An Efficient Face Recognition Model Using Multiple Angular Images and Deep Neural Network Architecture

by

Silma Subah
18201088
Arpita Nath
18201089
Mitheela das Armisha
18201101
Sumaiya Binte Sakhawat
18201076
Md Nuhas Alam Parbo
17201020

A thesis submitted to the Department of Computer Science and Engineering
in partial fulfillment of the requirements for the degree of
B.Sc. in Computer Science and Engineering

Department of Computer Science and Engineering
School of Data Science
Brac University
January 2023

© 2023. Brac University
All rights reserved.

Declaration

It is hereby declared that

1. The thesis submitted is my/our own original work while completing degree at Brac University.
2. The thesis does not contain material previously published or written by a third party, except where this is appropriately cited through full and accurate referencing.
3. The thesis does not contain material which has been accepted, or submitted, for any other degree or diploma at a university or other institution.
4. We have acknowledged all main sources of help.

Student's Full Name & Signature:

silma subha

Silma Subah
18201088

Arpita Nath
18201089

Mitheela Das

Mitheela das Armisha
18201101

Sumaiya Binte Sakhawat
18201076

Nuhas Parbo

Md Nuhas Alam Parbo
17201020

Approval

The thesis/project titled " An efficient face recognition model using multiple angular images and deep neural network architecture " submitted by

1. Silma Subah (18201088)
2. Arpita Nath (18201089)
3. Mitheela das Armisha(18201101)
4. Sumaiya Binte Sakhawat(18201076)
5. Md Nuhas Alam Parbo(17201020)

Of fall, 2023 has been accepted as satisfactory in partial fulfillment of the requirement for the degree of B.Sc. in Computer Science and Engineering on January 23 , 2023.

Examining Committee:

Supervisor:
(Member)

Dr. Md. Ashraful Alam
Assistant Professor
Department of Computer Science and Engineering
School of Data Science
Brac University

Program Coordinator:
(Member)

Md. Golam Rabiul Alam
Program Coordinator
Department of Computer Science and Engineering
School of Data Science
Brac University

Head of Department:
(Chair)

Sadia Hamid Kazi
Chairperson
Department of Computer Science and Engineering
School of Data Science
Brac University

Abstract

Face surface information in three dimensions is one of the promising biometric modality that can improve the identification and increase the accuracy of verification of face recognition systems in challenging situations.This study proposed a system that recognizes faces from multiple angular images and deep neural networks.The proposed model can be divided into three steps: image acquisition, processing, and recognition.In acquisition part we take multiple angular images of the face which was taken by us and the angle was (0° to 180°)whereas right side was considered as positive(0° to +90°) and left side was considered as negative(0°to -90°). After that the images using Haar cascade and MTCNN algorithm segment the image, specially the face area.Then we used deep learning model VGG16,VGG19,InceptionNetV3 and ResNet50 to determine the face of person where the accuracy were 97%,92%,98%and 98% respectively.This article aggregates data from openly available multiple angle face databases to enable future research easier. The proposed system achieved more accuracy than the existing face recognition models when angle or motion is considered. That's why we came up with an idea of various multiple angles which can detect a person in motion. The proposed system enables efficient face recognition in dynamic motion as well as with different angular deviations.It achieved higher accuracy than the existing 2D face recognition systems when the target object is in motion.

Keywords: Machine Learning; 3D Model, 2D Model, Multiple angle, VGG16, Resnet50, Inception Net V3, HaarCascade ,Training, Testing, MTCNN.

Table of Contents

List of Figures

List of Tables

Chapter 1

Introduction

Face recognition is a biometric process where a face can be detected by using different algorithms.Since face recognition is a desirable biometric modality, the field of biometric research has given it a lot of attention also due to the large range of potential application domains, several face recognition algorithms have been developed[12]In this research, angular faces are detected which refers to a face from a different angle. The four stages of the conventional face recognition pipeline are face identification, face alignment, feature extraction, and classification[7] Verification and identification are its two primary tasks. This method aids in the detection of moving persons at any angle.The biometric security checking method is extremely time-consuming in our garment industries and airports, or inspecting each individual is a laborious process. From this study, we demonstrate how to identify a moving person's face quickly. People do not need to wait in line for hours to get verified. The smooth movement of a person's face makes identification simpler and time saving. In the past 15 years, research has concentrated on developing techniques for fully automating face identification systems by learning how to detect faces in images or videos by extracting facial features like the eyes, lips, and other features[2]. Nowadays, facial recognition technologies are used in every industry to protect privacy or for other purposes. It is far more significant for all industries.That's why day by day this sector and research developed gradually. As a result, this study developed a workable method for everyone to recognize the person. One essential element for data analysis of face recognition is precise data on how well humans extrapolate to unique viewpoints[1].In previous literature, 2D frontal facial recognition methods were used which only identifies persons in standstill position ,not in motion. Thus, the concept of multiple angular face detection was generated in this paper. It demonstrates how deep learning methods may be used to accomplish facial recognition from various angular deviations.

1.1 Thoughts behind the Research

In this age of technological advancement, new levels of risks in the security system have increased and with that, the demand for individual identification is growing at the same pace. Most frequently in systems for monitoring or surveillance, homes, or any institution that have to go through with real-time human identification from photos or videos.It is very difficult to identify a moving person from various kinds of software. In the previous research almost all suggested frontal face detection which is the limitation for moving people detection. As a result, we develop our own dataset which consists of people with 10 degree variation and 5 degree variation.This dataset will limit the limitation. After taking the dataset firstly we train the model with the dataset of 10 degree variation,classify that data set and test the dataset with people of 5 degree variation .After that they compare that dataset to present condition for identification.In this process there are two stages recognition phase and testing those data set by comparing with the vast of angular data set. We trained the following model VGG16, VGG19, RESNET50, Inception Net for our dataset. Here, Haarcascade and MTCNN algorithm used to detect the face.The paper is organized as follows. Section 1 introduces the development of angular deviation with 180 degree and gives the purpose of this paper. An overview of the problem statement and our aims.Section 2 to is literature review. In Section 3, working process methodology.Section 4 discussed the result.

1.2 Research Objectives

This research aims to develop a suitable dataset in multiple angular deviations which helps human beings to recognize an individual face from different angles.

- In this paper we demonstrate a method for correctly identifying faces even in the presence of moving objects and various angle aberrations.Principal objective of this study is to ascertain moving objects from several angles in order to design a good face detect model.
- In this paper the accuracies were compared by using a variety of models. Moreover, It's also crucial to comprehend a few characteristics in order to find more precise results. This paper is expected to be a good resource for the future work in the sector of recognizing angular Images.

