Automatic Attendance System Based on Intelligent Face Recognition Scheme

Final year project report

Group Number

BSEE-FYP-B06-G01

Supervisor:

Dr Saqib Saleem

Co-Supervisor:

Moulana Muhammad Yasin

Department of Electrical & Computer Engineering
COMSATS University Islamabad, Sahiwal Campus

June 2021

Abstract

Face recognition is a biometric technology that is utilized to recognize human faces based on certain facial patterns as well as classify faces within various conditions. Recently face recognition has got popular and is utilized in various ways, especially in a security processes.The proposed project is an automatic attendance system based on intelligent face recognition scheme. Various methods of face detection as well as recognition are discussed in this report among them MTCNN and FaceNet are selected to detect face and generate face embeddings respectively. MTCNN is a Convolutional neural network. MTCNN consist of three neural network attached in cascaded form to detect face and generate bounding boxes. The FaceNet is based on deep-convolutional neural network as well as triplet loss training to generate embeddings from image dataset of faces. But training require complex computing hence takes a extended time. By integrating the Tensorflow module and pretrained weights training time can be reduced largly. And lastly Sikit-Learn library of Python is used to get cosine distance between input image and known dataset for classification purposes.

Acknowledgements

First and foremost we would like to thank Allah Almighty for blessing us with sound health, ability and courage to complete this project in successful manner. We would also like to thank our supervisor Dr Saqib Saleem for his invaluable supervision and support during the course of this project. Our acknowledgements also go out to our families and friends for supporting us throughout our studies.

Dedication

We would like to dedicate this project to our families and teacher for always being kind to us throughout our studies.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Introduction

The ability to identify individual face by examining the facial feature is known as face recognition. In last two decades there has been a rapid progress in advancement of facial recognition techniques. At first, the techniques were very slow and had low accuracy but introduction of deep learning in face recognition has solved these concerns. Lately, the face recognition techniques have been utilized to incorporate various duties s like confirmation of identity and accessibility of authority. There are numerous methods that can be used to develop a face recognition system. PCA (Principle component analysis) [1], LDA (Linear discriminant analysis) and deep learning are some widely used techniques.

PCA is widely used as an unsupervised technique for linear transformation. Its most prominent uses are dimension reduction and feature extraction. It is also utilized in denoising of signals and exploratory data analysis. Like PCA, LDA is also dimension reduction method but it uses supervised technique. Use of supervised techniques allow LDA to work effectively on datasets that contain specious amount of images. It uses closest average to identify classes of test objects. LDA's main benefit is that it is uncomplicated and fast but it is an aged techniques and only effective for classification.

Various deep learning techniques can also be utilized for development of face recognition system, such as Artificial neural network (ANN) and Convolutional Neural Network (CNN). FacNet is one method that utilize CNNs and Euclidean space to generate embeddings and similarity distance vectors for each face. In paper [5], FaceNet was able to achieve accuracy of 99.63 and 95.12 on two different datasets. The disadvantage of Face-Net is it's training is heavy and consume time and GPU is required to reach high speed.

1.2 Problem Statement

As the corona virus crisis arises its become very risky to use biometric attendance systems that require physical contact to the sensors. As physical contact to the sensors can cause the spread of the disease. Face recognition based attendance systems can be used to solve this problem as it is most effective biometric technique that can be used to mark attendance and it does not require any physical contact to sensors.

1.3 Introduction to FaceNet

Face-Net utilizes deep learning to produce embeddings, its an example of one-shot learning. In details, Face-Net develops the early model using small amount of face images. The initial model is then used without re-training. Euclidean space is utilized while the distance is composed of resemblance amongst face images. Once these similarity matrix are accumulated. Face recognition and classification becomes easy to perform. This project tend to propose FaceNet technique and test it on dataset containing images of student of Batch 6 of Dept of electrical and computer engineering.
Face-Net uses triplets by comparing faces to each other with the triplet-mining techniques. This triplet technique has a collection of anchor classes, wherein every anchor has positive as well as negative classes. This technique work by reducing the distance amongst images from positive class and anchor images, and increasing distance between anchor and negative class. The FaceNet trains directly into 128 dimensional vector called embeddings by using triplet loss method.

1.4 Multi-tasking Cascaded Convolutional Neural Networks (MTCNN)

MTCNN consist of 3 stages [6], that are used to generate the bounding-boxes of face locations in an image alongside 5 facial landmarks that are eyes, nose and mouth openings. Each stage improves the detection by passing input images through a CNN, that outputs candidate bounding-boxes confidence scores.In first stages image is resized multiple times to build an image pyramid shown in Figure 1.1.

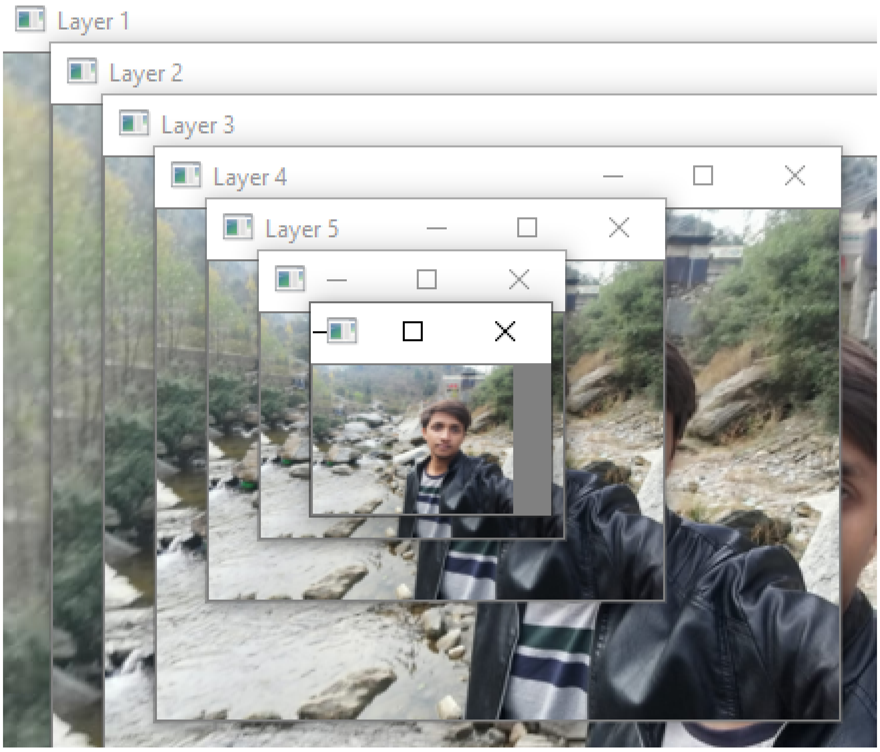


Figure 1.1: Image pyramid created by OpenCV

Each scaled image go through CNN. In stage 2 and 3 the patches containing faces in images are extracted for each bounding box and resized and sent to neural network of that particular stage.Other than the bounding boxes and confidence scores, stage 3 also detects 5 facial landmarks for each of bounding-boxes.

Chapter 2

Literature Review

2.1 Face detection approaches

Face detection is the first step in face recognition pipeline. The face detection methods are divided into 4 classes [28]. These classes are as follow.

2.1.1 Knowledge based approaches

These techniques are based on human knowledge of face detection. For example a face must have eyes, nose and mouth each at a certain distance from each other. The problem with these techniques is there could be many false positive as the set of rules are very generic. These techniques are not suitable for images that contain multiple faces. Multi-resolution rule based method is an example of these techniques [29].

2.1.2 Feature based approaches

These techniques work by locating faces by extraction of structural features of faces. These are trained to classify between facial and nonfacial regions in an image. These approaches show high accuracy even if an image has multiple faces. The examples of these techniques are SGLD of face face patterns [30].

