)	0
conv2d_377 (Conv2D)	(None, 22, 22, 64)	18496
max_pooling2d_377 (MaxPooling2D)	(None, 11, 11, 64)	0
dropout_752 (Dropout)	(None, 11, 11, 64)	0
flatten_125 (Flatten)	(None, 7744)	0
dense_500 (Dense)	(None, 32)	247840
dropout_753 (Dropout)	(None, 32)	0
dense_501 (Dense)	(None, 16)	528
dropout_754 (Dropout)	(None, 16)	0
dense_502 (Dense)	(None, 8)	136
dropout_755 (Dropout)	(None, 8)	0
dense_503 (Dense)	(None, 2)	18

Total params: 287,274
Trainable params: 287,274

Figure 2: The Current Architecture of the CNN model.

1.5 MODEL TRANING

1.5.1 INITIAL TRANING

Here the model was trained using the initial data of the size 52 which are 26 of them are one of the types and 26 belongs to another group. The key rule while training the model is data. If we have huge data then the model convers most of the cases, otherwise model won't be able to classify all cases. After training the existing dataset we have found the imperfection in accuracies and losses. As there is increment in epochs, fluctuations were found in the graphs of the accuracy.

.

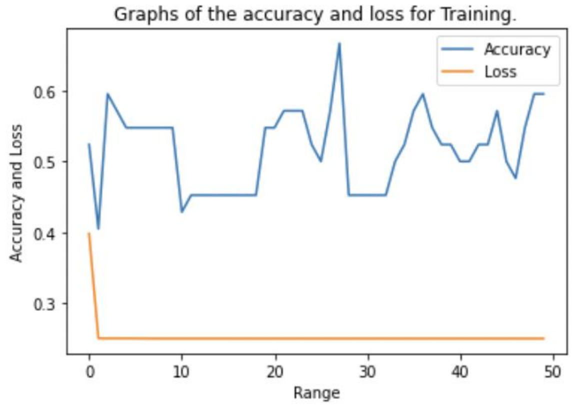


Figure 3: Accuracy and Loss of the initial Data.

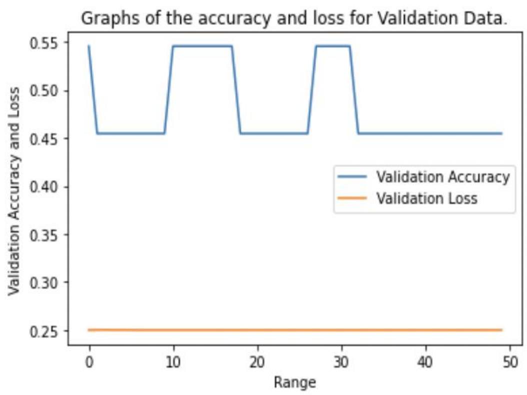


Figure 4: Validation Accuracy and Loss of initial data

1.6 FURTHER WORK ON ACCURACY IMPROVEMENT

Research Questions:

1. Do hyper Tuning bring any changes in the accuracy?
2. How we know that the given image is photo or not?
3. How can we increase the accuracy of the model?

1.7 QUESTION 1

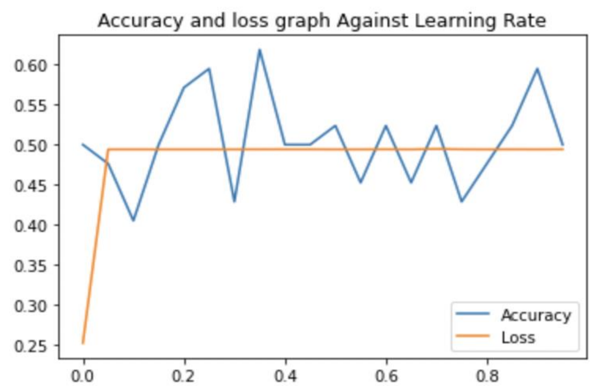
Does the Hyper parameter Tuning bring any changes in the accuracy?

The Hyper Tuning was done on parameters Activation Functions, Learning Rate, and number of cells in the layers

- **Hyper Tuning the number of cells and layers in the model.**
  While building the model there are several changes to the number of cells in each layer. We found that increase in the number of cells is directly proportional to the execution time of the model while training. So, after several observations in the time we decided to reduce the number of cells in the while increasing the number of layers. After few changes to the model after the observation, we came-up with a model that has of parameters around 290,000 which has decent execution time while performance.