1.3 Problem Statement

Face recognition is becoming a key component of security, access to personal data, enhanced human-machine interaction, and targeted advertising[29]. Detect a single face from a very large crowd or a massive rush is very difficult. Sometimes it is too time consuming which really affects people in working places. People are in a hurry during office hours. Also, it is a waste of their time when the attendance of the office counts manually one by one. Many are late for the calculation which results in deduction of that day's salary. Therefore, it is very necessary to develop a quick process that can help to identify a single face from a different angle. In our research, we proposed a technique that enables accurate face identification even when people are moving and have varied angular deviations. In addition, This method also guarantees excellent accuracy.

Chapter 2

Related Work

This paper presents a unique face segmentation approach to enhance face identification with wide stance angles and little training data[14]. As the population has grown, human face identification has become more difficult. This is very concerning given the focus on global security and crime that is currently prevalent. Due to different posture angles and occlusions, segmenting the face can be difficult in uncontrolled applications like cross-border security and surveillance. Instead of using 3D modeling, a variable reference angle (VRA) is created to enhance profile face segmentation, which causes less distortion when the database and input face photographs are aligned. Therefore, categorization significantly decreased intraclass variance. Besides, in this study, while using less frontal training data by including a range of angle face data VRA significantly increased face recognition accuracy. Also, over the half-face approach, VRA keeps more information and SVM is unable to accurately classify and scale with additional training samples. Fisher performs somewhat better than Eigen when applying VRA to several training sets. Furthermore, it is shown that while Eigen is unaffected by the segmentation technique, Fisher suffers when dealing with structured data. So, when the VRA segmentation approach was applied, the research problem was effectively solved, and the results were greatly improved.

This study implies that as society progresses, a deep learning-based multi-angle face recognition algorithm will gain the abilities required for high-precision face identification [22]. Face identification and recognition have advanced greatly as a result of this benefit of convolutional neural networks.A facial recognition software system should be able to distinguish a face in a photo efficiently. It asks for first splitting out its qualities, then classifying it regardless of the pose, brightness, mood, illuminating, maturing, and transformations of the snapshot (such as translation, rotation, and scaling)[5].Numerous eminent experts on a national and worldwide scale conducted substantial research on ways to improve deep learning face detection algorithms, leading to the invention of approaches like R-CNN, Fast R-CNN, and Faster R-CNN, among others, that have numerous benefits. In this research, they described a two-layer network as a crucial part of the convolutional neural network-based face identification system. The very first layer network is utilized to determine the initial large-scale placement of the face, and the second layer network mostly depends on that initial positioning to further realize the precise evaluation of the face. The R-FCN method has been extensively used in face identification, and it has yielded some outstanding results. In order to increase the generalizability of image feature extraction and widen the scope of feature perception, the ResNet architecture is mostly used in R-FCN for feature extraction by establishing a deep network. The feature picture obtained using the ResNet model and the anchor approach are used by RPN to produce ROI. The third and fourth convolutional layers of the R-FCN-FACE method's visual data are combined, and the extracted information mode of the convolutional layer of the R-FCN methodology has been adjusted. Using the FDDB and Wider Face two data sets, the suggested methodology is trained, examined, and validated. The AUC values for R-FCN-FACE, R-FCN, and R-FCN-FACE-A are 97percent, 95.8 percent, and 95.1percent, respectively, according to the article. In the detection of faces with many angles and several expressions, this work has produced positive detection results. Even though the R-FCN algorithm isn't specifically intended for face detection, this approach enhances the RFCN technique and proposes the R-FCN-FACE algorithm by using a multi-scale training regime and an online demanding sample mining strategy.

In this research, the author proposed a modal biometric system using PCA with a variety of distance-based classification methods to cluster photos, including the least mean of difference[6]. Face recognition for humans is one of the most important areas of biometric study. Biometrics refers to a technique for identifying a person based on their distinctive physical or behavioral traits. Utilizing biometric authentication for human face recognition has several benefits. The first reason is internet network security, the second is public safety, and the third is the high rate of human face recognition in biometric identification.Owing to it, face detection solutions are already being offered to clients beyond smartphones, for instance at sports activities, bands, and airline verification[19]. One of the robust statistical methods used to examine data is principal component analysis (PCA). Principal Component Analysis (PCA) is a dependable technique since the traits retrieved from facial photos are light-insensitive, distinctive, concealed, and helpful for biometric recognition. In comparison to the Euclidean distance method and City Block distance method, the test results show that the Squared Euclidean distance method with 100 percent recognition rate and PCA with mean clustering generated the best results..

The author suggests a trained auto-encoder to build a deep neural network architecture for reliably extracting data for SSPP face description[11]. Single sample power customization is a popular topic in computer vision. It helps us recognize passports, gate IDs, video surveillance, and other real-life scenarios.So, only a couple of individuals require to be detected for the id card system, while a very significant number of individuals might encounter or move through these security cameras.But while humans frequently dismiss new unexplored faces as outsiders, automated biometric technology has failed to deal with this issue[4].This project bridge was offered as a supervised auto-encoder, a new sort of deep architecture building component, in light of deep learning's significant success in computer vision. They start by mapping faces that are distinct from the person's canonical face. Then they make it look as if similar features are shared by the same person. Face identification gets simpler as a consequence of using the supervised auto inputter, which produces features that are resistant to changes in illumination, expression, occlusion, and position. This identification test feature was extracted using a text-supervised auto-encoder.

In this study, the author suggested an operational multi-angle face recognition technique for video sequences [27]. They propagate front face images using the GAN technique, and they localize using 2D facial feature point estimate methodologies. Besides, they address analysis using facial characteristics applied in low-cost visual sensors which is a problem because certain profiles of 70° or 80° will never be successfully identified. They start by taking the average of a 3D model to make an image, and then they use a statistical model to recreate feature mapping. Also utilized is a convolution neural network model (CNN). They use OpenPose to examine specific feature markings. TP-GAN is used to produce the front face. After that, they do extraction using ResNet-29. Their methods enhanced detection efficiency, which boosted identification accuracy by 1.3 times. Researchers took horizontal photographs of a single person from 18 different perspectives in order to evaluate several frontal face construction techniques. The orientation of the image is positive on the right and negative on the left, respectively. The results of this investigation show that the recommended technique outperforms the traditional facial feature-based detection method.

In this research , low-cost desktop embedded computing is used to detect and identify faces from a variety of perspectives and a sizable database.The facial expression, posture, and rotation detection approach proposed in this work is simple and rapid. The intended system's functioning was divided into two stages. Face detection comes first. Back and front of the face were discovered clipped face area from the image using the Viola-Jones method, and the right.One may tell which side of an image is the face by flipping the profile photo.The recognition phase uses the principal component analysis (eigenfaces) technique, which is dependent on database models that have been built and compared with test face images as input to the recognition process. For two sets of the image of the file exchange during the system's testing and training which is individually identified using the FEI database.[6]The trial's results show how effective and reliable the strategy is. Recognise faces has a high accuracy rate of 96 percent, which improves performance in recognition while operating rapidly.In addition, 35 training pictures' identification accuracy is 97.143 percent and average 0.323657s[6]Three steps make up a generic facial recognition (FR) system: detection, feature extraction, and recognition and this method is employed in several features like biometric system ,security system etc.This process shows how we can easily find closest range tested image and quickly find the person using Viola-Jones method.