2.1.3 Template Matching

Template matching techniques work by using predefined face template to locate faces in an image by finding corelation between the template and the image. These techniques does not work well with different posses of faces as well as multiple faces in an image. The example of these techniques is active shape model (ASM) [31].

2.1.4 Apperance based approaches

These techniques are based on training model on a set of set of faces images. These techniques are better than any other technique. Generealy these techniques depends upon the statistical analysis as well as machine learning to find the attributes of the face images. These techniques are also used in feature extraction while performing face recognition. Different PCA and neural network based techniques fall in this category. Example of these techniques are neural networks, eigen faces and inductive learning.

2.2 Face recognition approaches

There are 3 main groups of techniques that are used for face recognition on both still images and video frames that are holistic approach [7], feature based approach and hybrid approaches.

2.2.1 Holistic Approaches

In these techniques the complete face is considered as input for face recognition system. Eigen face, probabilistic eigen face, fisher face, SVM and NFL(nearest feature line) are examples of Holistic Approaches. These are all supported by PCA [1] techniques which are used to convert a dataset into lower dimension while holding the attributes of data-set.

2.2.2 Feature-based Approaches

As the name suggest in feature based approaches, facial features like nose, mouth openings and eyes are taken as input for the classifier. Likes of this category are Pure Mathematics, dynamic link design and hidden markov model. One amongst the majorly effective systems of this category is Elastic Bunch Graph matching system [8][9]. Wavelets, particularly Gabor wavelets plays an important role for face representation in such graph-matching techniques. The standard native feature illustration includes wave coefficients of various scales and rotations supported fixed wave bases. Such locally calculated wave coefficients play an important role in illumination customization, distortion, scaling and rotation.

2.2.3 Hybrid Approach

The concept of these techniques are based on human vision system, it perceives every holistic as well as native feature. The most effective factor that effect the efficiency of these methods incorporate a way to identify that options the ought to be combined, so their advantages are preserved and at the same time their disadvantages are averted. The task of recognizing the feature is very complex and remains unsettled even in techniques that have close relationship with hybrid approaches such as Multiple Classifier system (MCS)[10] and ensemble learning [11]. In regard to this various efforts has been made in this field to offer some insight toward such problem these lesson should be utilized to come up with a hybrid face recognition system. For example local and the global features posses different properties and hopefully can offer important information to solve the classification problem.

2.2.4 Hidden Markov Models Method

HMM is encouraging methodology that is effective for picture with variety in illumination, orientation and the expression of face. This method is collection of mathematical model that are used to define the attributes of the signal. It has great efficiency in speech and character recognition as long as the data is 1D. The model is presuppose to a markov process using unknown coefficients hence the main objective is to discover the unknown coefficients from the evident coefficients. Every HMM state contain a probability distribution across all potential results while in the regular markov model each state is obvious. In paper [8], the author based HMM on extraction of 2D discrete cosine transformation features. HMM generally contains unperceivable markoff chains with restricted numbers of status in model, initial state distribution ,the probability matrix B, state transition matrix A and set of PDFs.
A HMM is given as

$$\lambda = (\mathbf{A}, \mathbf{B}, \pi) \tag{2.1}$$

2.2.5 EigenFace

Turk Pentland developed a system named EigenFace in 1991, This system was able to perform face recognition using the Euclidean similarity [18] by utilizing PCA projection as the feature vector. This was the first eigen-space based face recognition system and, from then onward, various eigen-space based techniques were developed. Particular in 1997, FLD was used as projection algorithm in Fisher-faces system[19]. In all the eigen-space based techniques a similarity function is utilized to operate as nearest neighbor classifier[20].

2.2.6 FisherFace Technique

Fisher-face technique [19] is an algorithm that work on basis of the ratio between the change of one person's face vector with another person. It maximizes the determinant of class scatter matrix concurrently lowering the determinant of within class scatter matrix. Scatter matrix for a class c problem is given below:

$$S_b = \sum_{i=1}^{c} \Pr\left(\Omega_i\right)\left(\mu_i - \mu\right)\left(\mu_i - \mu\right)^T \tag{2.2}$$

$$\begin{gathered} S_w = \sum_{i=1}^{c} \Pr\left(\Omega_i\right)\Sigma_i \\ \text{where } \Sigma_i = \frac{1}{N}\sum\left(x_k - \mu_i\right)\left(x_k - \mu_i\right)^T \end{gathered} \tag{2.3}$$

is co-variance matrix of in-class samples. N_i is the amount of samples in class Ω_i Fisher criteria function is outlined as:

$$J(W) = \frac{\left|W^T S_b W\right|}{\left|W^T S_w W\right|}$$

Then the projective matrix W_{fd} can be chosen as follows:

$$W_{fd} = \arg\max_{W} \frac{\left|W^T S_b W\right|}{\left|W^T S_w W\right|}$$

W_{Ad} is calculated using eigen-value problem:

$$S_b W = S_w W \Lambda$$

2.2.7 Neural Network Method

Neural network based methods utilize different techniques from machine learning to disover the appropriate attributes of facial imagery. These attributes, in terms of discriminant function are eventually utilized for face recognition. Traditionally, face imagery is estimated in a low dimensional feature space and neural networks are used to develop nonlinear decision surface for classification[12].

The benefits of neural networks are good abstraction can be gained as the system is trained to get information about varity of facial patterns[13]. The main downside of these techniques is that the neural networks requires to be carefully tuned to achieve exemplary efficiency. Multilayer perception FaceNet and vggnet are mostly used neural network techniques.

2.2.8 FaceNet

As discussed in chapter 1 Face-Net utilizes deep neural networks to produce its embedding . In Face-Net the Euclidean space is utilized while the distance is composed of resemblance amongst face images. Once these similarity matrix are accumulated. Face recognition and classification becomes easy to perform. This project tend to propose FaceNet technique and test it on dataset containing images of student of Batch 6 of Dept of electrical and computer engineering. Face-Net uses triplets by comparing faces to each other with the triplet-mining techniques. This triplet technique has a collection of anchor classes, wherein every anchor has positive as well as negative classes. This technique work by reducing the distance amongst images from positive class and anchor images, and increasing distance between anchor and negative class. The FaceNet trains directly into 128 dimensional vector called embeddings by using triplet loss method.

Chapter 3

Methodology

3.1 Convolutional Neural Networks

Convolutional neural networks try to mimic the structure of the human brain. Human brain is constantly analysing things around a the person. Without any deliberate effort, humans make predictions about everything they see. They label everything on basses of what they have learned in the past. The alliance amongst the eyes and human brain, known as the primary vision pathway, allow humans to make sense of the world around them.