- **Hyper Tuning the learning rate on the model created.**

  The model was checked with learning rates which are in range (0,1) with difference of 0.05. An iteration was performed on model with different learning rates and parallelly noted the accuracy and the loss values. Here the learning rate was taken with max accuracy rate.




```
The best learning rate: 0.35000000000000003
```

Figure 5: Validation Accuracy and Loss of initial data.

- **Hyper Tuning the Activation Functions.**
  There are six layers in the model without the output layer. Here the max activation function for each layer was initialized as "Relu" and the activation function of the output layer is "softmax". There are set of activation functions which are declared explicitly and passed while calling to the model. Few iterations were done on the model and checked the highest number of activations repeated with highest accuracy.

```
# Predefining the Variables
activation = [["relu","relu", "relu", "relu", "relu", "relu", "softmax"],
              ["relu", "relu", "relu", "LeakyReLU", "LeakyReLU", "LeakyReLU", "softmax"],
              ["relu", "LeakyReLU", "relu", "LeakyReLU", "relu", "LeakyReLU", "softmax"],
              ["LeakyReLU", "LeakyReLU", "LeakyReLU", "LeakyReLU", "LeakyReLU", "LeakyReLU", "softmax"],]
```

⇨ **Training the Model with the Hyper Tuned Parameters.**

After the choosing the parameters in the Hyper-Parameter tuning. There are set of parameters in the which has been updated in the model and trained. The accuracies of the model along with the validation data was shown below.

1.8 QUESTION 2

How we know that the given image is photo or not?

Here the images in the data are of human faces. How the model will know that the orientation of the face should same as humans. So, if the picture with separate of facial organs, means, eyes, nose, and lips are oriented differently, then how to know they are oriented different.

New Method:

The image point was replaced by the average of the eigen values neighbors. By this way we can store the feature of the neighbors in the point. If all points are replaced by same method, then every point holds the features of the relative points. By this way, we replaced the Image[I,J] value with the relative values and formed a relation with neighbors.

The model was trained with the present data and the graphs are show below.

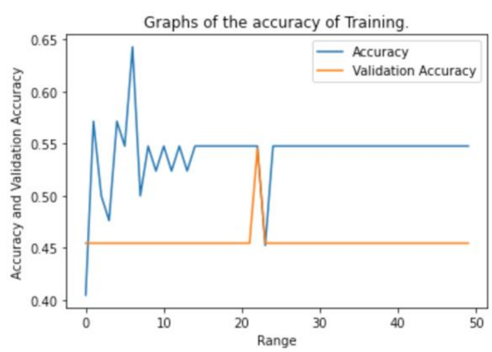


Figure 5: Training and Validation Accuracy

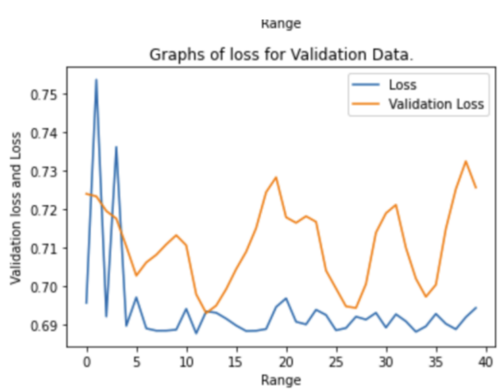


Figure 6: Training and Validation Loss

1.9 QUESTION 3

Data Augmentation.

Data Augmentation is the process which increases the size of the data by recreating form original data. There are few arguments the data augmentation which are,

1. Rotation.
2. Zoom.
3. Horizontal Flip.
4. Brightness Range

There arguments are used to generate the data.