Image of frontal 2D face building approach from a side face picture with its boundaries is covered by the writers in this article.We all know various kind of application which are dealing with find people from image or videos is very tough.That's why author shows we can identify people from any angle or one sided view.In the process first take input image which must be visible every parts of the face.After that by using Viola-Jones method feature extraction.[8]A human face's traits may be determined based on a specified restriction by using "Haar features," which are common qualities shared by all human faces.It provides nose and mouth coordinates using the Viola-Jones technique. The virtual line dividing the face into two pieces may be marked by taking the center point of the nose and mouth using the "Haar feature." Changing the virtual line's angle such that it is not exactly 90 degrees is possible if the face orientation is not extremely straight.Select the best piece that has a clear representation of a human face since it will help create a 2D face.Finally listed five images which are matched and identified. Author's target is to increase facial recognition's rate of success when used to identify people. Here, the face's angle is crucial to a successful outcome.

Turk and Pentland's [9] proposal was to use PCA to recognize human faces was discussed by Yang and Hang a typical instance of facial detection based on prior knowledge in 1994.Active appearance models (AAM) and shape models (ASM) (AAM),using two traditional techniques for deformable template.To find human activity, Support Vector Machine (SVM) was utilized.face to identify face and non-face photos initially by Osuna et al. In 2001, Viola and Jones[8] introduced an Adaboost algothe detection of faces and non-faces.The first step of Adaboost algorithm is detect front face and after that classifier fusion design those front face profile and detect multi-angle face.When matched the human face sometimes discover fake faces. By using DP-Adaboost algorithm we can drop those fake images.There is two benefit found which is two eyes may be distinguished, and the highest value in the rise in the horizontal differential projection picture results from weighted approach.Another is erroneous detection brought on by the issue that the two eyes do not lock onto the horizontal axis.But this problem has a solution which is improved method.For the dataset this research paper use (MIT) face library and face recognition technology (FERET) face library including face and non-face images as training samples which size is (20*20).In the end we can say that the author shows overlap separation technique to detect multi-angle faces is a novel approach to classifier fusion presented in this study and performance is quite better.[9]

Now-a-days face recognition improve a lot in different places which we ca as a security system purpose or ticket showing or working place.We can do this thing with long face angle method.Previously people try to established or trying to work with small face angle but author trying to show us long face angle face with multi view.Author use MFF framework and create large number of data set.This MFF framework with pixel intensity and the 3DMM prior parameters to the BFM.[21]

IN our facial components getting side view image is very difficult.That's why author trying to show us how we can get various kind of angle using machine learning ang VGG16 method.VGG16 method is very important method of neural network which helps ImageNet,LeNet dataset to calculate accuracy.VGG16 has fixed size 224 x 244 And its accuracy 92.7 percent [31].After that we simplify our dataset using Haar cascade classifier.This classifier works on different region and implement their three feature with help of pixel to find rectangular regions in a detection window.After that we calculate integral image.Next train this dataset features for a window of 24 x 24 pixels, by ADABOOST training software[31].We can conclude that the accuracy dropped during the testing on low contrast or dark images when compared to the normal images.

In this paper author proposed that an efficient and lightweight deep convolutional network which is inception network.This network is designed to decrease the depth and width of the state-of-the-art networks while maintaining the high-performance. Basically this is use processing in high challenging and multimedia data like video.And if u extract data from video it will also help.Inception architecture has two levels of Residual-Inception combination and successfully tested on two different multimedia.But this model doesn't fulfill all criteria.That's why author claims that this is a disaster network which is input is 224 $\times$ 224 images channel-wise (pixel) mean is used instead of mean image.[13]The learning rate is set to 0.0001 to train the network slowly and avoid overfitting.After processing we can see that this network does not give actual accuracy it is effectiveness from others network.

In this study, the researcher introduced the EfficientNetV2 family of faster, more compact neural networks for image recognition[26]. They develop new systems by combining training-aware neural network model search with scaling, which aims to concurrently improve classification performance and parametric efficiency. This experimental test demonstrates that EfficientNetV2 models trained up to 6.8 times quicker than state-of-the-art methods. In order to increase training performance and effectiveness, they combine scaling and training-aware neural architecture search (NAS). Compared to prior versions, they found EfficientNetV2 trains up to four times faster and also found 84.6 percent accuracy and 19B FLOPs efficiency which performs better than ImageNet, CIFAR, Flowers, and Cars. This model performs better than Inception Net.

Blenz and Vetter created history two decades ago when they proved that 3D facial geometry can be rebuilt from a single photograph[15]. The 3D morphable model, a linear statistical model of 3D spatial form and texture, was used to solve a nonlinear optimization problem in which the whole solution space was limited. Deep Convolutional Neural Networks were utilized to train non-linear mappings from images to model parameters using 3DMMs as generative models. Previous research has proposed both linear and nonlinear 3DMM representations. A principal component analysis is a method for describing and producing 3D faces. Several research has employed principal component analysis to generate statistical 3D shape models. PCA has been used to create large-scale statistical models of the human face and head. As a result, statistical blend shape models that exclusively explain expression variation were created using principal component analysis. Image encoders with DCNNs are known as nonlinear CDMMs. This technique established a new decoder with simply linked layers that can reconstruct every picture using a linear TDMM and unique architectures. These methods ignore the 3D face structure's local geometry. As a result, encoders that use convolutions directly in the non-euclidean face mesh domain should be developed. Geometric deep learning is a subset of deep learning that concentrates on non-Euclidean domains. This model gives an accurate result which is better than the other two models.

In this research the door lock mechanism was integrated with the servomotor, and the signal output was employed to boost the mechanism's operation[24].In today's culture, life security is more important than anything else. That's why we came up with a modern innovation which is door security.Traditional door locks offer excellent security but are inconvenient since they require a key to open. As a result, we looked into ways to improve this type of door lock.This door lock will work by biological characteristic recognition.Contact recognition is used in a wide range of devices.It has the drawback of requiring people to approach the system in order to be detected, which is cumbersome.the first method of this model is Face recognition and data collection after that training model testing and get result.Second method will be Electro-mechanical integration between app and Raspberry Pi which is app processing and interface and connect them with pi.Last method will be Door lock mechanism. The ResNet50 model also gives accurate results.