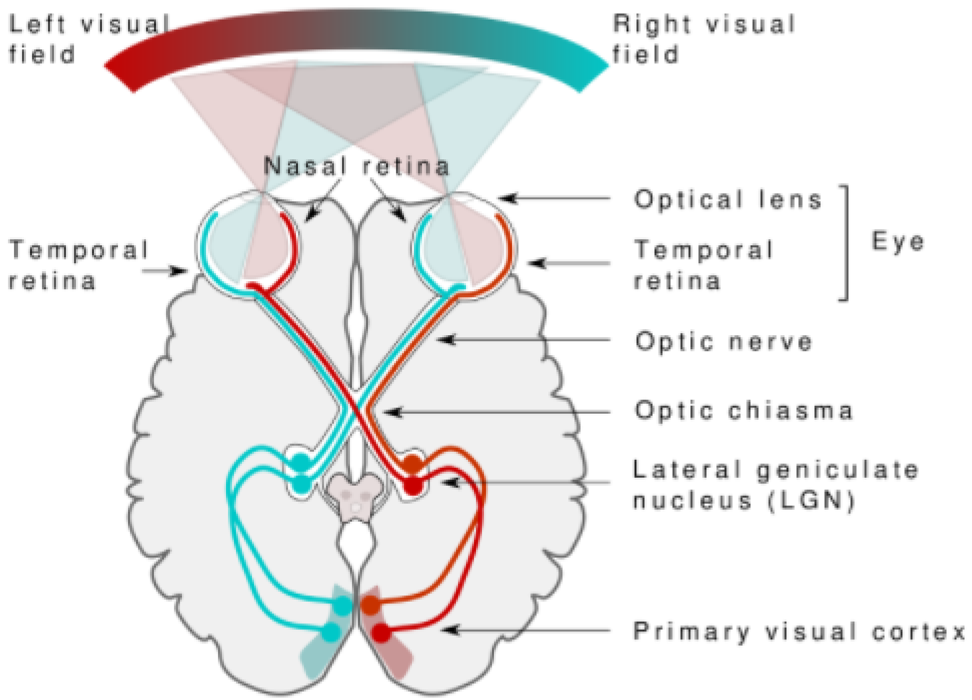
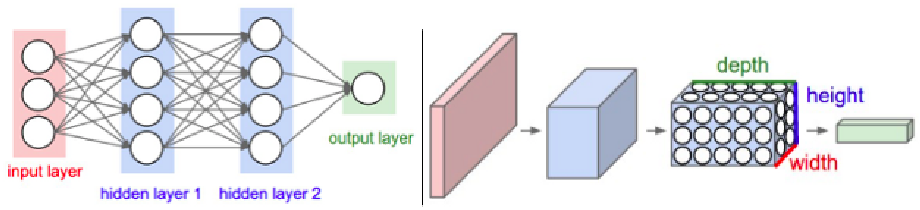


Figure 3.1: Primary vision pathway (Adapted from [26])

The vision starts at the eyes the actual analysis take place in brain, in primary visual cortex. A deeply complex hierarchical structure of neurons in the brain plays an important part in processing and labelling the objects.
Contemplate how human children learn to recognize different object. In the starting, their parent of elderly tell them names of different objects in direct environment. And slowly these objects start becoming recognizable. After some time they become so common the human recognize them instantly.
Similarly in CNN model have to be shown large amount of photos of an object for it to be able to recognize that object. For the computer the world only consists of numbers. Each image can be presented in a 2D arrays of numbers, known as pixels. In the regular neural networks these pixels goes thorough series of hidden layers of neurons where each layer is fully connected to all the neurons in the next layer. And the last FC layer represents prediction.
CNN are bit different, in CNNs the layers are arranged in three dimensions height, width and depth.And the layers are not fully connected with each other but only small regions of them are.



Left: A regular 3-layer Neural Network. Right: A ConvNet arranges its neurons in three dimensions (width, height, depth), as visualized in one of the layers. Every layer of a ConvNet transforms the 3D input volume to a 3D output volume of neuron activations. In this example, the red input layer holds the image, so its width and height would be the dimensions of the image, and the depth would be 3 (Red, Green, Blue channels).

Figure 3.2: Regular NN vs CNN (Adapted from [27])
The CNNs have two components the hidden layers and the the classification part. In the hidden layers CNNs perform multiple convolution and pooling operations to recognize patterns and extract features. If the input is an image of zebra this part will recognize stripes, ears and legs. In the classification part the last fully connected layer will work as an classifier on top of extracted features. It will assign probability for the objects in the input.

3.1.1 Feature extraction with CNNs

Convolution is the most important building block of CNN. The mathematical combination of 2 functions that produce a third function is called convolution. In CNN , Convolution used on the input data with a filter(kernal) to create a feature map. The execution is performed by sliding the kernal over the input image. At every stride a matrix multiplication is done and sums the results on the feature map.

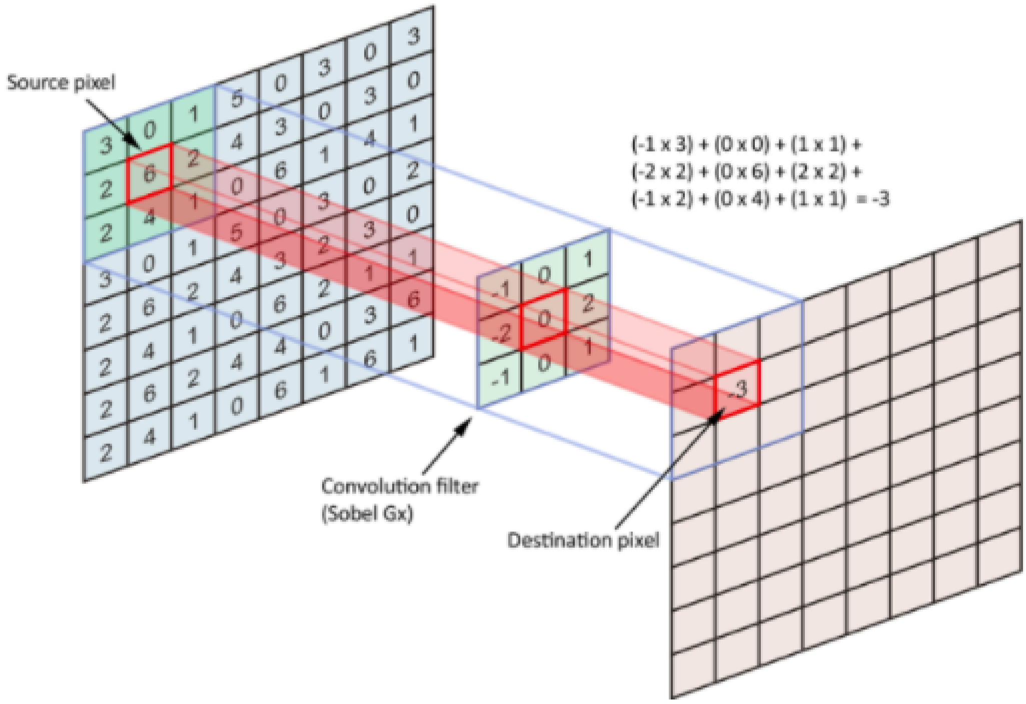


Figure 3.3: Convolution and the feature map (Adapted from [28])

As the size of the filter is always less than the input. A layer of zeros is added on the sides of the input to prevent feature map from shrinking. This phenomenon is called padding.
Mostly after a convolution layers, a pooling layer is added.The function of this layer is to reduce the dimensions to reduce the amount of parameters and computation in the network. The most common type of the pooling is max pooling which takes the maximum value of the window.

3.1.2 Classification with CNN

After the convolution and the pooling layers a fully connected layer is added in the end of the network. But these FC layers only accept 1 dimensional data. To convert the 3 dimensional data into 1 dimensional data flatten function of Python is used. It arranges the three dimensional data into 1 dimensional data.

3.1.3 Training with CNN

Training in CNN is performed using the gradient descent of backpropagration techniques.

3.1.3.1 Gradient Descent

Gradient descent is an algorithm used for optimization, it finds a set of input variables for a function that results in minimum value of the function.It involves finding the gradient of the function. The gradient descent algorithm require calculation of the gradient with respect to specific values of the variables. As the gradient points uphill, hence the negative gradient point toward downhill to result in new values for each variable that result in lower evaluation of the target.

3.1.3.2 Back-propagation

Back propagation also find gradient with respect to functions variables. The back part of the name suggest that the propagation proceeds backwards. Gradients for final layer is calculated first and the first layer at the end.The first-order derivative of a function for a specific value of an input variable is the rate of change of the function for that input. When there are multiple input variables for a function, they form a vector and the vector of first-order derivatives is called the gradient.Backpropagation is used while training CNN models to calculate the gradient for all weights in the network model. The gradient is then utilized by an optimization algorithm to update the weights of the model.

3.2 Face recognition pipeline

The problem of identifying different faces in a still image or video frame is known as face recognition. The recognition process can be divided into multiple steps. The following image show example of face recognition pipeline.