While generating the data, each image was recreated to 10 images by applying the properties mentioned above. With the help of the data Augmentation, we tried to create the data or more than 5000 images from initial 52

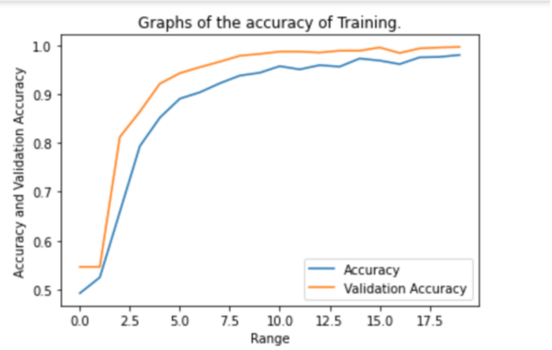


Figure 7: The Training and validation Accuracies of the data

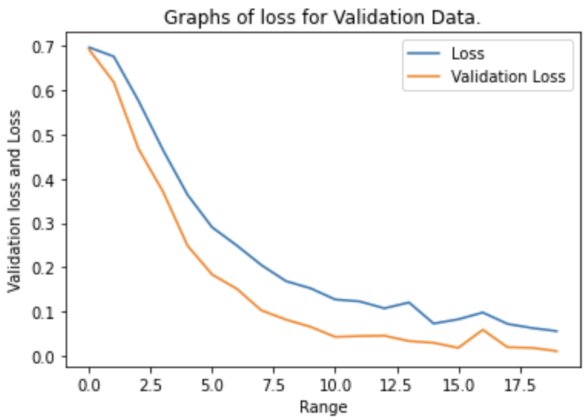


Figure 8: The Training and validation loss of the data

```
Epoch 12/20
43/43 [==============================] - 104s 2s/step - loss: 0.1234 - accuracy: 0.9504 - val_loss: 0.0453 - val_accuracy: 0.9869
Epoch 13/20
43/43 [==============================] - 103s 2s/step - loss: 0.1079 - accuracy: 0.9591 - val_loss: 0.0461 - val_accuracy: 0.9850
Epoch 14/20
43/43 [==============================] - 104s 2s/step - loss: 0.1211 - accuracy: 0.9561 - val_loss: 0.0340 - val_accuracy: 0.9887
Epoch 15/20
43/43 [==============================] - 104s 2s/step - loss: 0.0735 - accuracy: 0.9728 - val_loss: 0.0302 - val_accuracy: 0.9887
Epoch 16/20
43/43 [==============================] - 103s 2s/step - loss: 0.0831 - accuracy: 0.9685 - val_loss: 0.0187 - val_accuracy: 0.9953
Epoch 17/20
43/43 [==============================] - 105s 2s/step - loss: 0.0985 - accuracy: 0.9612 - val_loss: 0.0595 - val_accuracy: 0.9840
Epoch 18/20
43/43 [==============================] - 104s 2s/step - loss: 0.0728 - accuracy: 0.9751 - val_loss: 0.0202 - val_accuracy: 0.9934
Epoch 19/20
43/43 [==============================] - 103s 2s/step - loss: 0.0634 - accuracy: 0.9760 - val_loss: 0.0187 - val_accuracy: 0.9953
Epoch 20/20
43/43 [==============================] - 105s 2s/step - loss: 0.0565 - accuracy: 0.9800 - val_loss: 0.0114 - val_accuracy: 0.9962
34/34 [==============================] - 5s 145ms/step - loss: 0.0114 - accuracy: 0.9962
[0.011422784999012947, 0.9962441325187683]
```

Figure 9: The Accuracies and the loss of the Model with final data

Let's see the images of Augmented data,

Figure 10: Augmented Images of the Initial Figures

1.10 Conclusion

This project describes a Facial Detection that employs to detect the faces of the humans. Many current studies have investigated the problems in face emotion recognition and face detections in general, and relative face sensors have been tested on normal faces datasets. Few face detectors attain extremely high performance on the current datasets and improving them appears to be challenging. However, storing faces taken at unexpected resolution, luminance, and shadow is much more difficult in real life. Face detection is a critical task. The proposed CNN model uses user-build data set which has around 5000 images after data augmentation of two face photos of us for the Training of CNN model. The proposed model achieved an accuracy of 98% during the training phase and 97% during validation phase. Face recognition could make security checkpoints at airports less intrusive to passengers. In addition to this, they can be used by humanoid robots to analyze the emotions of the other people and for public security in analyzing the criminals.

1.11 References

[1] https://ieeexplore.ieee.org/document/8027997
[2] https://ieeexplore.ieee.org/document/9432156
[3] https://ieeexplore.ieee.org/document/9297483
[4] https://ieeexplore.ieee.org/document/9154121
[5] https://ieeexplore.ieee.org/document/8124509
[6] https://ieeexplore.ieee.org/document/8981757
[7] https://ieeexplore.ieee.org/document/8612852