This paper says how we easily can Identify biometric system face recognition (retina, fingerprint and testing)using VGG16 model.It has four four stage face detection, alignment , extraction and classification.This process slightly various from different background position or posture.VGG16 plays here important part which is validate image for classification and then analysis.This model did not need any kind extra method to extract the feature it has build up method which helps doing the part.And good part is it has large number of parameter which can train large number of data set.ImageNet help this model to evaluate and recognition data set.VGG16 layers kernel followed by max-pooling.VGG16's one layer from the 64 filters convolutional layers is removed and the 256 filters and 512 filters layers are completely removed.After removing it becomes right and compact[16].Second convolutional layers have 128 filters with 3 x 3 filter size followed by max pooling layer with 2 x 2 filtersize.Final layer has two connected layers which is 512 neurons and 30 neurons.The researcher use 30 labels and consists of 7,250 face images with 5,075 images for training and 2,175 images for validation**15**.After VGG16 model train author get accuracy 94.5 percent. Finally, this model has a better accuracy rate than other models.[16]

VGG-19 is a convolutional neural network which uses 19 layers on the basis of multiple image samples and applied architectural style for Zero-Center normalization on Image Convolution, ReLu, Max Pooling, Convolution etc. This (CNN) architecture utilizes deep learning[25]. In this model Zero-Center normalization is the Centralization and Normalization. In the process of Convolutions generalizes and decreases dimension to keep measure centralized. Normality condition is important because this brings everything down in line in normalized, streamlined and orderly fashion. In order to have a pre-definite boundary we need a normal structure of what we are parsing to be normalized and centralized.

Chapter 3

Methodology

3.1 Working process

Firstly, image acquisition is applied throughout this study. In this step, multiple angles of individual images were captured and a total of 19 images for a single face were captured. Around 174 people's pictures were collected at a 10 degree angle for training and 5 degree angle for testing.Then in the pre-processing stage of the image we try to eliminate the background or any kind of distortions that affect the image identification. For this, an input image is needed with 224*224 resolutions. During this research, several models like VGG16, VGG19, ResNet50 and InceptionNet are used. After preprocessing at 10 degree intervals data are trained in different models . Moreover, at 5 degree intervals data is also tested in different models.
Followed by, accuracy for this system will be verified. On the result of the accuracy rate, the best method for a better outcome will be applied. Finally, we will be approaching a deep learning-based method for image classification.

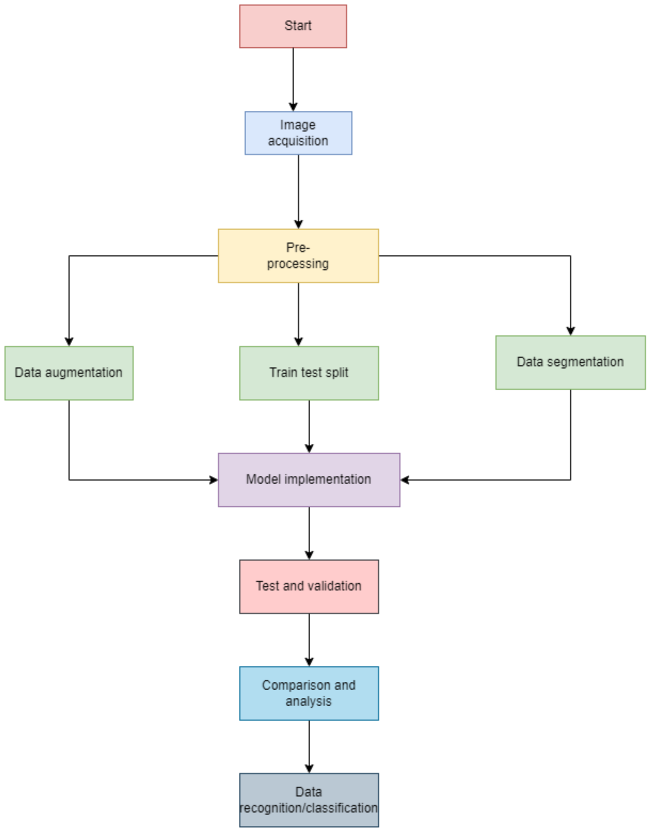


Figure 3.1: Workflow

Followed by, we will verify our model for accuracy. On the result of the accuracy rate, we will apply the best method for a better outcome. Finally, we will be approaching a deep learning-based method for our image classification.
The workflow diagram in Figure 3.1 below shows an overview of each step we will take to train and validate our models

3.2 Used Architectures

This research uses four architectures, which are VGG-19,VGG-16, ResNet 50, and Inception Net V3w.

3.2.1 Visual Geometry Group 16 (VGG 16)

VGG16 consists of 13 convolutional layers, 5 pooling layers, and 3 fully connected layers[18].The fixed size of input to the VGG16 model is 224 by 244.This input image has to deal with the convolution layers of 1 and 2 of 64 channels of 3 by 3 kernel with padding One, stride One. The dimension of this is 224 by 224 and 64. The output of max pooling is 2 by 2 pixel window with stride 2.Again, max pooling reduces the size of the image to half. The output of size 56 x 56 x 128 is given to layers 4,5 and 6 of 256 channels and max pooling is carried out.[11] This gives an output of 28 x 28 x 256 which is given to layers 8,9 and 10 of 512 channels. Max pooling is again carried out which produces a 14 x 14 x [11] 512 which is given to the convolution layers 11,12 and 13 of the 512 channel. Max pooling is again to reduce the size to half and finally produces an image of size 7 x 7 x 512.

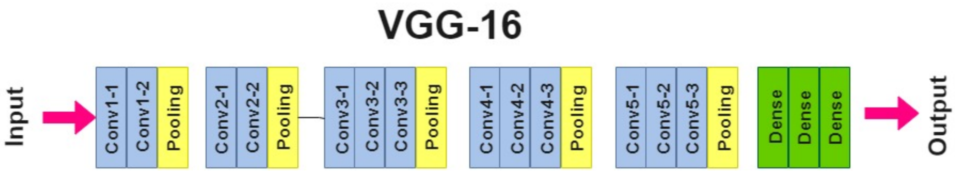


Figure 3.2: Internal Architecture of Vgg-16

3.2.2 Visual Geometry Group 19 (VGG 19)

VGG-19 is a 19 layer convolution neural network.First this model trains over a million pictures from the IMageNet database which is a pre-trained version. It utilized just 3v3 filters with stride and pad of 1 as well as 2*2 max-pooling layers with stride of 2.And error rate 3.3. In second position 138 parameters were included in classification another is localization.This network is working and processing using ImageNet.