- Face detection — Finding different faces in an image..
- Feature extraction — Extracting important feature from face imagery.
- Face classification — Using the extracted features to classify faces.

Figure 3.4: Face Recognition pipeline

3.3 Face Detection

MTCNN is used to perform the face detection. As the name suggest MTCNN is model containing three neural networks (P-net, R-net and O-net) connected to each other[6]. A video frame or an image is passed to MTCNN. In order to detect faces of different sizes, MTCNN creates an image pyramid by creating copies of different sizes of input image.

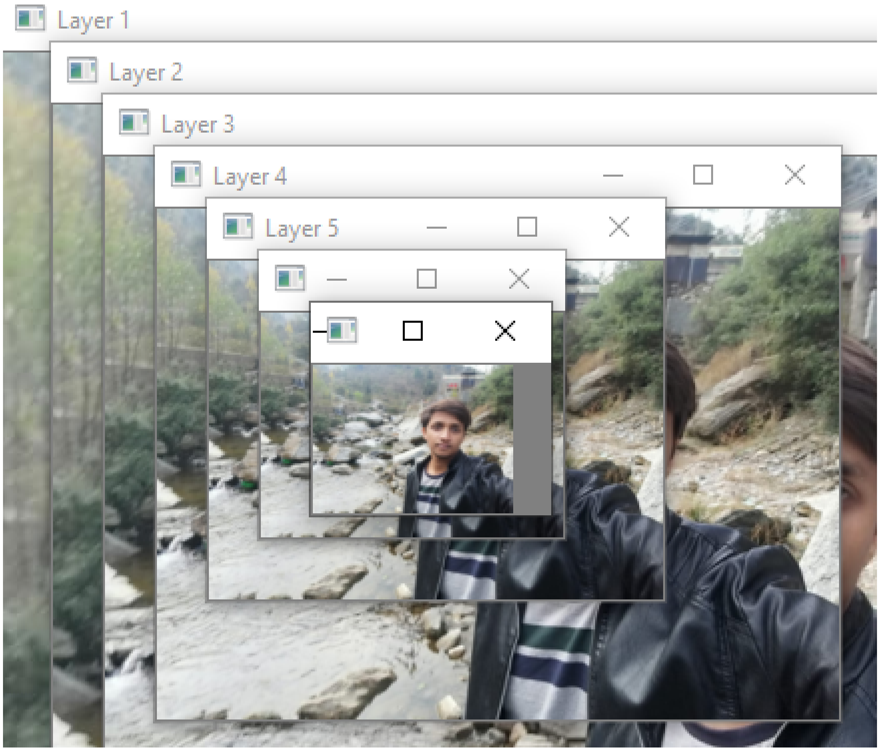


Figure 3.5: Image pyramid created by OpenCV

MTCNN is used to perform the face detection. As the name suggest MTCNN is model containing three neural networks (P-net, R-net and O-net) connected to each other[6]. A video frame or an image is passed to MTCNN. In order to detect faces of different sizes, MTCNN creates an image pyramid by creating copies of different sizes of input image.

MTCNN creates a 12 x 12 kernal that go through every part of each scaled copy scanning for faces. It starts from (0,0) to (12, 12). P-net takes this portion as an input to return bounding boxes if a face is found. Then the filter will move with a stride of 2 and move to (0+2a, 0+2b) to (12+2a, 12+2b), now this portion will be passed to the P-net to scan for faces. Sometimes, instead of full face an image might contain a part of face peeking in from side of image. In that case P-Net might return coordinates of bounding boxes that is partially outside of the frame. For each bounding box array of same size is created then MTCNN copy this array to a new array. If the bounding box is partially outside the frame MTCNN only copy part of frame to the new array and everything else is replaces with zero. This process is known as padding.

Once each of the bounding boxes are resized, they are feed to R-net. Output of R-net is similar to P-net, It contains new and more accurate coordinate of bounding boxes, and additionally it also contain confidence level for each bounding box.

Then these bounding boxes are feed to the O-net. The output of O-net is slightly different from other networks. It contain 3 outputs the coordinates of bounding boxes, 5 facial landmarks(mouth openings, nose and eyes) and confidence level for each of the bounding boxes.

The Structure MTCNN is show in fig3.3.

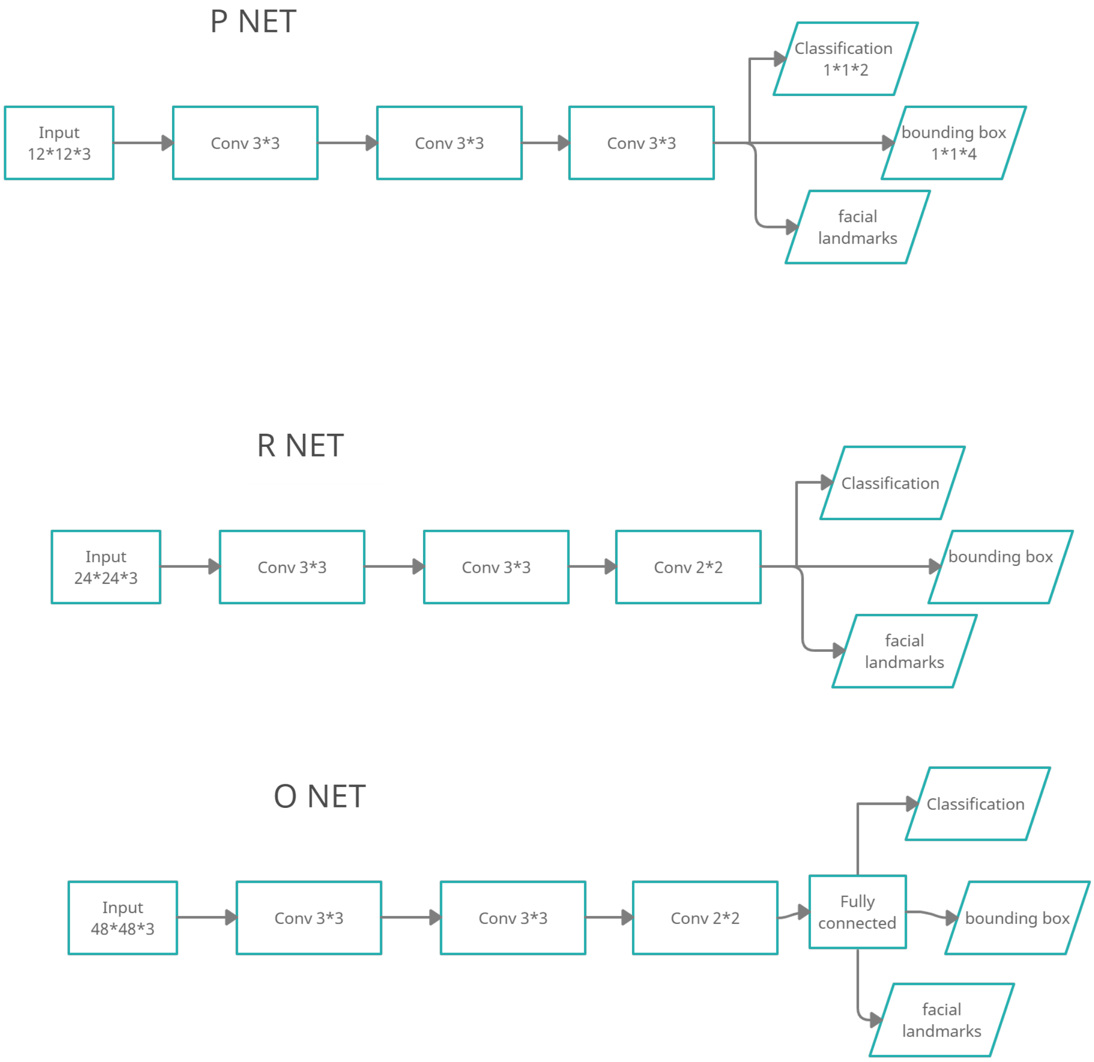


Figure 3.6: Structure of MTCNN

3.4 Feature Extraction with FaceNet

FaceNet is used as feature extractor. FaceNet is a inception based pretrained convolution network that is used for extraction of facial features from facial imagery. It was created by Google researchers in 2015. It takes an image of the face as input and return a vector of 128 numbers. The output is called embedding as all the important features of the face are embedded into a single vector. Ideally the embeddings of two similar face must be similar. Facenet calculates

embedding of each face and calculate the distance of these embeddings with embeddings of each known face present in the dataset. And if the embeddings are closer than a certain treshhold to a known person A. The FaceNet classify the face as face of person A.
FaceNet uses following steps to learn:

- Select a random image.
- Select an image of same person (positive class).
- Select image of different person.
- Adjusts the Face-Net network parameters so that the positive class is closer to the anchor than the negative class.

These steps are repeated until there are no more changes to be made. And all picture of a person are closer to each other and far from others. This method is known as triplet loss method.
The following equation describe the triplet loss function.

$$\mathcal{L}(A, P, N) = \max\left(\|\mathrm{f}(A) - \mathrm{f}(P)\|^2 - \|\mathrm{f}(A) - \mathrm{f}(N)\|^2 + \alpha, 0\right) \tag{3.1}$$

A is an input, P is positive, N is negative and alpha is a margin between the positive and the negative. The parameter summary of the FaceNet is given below.

Layer	size-in	size-out	kernel	param	FLPS
conv1	$220 \times 220 \times 3$	$110 \times 110 \times 64$	$7 \times 7 \times 3, 2$	9 K	115M
pooll	$110 \times 110 \times 64$	$55 \times 55 \times 64$	$3 \times 3 \times 64, 2$	0	
morm 1	$55 \times 55 \times 64$	$55 \times 55 \times 64$		0	
conv2a	$55 \times 55 \times 64$	$55 \times 55 \times 64$	$1 \times 1 \times 64, 1$	4 K	13M
conv2	$55 \times 55 \times 64$	$55 \times 55 \times 192$	$3 \times 3 \times 64, 1$	111 K	335M
morm2	$55 \times 55 \times 192$	$55 \times 55 \times 192$		0	
pool2	$55 \times 55 \times 192$	$28 \times 28 \times 192$	$3 \times 3 \times 192, 2$	0	
conv3a	$28 \times 28 \times 192$	$28 \times 28 \times 192$	$1 \times 1 \times 192, 1$	37 K	29M
conv3	$28 \times 28 \times 192$	$28 \times 28 \times 384$	$3 \times 3 \times 192, 1$	664 K	521M
pool.	$28 \times 28 \times 384$	$14 \times 14 \times 384$	$3 \times 3 \times 384, 2$	0	
conv4a	$14 \times 14 \times 384$	$14 \times 14 \times 384$	$1 \times 1 \times 384, 1$	148 K	29M
conv4	$14 \times 14 \times 384$	$14 \times 14 \times 256$	$3 \times 3 \times 384, 1$	885 K	173M
conv5a	$14 \times 14 \times 256$	$14 \times 14 \times 256$	$1 \times 1 \times 256, 1$	66 K	13M
conv5	$14 \times 14 \times 256$	$14 \times 14 \times 256$	$3 \times 3 \times 256, 1$	590 K	116M
conv6a	$14 \times 14 \times 256$	$14 \times 14 \times 256$	$1 \times 1 \times 256, 1$	66 K	13M
conv6	$14 \times 14 \times 256$	$14 \times 14 \times 256$	$3 \times 3 \times 256, 1$	590 K	116M
pool4	$14 \times 14 \times 256$	$7 \times 7 \times 256$	$3 \times 3 \times 256, 2$	0	
concat	$7 \times 7 \times 256$	$7 \times 7 \times 256$		0	
fcl	$7 \times 7 \times 256$	$1 \times 32 \times 128$	maxout p = 2	103M	103M
fc2	$1 \times 32 \times 128$	$1 \times 32 \times 128$	maxout p = 2	34M	34M
fc7128	$1 \times 32 \times 128$	$1 \times 1 \times 128$		524 K	0.5M
L2	$1 \times 1 \times 128$	$1 \times 1 \times 128$		0	
total				140M	1.6 B

Table 3.1: Structure of FaceNet

3.5 Face classification

Now its time to compare the embedding of unknown face with known faces. Cosine Distance similarity function of Sikit-Learn library is used to classify faces. First the embedding vectors are normalize.And then passed to sklearn to classify by comparing to known faces

Chapter 4

System Design and Development

4.1 System Design

Face recognition is computer vision task to recognize various faces in an image or video frame. To perform this task different modules of python have been used. OpenCV is used to capture video from the webcam. Tensorflow and Keras are used to implement the neural network models MTCNN and FaceNet which are used to detect faces from images and generate embeddings containing facial features respectively.Pickle library is used to store embedding of datasets. Once the dataset is created SKlearn library is used for classification. Attendance database is created using Pandas library.
Pipeline for system design is as folllow.

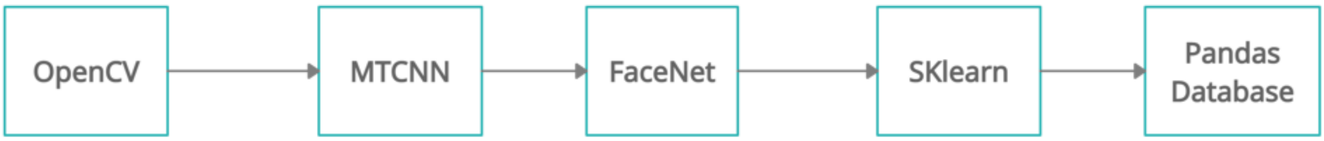


Figure 4.1: Structure of MTCNN

4.1.1 OpenCV

OpenCV is a library of programming functions mainly used for real time computer vision. In this project OpenCV is used to capture video using webcam and pass each frame of the video to neural networks for further processing. Once all the processing is done OpenCv is utilized to play the real time video with outputs of face detection and recognition system drawn on each frame of the video[21].

4.1.2 Tensorflow

Tensorflow is an open source library used for machine learning. A range of task can be performed using tensorflow but it is perticularly used for training the deep neural networks. Initially It was developed by the Google Brain team for only internal use by google. But in 2015 it was released as an open source library. Tensoflow can run on mulitiple CPUs and GPUs and has CUDA compatibility that can spped up the training process.
In tensorflow the computations are expressed in terms of the stateful data-flow graphs. The name tensorflow is inspired by the fact that the opertion performed to multi-dimensional arrays by neural networks are knowns as tensors.In 2016 just one year after the release of tensorflow there were 1500 repositories on github based on tensorflow and only 5 of them were from google.[22]
Tensorflow is used in this project to implement MTCNN face detection system as well as feature extraction using the FaceNet model.

4.1.3 Keras

Keras is an open source library used as interface for Artificial neural networks. Tensorflow library is used to create ANNs and Keras is used as an interface for Tensorflow. Keras has implementation of building block of numerous commonly used networks. These blocks contain layers, activation functions and optimizers etc.Keras also contain tools that make working with image and text data easier. In addition to ANN blocks it also contain dropout, pooling and batch normalization block that are used in implementation of convolutional neural networks.
FaceNet model is developed using Keras interface hence require Keras compatability.[23]

4.1.4 SKlearn

SKlearn is an open source machine learning library. It includes numerous implementations of classifiers such as SVM,KNN and cosine distance classifier. It also include implementation of various dimension reduction algorithm such as LDA and PCA. In this project Sklearn is used to implement face classifier.[24]

4.1.5 Pandas

Pandas is data manipulation library written in python language. Particularly, it offers operation for manipulating numerical data as well as time series.In this project Pandas is used to build an attendance database.[25]

Chapter 5

Results And Conclusion

5.1 Results

5.1.1 Face Detection

Convolutional Neural networks model MTCNN is used for face detection which takes a video frame as an input an provide bounding boxes with coordinates of faces and facial landmarks.