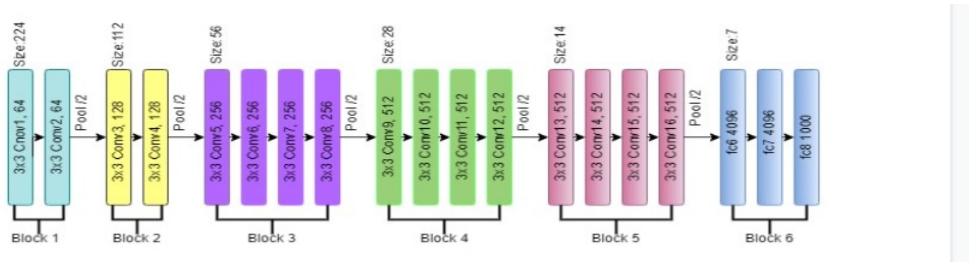


Figure 3.3: Internal Architecture of Vgg-19

3.2.3 ResNet50

ResNet-50 resolves the vanishing gradient issue, making it possible to train insanely deep neural networks quickly. The framework of a skip link is initially illustrated in the model in figure 3.3.

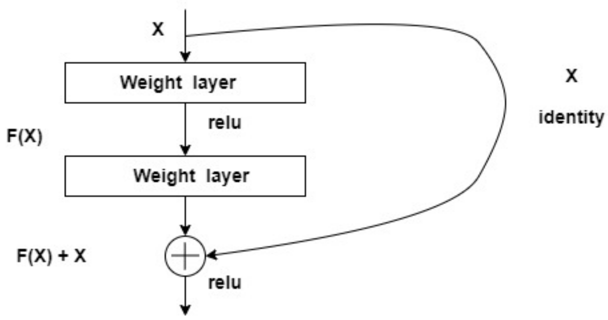


Figure 3.4: Residual Learning

Since a layer's output may be communicated to both the local layer and a remote layer, this approach shows that all levels operate at the same power level.

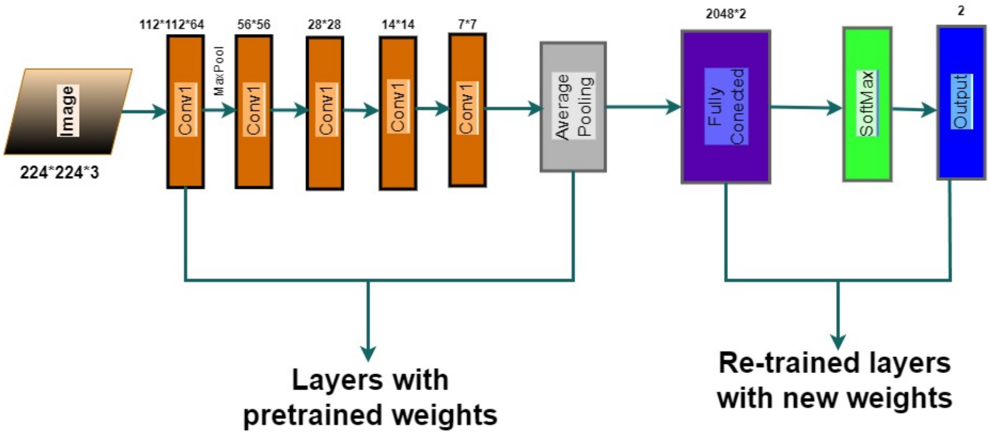


Figure 3.5: Internal Architecture of ResNet50

This model is formed by five steps, each with a unique configuration of convolutional layers.
EfficientNetV2 advises a lesser expansion ratio for MBConv because smaller expansion ratios frequently have lower memory access costs[26].

3.2.4 InceptionNetV3

Convolutional neural networks are the foundation of the deep learning algorithm known as Inception Net, which is used to categorize images.On the ImageNet dataset, it has been demonstrated that the image authentication method Inception Net can achieve greater than 78.1 percentefficiency. Convolutional neural network Googlenet created Inception Net as a plugin to help with object detection and picture analysis.The main goal of Inception Net is to use less processing power by modifying the earlier Inception structures[17].A suggested inception model by Inception Net combines many convolutional filters of various sizes into a single filter. The addition of an auxiliary classifier and the use of Label Correction factorized 7 x 7 convolutions both represent advancements over the earlier members of the Inception family.

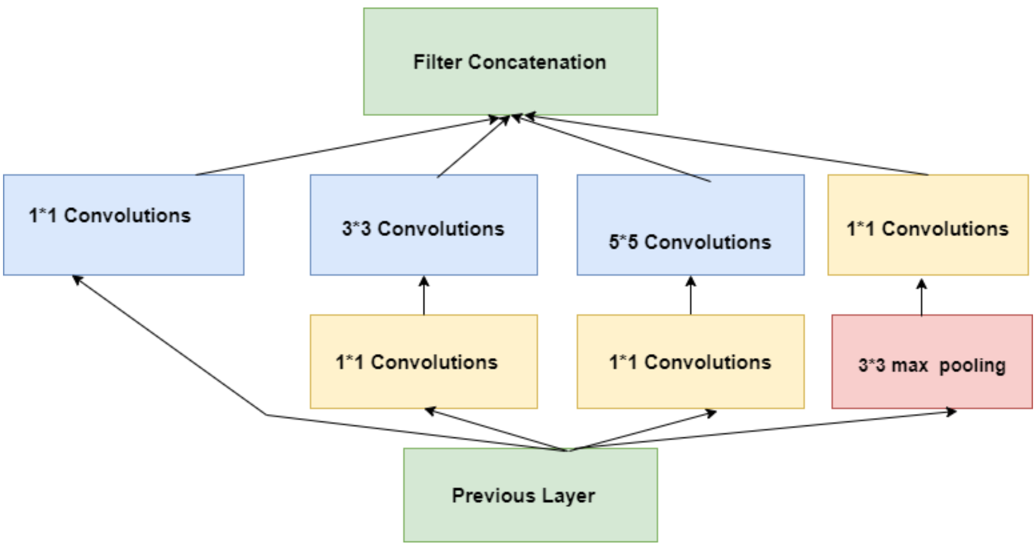


Figure 3.6: Internal Architecture of Inception net

efficiency. Convolutional neural network Googlenet created Inception Net as a plugin to help with object detection and picture analysis. The main goal of Inception Net is to use less processing power by modifying the earlier Inception structures[10]. A suggested inception model by Inception Net combines many convolutional filters of various sizes into a single filter. The addition of an auxiliary classifier and the use of Label Correction factorized 7 x 7 convolutions both represent advancements over the earlier members of the Inception family.