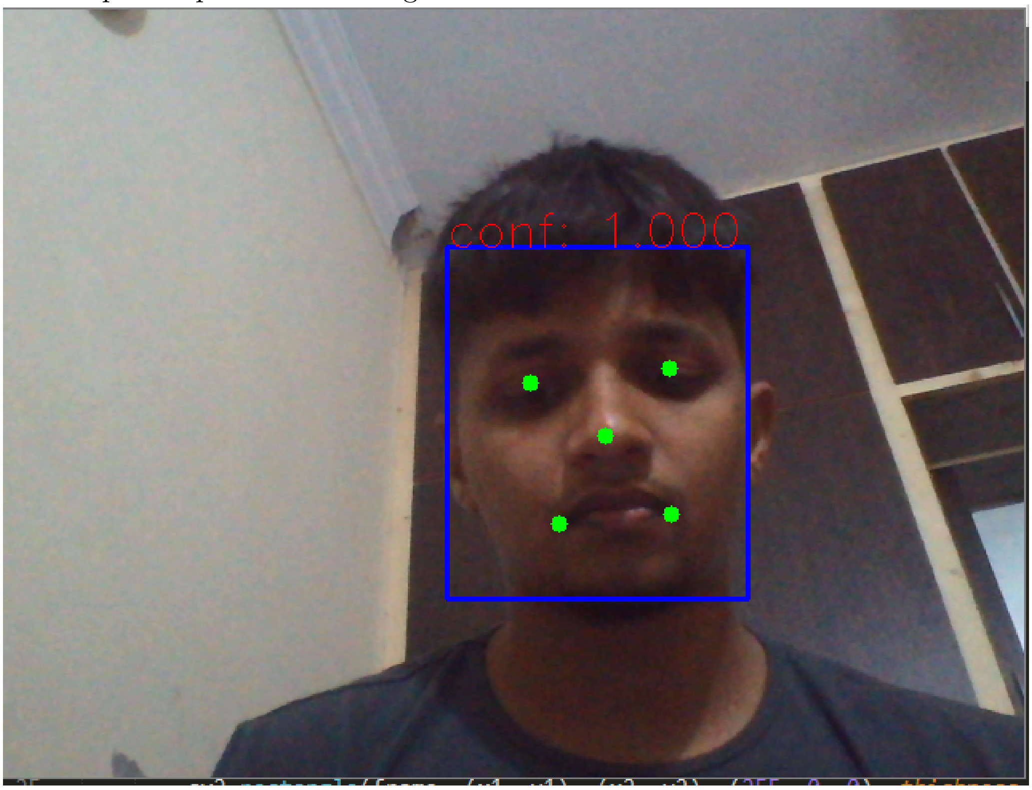


Figure 5.1: Face detection using MTCNN

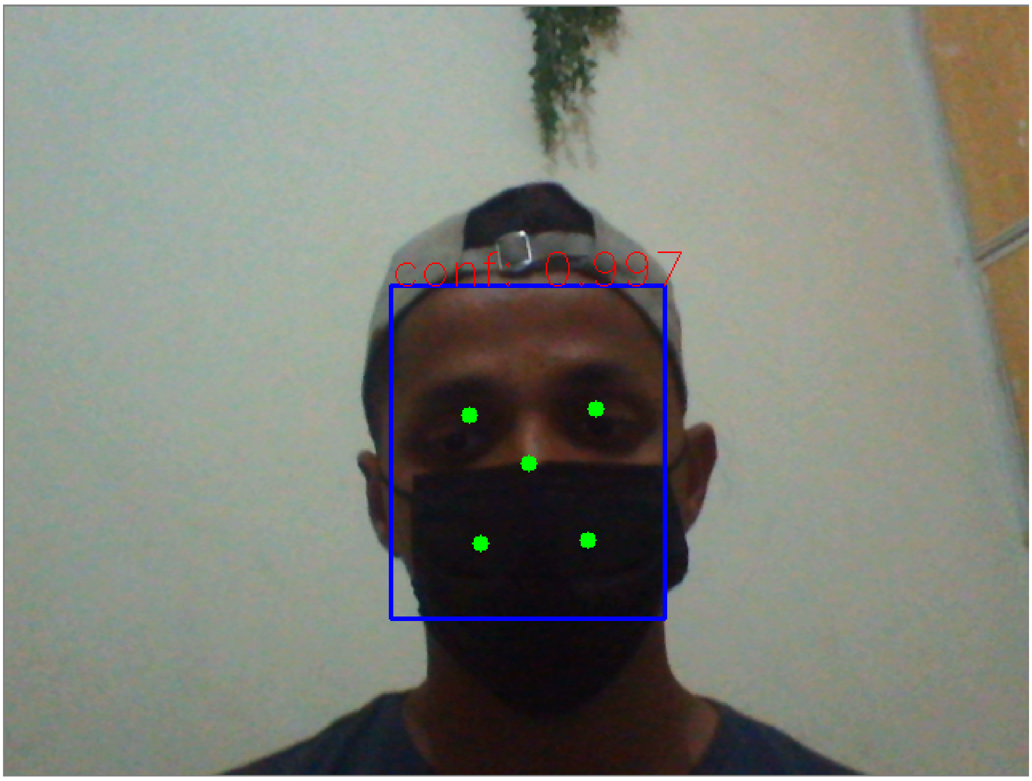


Figure 5.2: Face detection while wearing a mask using MTCNN

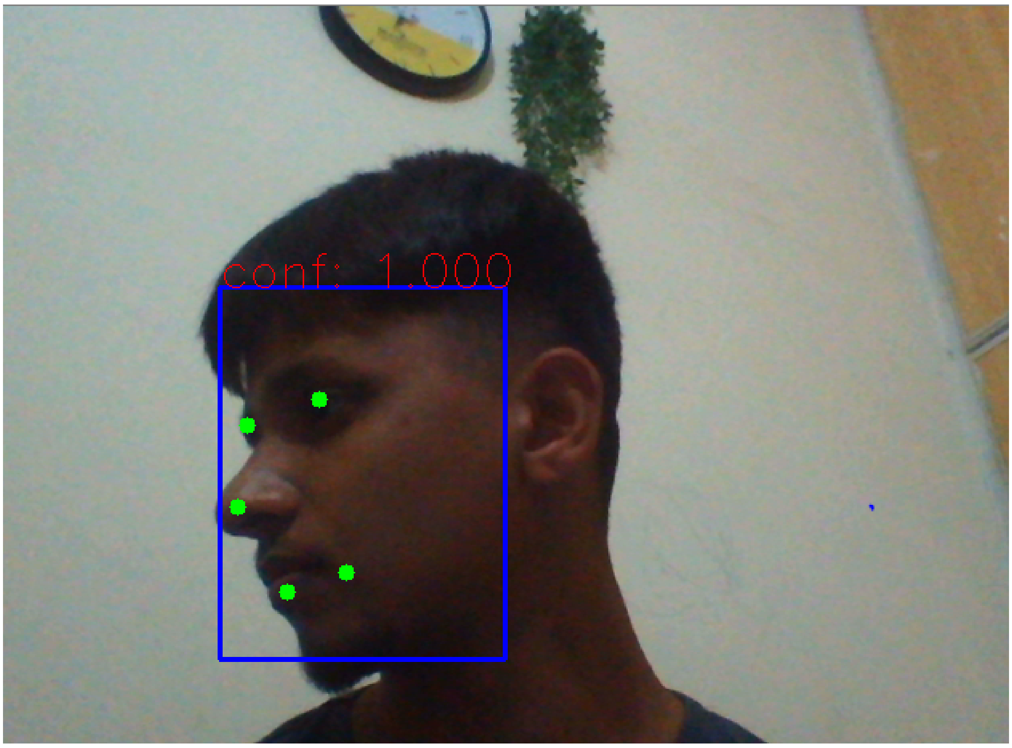


Figure 5.3: Face detection from side angle using MTCNN

Figure 5.4: Multiple Faces detection using MTCNN

5.1.2 Face Recognition

Images of faces are fed to the FaceNet model which returns embeddings containing facial landmarks which are compared to the database using the classifiers to identify the person.