Chapter 4

Implementation

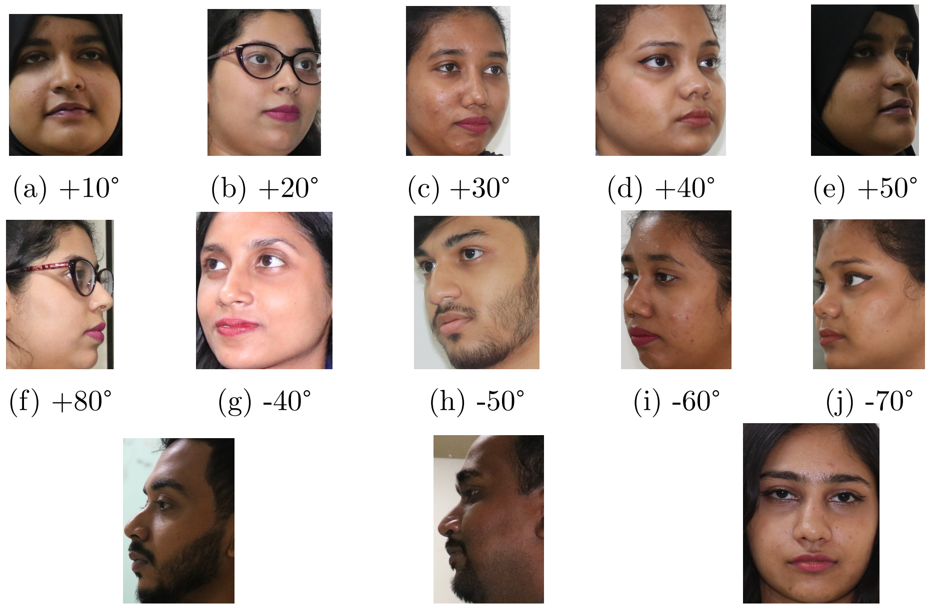
4.1 Dataset

Our dataset consists of 3914 images of the 174 human faces from different angles.

4.1.1 Source

We collected our data by ourselves. Our friends and family helped to collect data by allowing us to take their pictures. The images were taken from 0 to 180 degrees at a distance of 10 degrees. And some extended data were taken from 0 to 180 degrees at a distance of 5 degree for testing. At 10 degree deviation there are 19 images in total for each person and at 5 degree deviation there are 37 images for each person.
Here is the source of our dataset which has been used in our thesis..
shorturl.at/adjlT

4.1.2 Data Sample



(a) +10° (b) +20° (c) +30° (d) +40° (e) +50°

(f) +80° (g) -40° (h) -50° (i) -60° (j) -70°

(k) -80° (l) -90° (m) 0°

Figure 4.1: Sample Images

In the above figure, these images are from our dataset. All these sample data we collected from 0° to 180° at 10° intervals. First six images were taken +10°, +20° , + 30°, +40°, +50° and +80° respectively. Last six images were captured at -40°, -50°, -60°, -70°, -80° , -90° and 0° sequentially.

4.1.3 Chart of Data sample

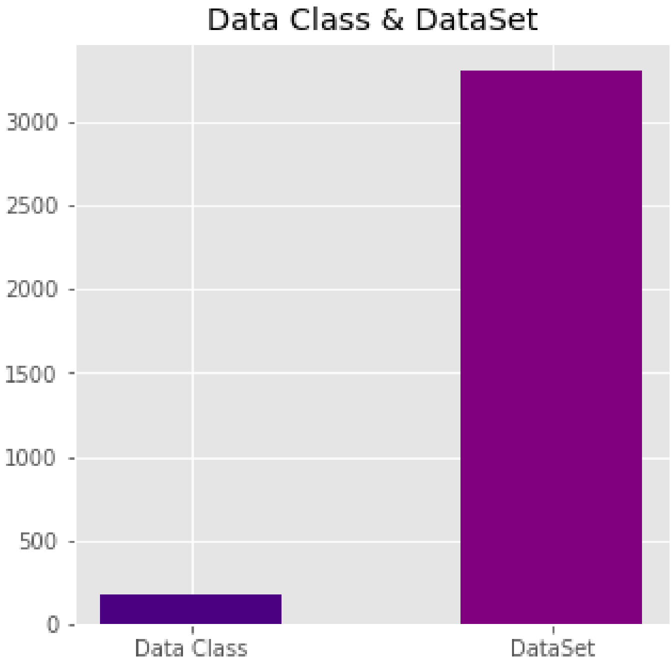


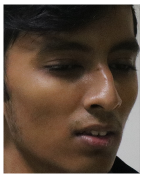
Figure 4.2: Data sample chart

This chart describes the dataset of 10 degree variation where 174 classes of our dataset are labeled as data class and total 19 images for each class which is equal to 3306 data are labeled as dataset

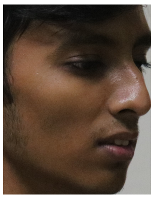
4.1.4 Extended Data Sample


(a) +5°


(b) +15°

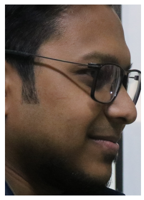

(c) +25°

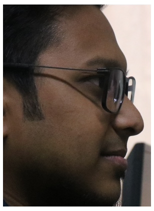

(d) +35°

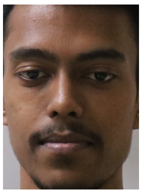

(e) +45°

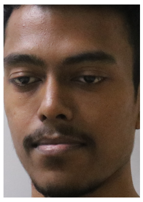

(f) +55°


(g) +65°


(h) +75°

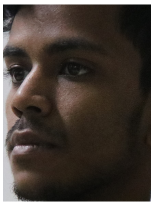

(i) +85°

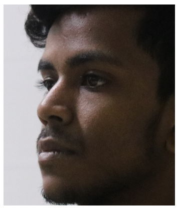

(j) -5°

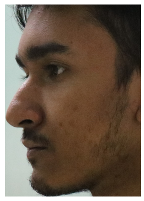

(k) -15°


(l) -25°


(m) -35°


(n) -45°


(o) -55°


(p) -65°


(q) -75°

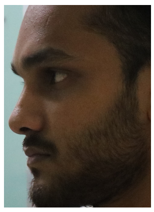

(r) -85°

Figure 4.3: Sample Images

In the above figure, these images are from our data set. All these sample data we collected from 0° to 180° at 5° intervals. First nine images were taken +5°, +15° , +25°, +35°, +45°, +55°, +65°, +75° and +85°. Last nine images were captured at -5°, -15°, -25°, -35°, -45°, -55° , -65°, -75°and -85°.

4.1.5 Chart of Extended data Sample

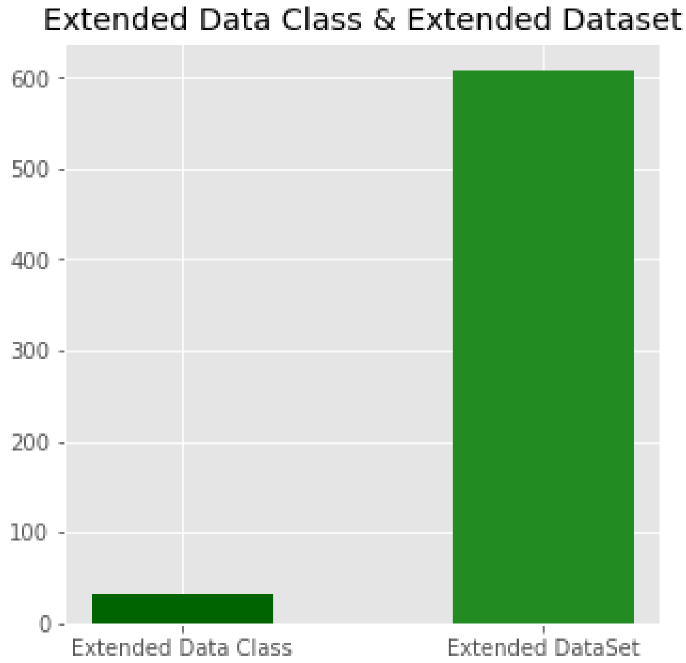


Figure 4.4: Extended data sample chart

This extended chart describes the dataset of 5 degree variation where 32 classes of our dataset are labeled as extended data class and total 19 images for each class which is equal to 608 data are labeled as extended dataset

4.1.6 Data Classification

Model Name	Training Set	Validation Set	Testing Set
VGG16	2314	992	608
VGG19	2314	992	608
ResNet50	2314	992	608
InceptionNet	2314	992	608

Table 4.1: Data Classification

Training Set: The process of feeding samples containing machine learning algorithm output labels or results into the system. into the system. 69.32 of data percentage uses as our training data set

Testing Set: The algorithm may learn certain features of the training set as it converges to improve performance by using a succession of real-world samples. Positive results will boost the algorithm's credibility in the real world for an unidentified test collection.16.11 of data percentage uses as our testing set

Validation Set: A sample of data is utilized to offer an indication when modifying the model's hyperparameters. unbiased evaluation of model fit on training data.After competence when the selected model takes into account the validation dataset, the evaluation changes.21.30 percentage of data uses as our validation set

4.2 Data Pre-processing

The development of appropriate measurements for similarity or dissimilarity is a crucial first step in face recognition. Human faces differ depending on lighting, viewing angle, position, facial expression, occlusion (such as glass and facial hair), face aging and this makes it difficult to identify a human face, so a reliable measure ought to be able to capture as much variation across several topics as is practical while remaining unaffected by variations unrelated to the subjects[3]. Pre-processing is performed to improve the image's quality so that we can perform a more in-depth analysis of it.While sending photos to the classifier, the Keras image module provides tools for integrating and quickly pre-processing them [30]. By using the load img() method an image file downloaded as a PIL image and this downloaded image transformed into a NumPy array. Images that have undergone the RGB to BGR conversion are given back as a NumPy array and the classifier-based observed values' labels are estimated using the model prediction.Remove the Red marked part, we used RGB images (all the 3 channels) and didn't convert it to grayscale. The images were resized to 128x128, and the pixel values of the images were divided by 255 to make the values fall between a range of 0 to 1 (floating point values) because we need to use floating values when training the models to improve the precision and other calculations etc.

4.2.1 Creation of Dataframe:

Data frame has two dimensions: one is row, another is colomn.By using Pandas data frame pre-process dictionary created.Then data frame saved in a google drive so that we can easily run code with the data set.

4.2.2 Face Detection Using Haar Cascade algorithm:

Cascade function doing important work in this process.This function detects the face from the image after that opens it OpenCV.when it detects automatically it creates a shape around the face. pseudocode of haarcascade algorthim is given below

```
import cv2
import numpy as np
face_classifier=cv2.CascadeClassifier('/haarcascade_frontalface_default.xml')
gray = cv2.cvtColor(resized, cv2.COLOR_BGR2GRAY)
faces = face_classifier.detectMultiScale(gray, 1.0485258, 6)
if faces is ():
print("No faces found")
for (x,y,w,h) in faces:
cv2.rectangle(resized, (x,y), (x+w,y+h), (127,0,255), 2)
cv2.imshow('Face Detection', resized)
```

```
cv2.waitKey(0)

cv2.destroyAllWindows()
```

4.2.3 Face detection using MTCNN algorithm:

To identify faces in photos and define and articulate them, a neural network known as MTCNN—Multi-Task Cascaded Convolutional Neural Networks—is employed. The MTCNN employs a three-stage convolutional neural network that aids in the localisation of facial landmarks such the eyes, nose, and mouth. This architecture suggests a CNN neural network with strong recognition and quick facial detection abilities. The Proposal Network (P-Net), Refinement Network (R-Net), and Output Network are three cascade networks that CNN used to recognize objects (O-Net). Below is a description of these three classes:

1. **Proposal network (P-Net):** By resizing the image, the image pyramid is constructed in this stage[17]. Through the use of this pyramid, P-Net generates candidate frames and merges the candidate bounding box (NMS).

2. **Refinement Network (R-Net):** This R-Net then receives candidate frames from the P-Net and refines the potential bounding boxes for categorization.

3. **Output Network (O-Net):** This network creates the final classification findings as well as face landmarks.

Following that, MTCNN produced three results: facial landmark localization, bounding box regression, and face classification.

1. **Face classification:** This determines if a face is there or not using the entropy loss function.

2. **Bounding box regression:** The offset between a candidate and its closest neighbor is determined for each candidate window.

3. **Facial landmark localization:** The localization of face landmarks is determined using the loss function, which is Euclidean distance.[28]

The pesudocode for MTCNN algorithm is attached below

```
from mtcnn import MTCNN
import cv2

detector = MTCNN()

img = cv2.imread("img.jpg")
detections = detector.detect_faces(img)

for detection in detections:
 score = detection["confidence"]
if score & amp;amp;gt; 0.90:
x, y, w, h = detection["box"]
```

```
detected_face = img[int(y):int(y+h), int(x):int(x+w)]
keypoints = detection[”keypoints”]
left_eye = keypoints[”left_eye”]
right_eye = keypoints[”right_eye”]
def alignment_procedure(img, left_eye, right_eye):
 left_eye _x, left_eye_y = left_eye
 right_eye_x, right_eye_y = right_eye

if left_eye_y & amp;gt; right_eye_y:
 point_3rd = (right_eye_x, left_eye_y)
direction = -1
else:
point_3rd = (left_eye_x, right_eye_y)
direction = 1

a = distance.findEuclideanDistance(np.array(left_eye), np.array(point_3rd))
b = distance.findEuclideanDistance(np.array(right_eye), np.array(point_3rd))
c = distance.findEuclideanDistance(np.array(right_eye), np.array(left_eye))

if b != 0 and c != 0:

cos_a = (b*b + c*c - a*a)/(2*b*c)
angle = np.arccos(cos_a)
 angle = (angle * 180) / math.pi

if direction == -1:
 angle = 90 - angle

img = Image.fromarray(img)
img = np.array(img.rotate(direction * angle))
return img
```

4.2.4 Testing Face Detection and Recognition Test Dataset:

Features are compared to the training images while the learned data structure is examined. The training picture is evaluated to the sample image once it has been extracted. When a threshold is established, it will be compared to the necessary minimum distance. If it’s lower than the minimum distance it will be classified as not being part of our data source.