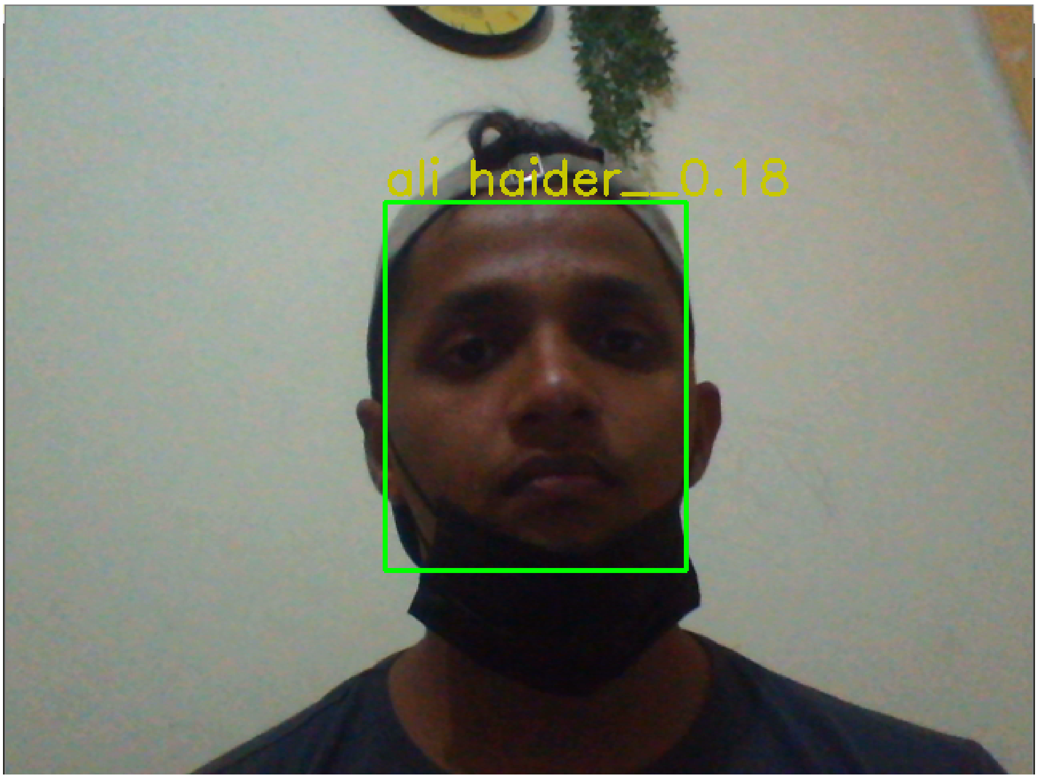


Figure 5.5: Face recognition using FaceNet

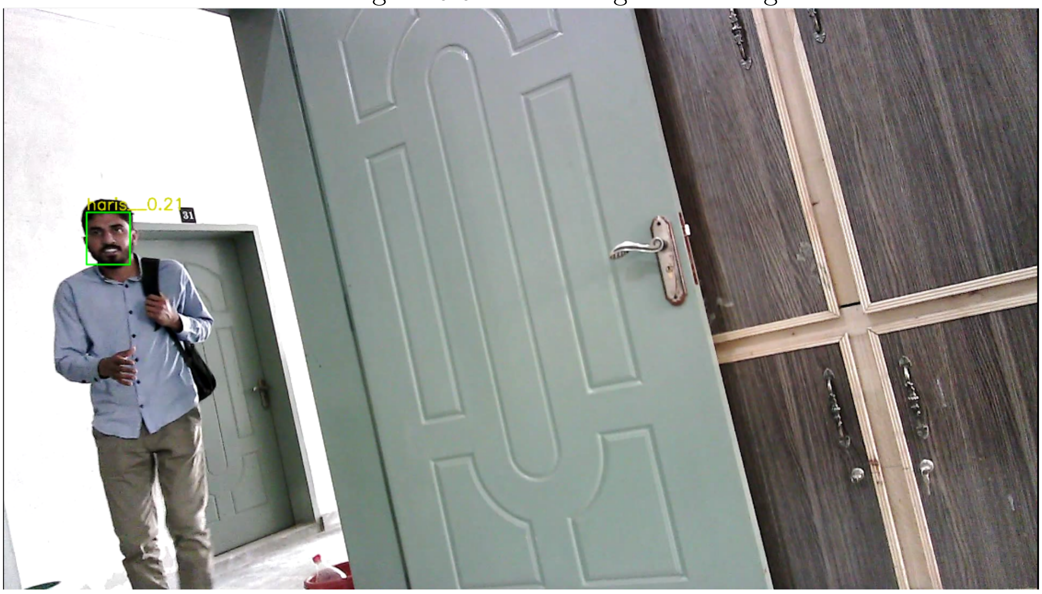


Figure 5.6: Face recognition using FaceNet

Figure 5.7: Multiple Faces recognition using FaceNet

Figure 5.8: Multiple Faces recognition using FaceNet

Figure 5.9: Unknown Faces recognition using FaceNet

5.2 Accuracy

While testing on 74 images of Batch 6 of ECE department of CUI sahiwal the system showed an accuracy of 0.958904109589041.

```
perdicted output
['aoun', 'arbab', 'arbab', 'azham', 'azham', 'hamza asghar', 'hamza asghar', 'hamza gardazii', 'hamza gardazii', 'Hamza lati
f', 'hamza rao', 'hamza rao', 'hamza sarwar', 'hamza sarwar', 'Haris', 'Haris', 'haris dogar', 'haris dogar', 'jangsher', 'ja
ngsher', 'kashif', 'kashif', 'muneeb', 'muneeb', 'nabeel', 'nabeel', 'naeem', 'naeem', 'naeem saleem', 'naeem saleem', 'sulam
an', 'nouman sarfaraz', 'sulaman', 'nouman khan', 'nouman khan', 'nouman sarfaraz', 'nouman sarfaraz', 'rayast', 'rayast', 's
ameer', 'sameer', 'shabir', 'shabir', 'shahid', 'shahid', 'shazad', 'shazad', 'shazad ahmad', 'shazad ahmad', 'shoiab', 'shoi
ab', 'sulaman', 'sulaman', 'tallat', 'tallat', 'tazeem', 'tazeem', 'usama irshad', 'usama irshad', 'usama khalid', 'waqar',
'waqar', 'Waqas', 'Waqas', 'waseem', 'yaseen', 'yaseen', 'zahid', 'zahid', 'zaid maqsood', 'zaid maqsood', 'zain', 'zain']
expected outputs
['aoun', 'arbab', 'arbab', 'azham', 'azham', 'hamza asghar', 'hamza asghar', 'hamza gardazii', 'hamza gardazii', 'Hamza lati
f', 'hamza rao', 'hamza rao', 'hamza sarwar', 'hamza sarwar', 'Haris', 'Haris', 'haris dogar', 'haris dogar', 'jangsher', 'ja
ngsher', 'kashif', 'kashif', 'muneeb', 'muneeb', 'nabeel', 'nabeel', 'naeem', 'naeem', 'naeem saleem', 'naeem saleem', 'noman
60', 'noman 60', 'noman 60', 'nouman khan', 'nouman khan', 'nouman sarfaraz', 'nouman sarfaraz', 'rayast', 'rayast', 'samee
r', 'sameer', 'shabir', 'shabir', 'shahid', 'shahid', 'shazad', 'shazad', 'shazad ahmad', 'shazad ahmad', 'shoiab', 'shoiab',
'sulaman', 'sulaman', 'tallat', 'tallat', 'tazeem', 'tazeem', 'usama irshad', 'usama irshad', 'usama khalid', 'waqar', 'waqa
r', 'Waqas', 'Waqas', 'waseem', 'yaseen', 'yaseen', 'zahid', 'zahid', 'zaid maqsood', 'zaid maqsood', 'zain', 'zain']
test accuracy on 74 test images =  0.958904109589041
```

Figure 5.10: System Accuracy

5.3 Conclusion

The main purpose of the proposed project is to implement biometric attendance system based of facial recognition. Different face recognition techniques are discussed in this write-up but FaceNet based technique is implemented which directly learn face embeddings, which separates it from other techniques that uses CNN bottle-nect layer, or require addition processing through PCA , LDA and SVM classifications. FaceNet performs end to end training and directly optimize the loss function which simplifies the implementation and improves accuracy. FaceNet does not require face alignment as experiments[5] with similarity transform alignment does not effect the performance much.

The model performed well under varying condition. However few problems were found while testing such as : 1) uncommon poses sometime make model to give unaccurate output. 2) blurry frames also effects models accuracy. However as the face recognition is performed on real time video (there are multiple frames containing same face) these problems can be solved by decreasing the threshold of the classifiers. Training time is an other big issue. If GPU is not available it can take hours to train the neural networks model on a large dataset. Future work should focus on solving these problems

Bibliography

[1] X. Chen, L. Song and C. Qiu, ”Face Recognition by Feature Extraction and Classification,” 2018 12th IEEE International Conference on Anti-counterfeiting, Security, and Identification (ASID), pp. 43-46, 2018.

[2] M. E. Rane and A. J. Pande, ”Multi-Modal Biometric Recognition of Face and Palm-Print Using Matching Score Level Fusion,” 2018 Fourth International Conference on Computing Communication Control and Automation (ICCUBEA), pp. 1-6, 2018.

[3] M. Korkmaz and N. Yilmaz, ”Face Recognition by Using Back Propagation Artificial Neural Network and Windowing Method,” 2015 2nd International Conference on Artificial Intelligence (ICOAI 2015), vol. 4, no. 1, pp. 15-19, 2015.

[4]] O. M. Parkhi, A. Vedaldi, and A. Zisserman, ”Deep Face Recognition,” Proceedings of the British Machine Vision Conference 2015, no. Section 3, pp. 41.1-41.12, 2015.

[5] F. Schroff, D. Kalenichenko and J. Philbin, ”FaceNet: A Unified Embedding for Face Recognition and Clustering,” 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 815-823, 2015.

[6] TY - JOUR AU - Zhang, Kaipeng AU - Zhang, Zhanpeng AU - Li, Zhifeng AU - Qiao, YuPY - 2016/04/11 SP - T1 - Joint Face Detection and Alignment Using Multitask CascadedConvolutional Networks VL - 23 DO - 10.1109/LSP.2016.2603342 JO - IEEE Signal ProcessingLetters ER -

[7] Zhao W., Chellappa R., Phillips P. J., and Rosenfeld A., Face recognition: A literature survey, ACM Computing Surverys (CSUR), 35(4), 2003, 399- 458

[8] Okada K., Steffans J., Maurer T., Hong H., Elagin E., Neven H., Andmalsburg C. V. D., (1998). The Bochum/USC Face Recognition System and how it fared in the FERET Phase III Test. In Face Recognition: From Theory to Applications, Eds. Springer-Verlag, Berlin, Germany, (1998) 186-205.

[9] Wiskott L., Fellous J.-M., and Von dermalsburg C., (1997), Face recognition by elastic bunch graph matching. IEEE Trans. Patt. Anal. Mach. Intell. 19, (1997) 775-779.

[10] Kittler J., Hatef M., Duin R.P.W., and Matas J., On combining classifiers. IEEE Trans. Pattern Analysis and Machine Intelligence, 20(3), 1998, 226- 239.

[11] Zhou Z.H., Wu J., and Tang W., Ensembling neural networks: Many could be better than all. Artificial Intelligence, 137(1-2), 2000, 239-263

[12] Rowley H., Baluja S., and Kanade T., Neural network-based face detection, IEEE Trans. Pattern Anal. Mach. Intell, 20, 1998, 23-38

[13] Er M.J., Wu S., Lu J., Toh H.L., Face recognition with radial basis function (RBF) neural networks, IEEE Trans. Neural Net,13, 2002, 697- 710.

[14] Osuna E., Freund R., Girosi F., Training support vector machines: an application to face detection, in: Proceedings of the IEEE Conference Computer Vision and Pattern Recognition,1997, 130-136.

[15] Yang F., Paindovoine M., Implementation of an RBF neural network on embedded systems: real-time face tracking and identity verification, IEEE Trans. Neural Network, (14), 2003, 1162-1175.

[16] Valentin D., Abdi H., O'Toole A.J., and Cottrell G.W., Connectionist models for face processing: a survey, Pattern Recognition, 27, (1994), 1209–1230

[17] Q. Cao, L. Shen, W. Xie, O. M. Parkhi and A. Zisserman, "VGGFace2: A Dataset for Recognising Faces across Pose and Age," 2018 13th IEEE International Conference on Automatic Face Gesture Recognition (FG 2018), Xi'an, China, 2018, pp. 67-74, doi: 10.1109/FG.2018.00020.

[18] Turk M. and Pentland A., Eigenfaces for Recognition, J. Cognitive Neuroscience, 3(1), 1991, 71-86.

[19] Belhumeur P.N., Hespanha J.P., and Kriegman D.J., Eigenfaces vs. Fisherfaces: recognition using class specific linear projection, IEEE Trans. Pattern Analysis and Machine Intelligence, 19(7), July 1997, 711-720.

[20] Duda R.O., Hart P.E., and Stork D.G., Pattern Classification, Second Edition, 2001.

[21] https://opencv.org/

[22] https://www.tensorflow.org/

[23] Ketkar, Nikhil. (2017). Introduction to Keras. 10.1007/978-1-4842-2766-$4_7. MatthieuDuchesnay, EdouardLouppe, Gilles.(2012).Scikit - learn : MachineLearninginPython.JournalofMachineLearningResearch.12.$

[24[24] https://pandas.pydata.org/

[25] https://www.freecodecamp.org/news/an-intuitive-guide-to-convolutional-neural-networks

[26] https://commons.wikimedia.org/wiki/File:Human$_visual_pathway.svghttp$: $//cs231n.github.io/convolutional - networks/$

[27][27] https://towardsdatascience.com/applied-deep-learning-part-4-convolutional-neural-networks-584bc134c1e2

[28] Yang, Ming-Hsuan Kriegman, David Ahuja, Narendra. (2002). Detecting Faces in Images: A Survey. Pattern Analysis and Machine Intelligence, IEEE Transactions on. 24. 34 - 58. 10.1109/34.982883.

[29] San, Tang. (2013). Human Face Detection Method Based on Skin Color Model. Advanced Materials Research. 706-708. 1877-1881. 10.4028/www.scientific.net/AMR.706-708.1877.

[30] Dai, Y. and Y. Nakano. "Face-texture model based on SGLD and its application in face detection in a color scene." Pattern Recognit. 29 (1996): 1007-1017.

[31] Le, Thai Truong, Vo. (2011). Face Alignment Using Active Shape Model And Support Vector Machine. International Journal of Biometric and Bioinformatics.