Chapter 5

Result and Analysis

5.1 Result Analysis

:In our result analysis, the performance information such as Validation accuracy,recall, precision, and F1 score for each model.Then we evaluated all of these and used for performance measurement.In our first stage, we have implemented the VGG16, VGG19, ResNet50 and InceptionNetV3 architecture and obtained the following results along with their respective graphs. Here, after implementing VGG16, VGG19, ResNet50 and InceptionNetV3 architectures, we achieved an accuracy of 97%, 92%, 98% and 98% respectively.We used Accuracy, Precision, F1 Score and Recall to evaluate the performance of the model.

5.1.1 Confusion Matrix

Confusion matrix is a representation of the classifier model's predicted value and real value in a table-like structure, where all rows denote predicted value and all columns denote real value. All values presented in the table are the number of inputs given to the model for classification. In the confusion matrix, four terms are used: True positive (TP), True Negative (TN), False positive (FP), and False Negative (TN).

True Positive(TP):- All of the values predicted by the model that matches with the real values indicating that the model predicted them correctly.

True Negative(TN):- Although a positive observation was predicted, it turned out to be negative.

False Positive(FP):- The actual outcome confirms the negative prediction.

False Negative(FN):- The study turns out to be positive though projected to be negative.

Accuracy: Accuracy is the proportion of properly categorized data instances over the number of all data instances.[23]

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{5.1}$$

Precision: When all positive results, including those that were wrongly identified, are taken into account, precision is defined as the ratio of correctly recognized positive results to the number of all positive results.[20]

$$Precision = \frac{TP}{TP + FP} \tag{5.2}$$

Recall: Recall can be described as the ratio of the all positive findings that were correctly identified and the total number of samples that should have been identified as positive.[20]

$$Recall = \frac{TP}{TP + FN} \tag{5.3}$$

F1-Score: The harmonic mean or weighted average of precision and recall is called F1-score.[20]

$$F1 - Score = \frac{2 * TP}{2 * TP + FP + FN} \tag{5.4}$$

5.1.2 Performance analysis with the learning curves VGG16

During the VGG16 model training period, the curves were generated which is accuracy and loss curves. After that we ran another CNN model with a batch size of 32 ,learning rate set to 0.00001. We loaded the weights from the model trained on ImageNet dataset using Keras library and we ran the model for 500 epochs. Figure 5.4 shows the training and validation losses of the model and Figure 5.1 shows the training and validation accuracies of the model during the training process. We used Keras Checkpoint to save the model with the highest validation accuracy around 0.8980 and training accuracy around 0.99. However, this model is slightly overfitting.

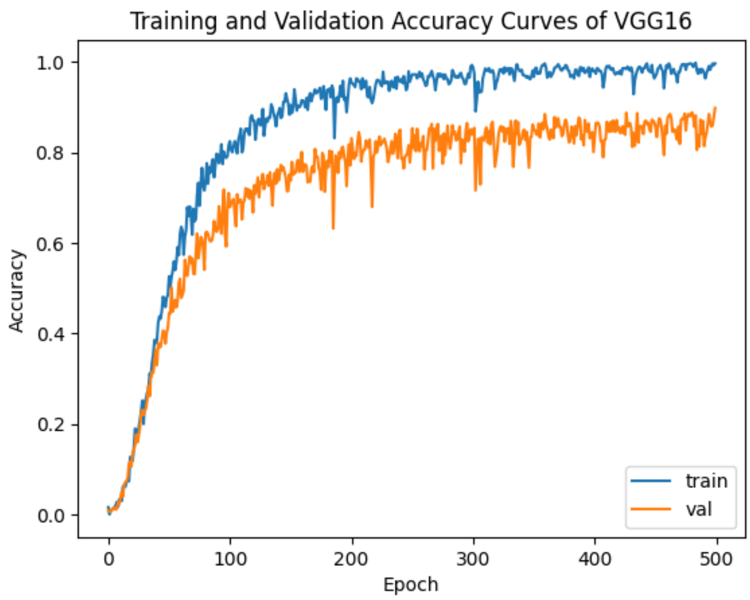


Figure 5.1: Accuracy curve(s) of VGG16 architecture

5.1.3 VGG16 angular prediction analysis:

From this graph we can see, from +15° to +45°, -15° to -25°, it has predicted highest number of person which is 34. From -65° to -85° it has predicted lowest number of person in decreasing order which start from 31.5 and end to 30. And in other degrees, it fluctuates either in increasing order or in decreasing order.

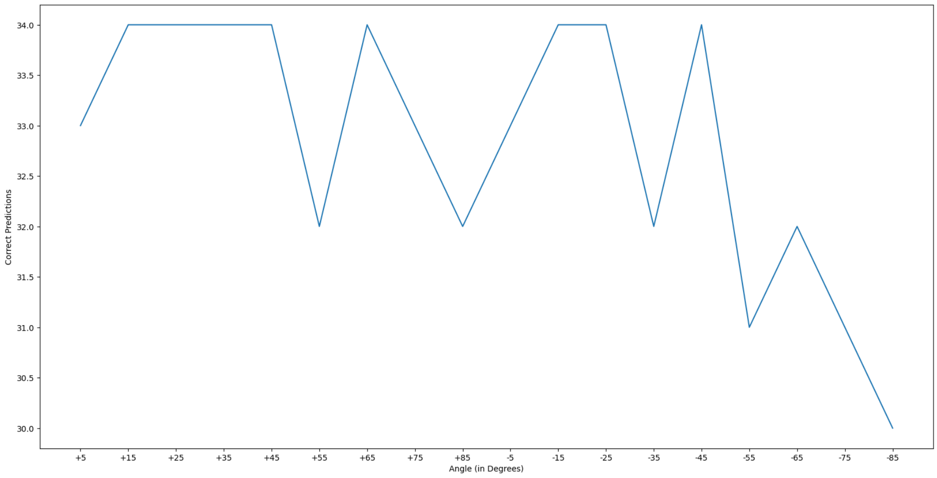


Figure 5.2: Correct predictions of VGG16 on 5 degree testing

5.1.4 Confusion Matrix of VGG16:

After analysing the figure, we can summerize that VGG161 True positive for each class is 18 and in some classes it is 17,16.For each class it does not falsey recognized other person which is true negative.False positive and false negative are 1 in some cases and inmost cases it is quite statisfactory.

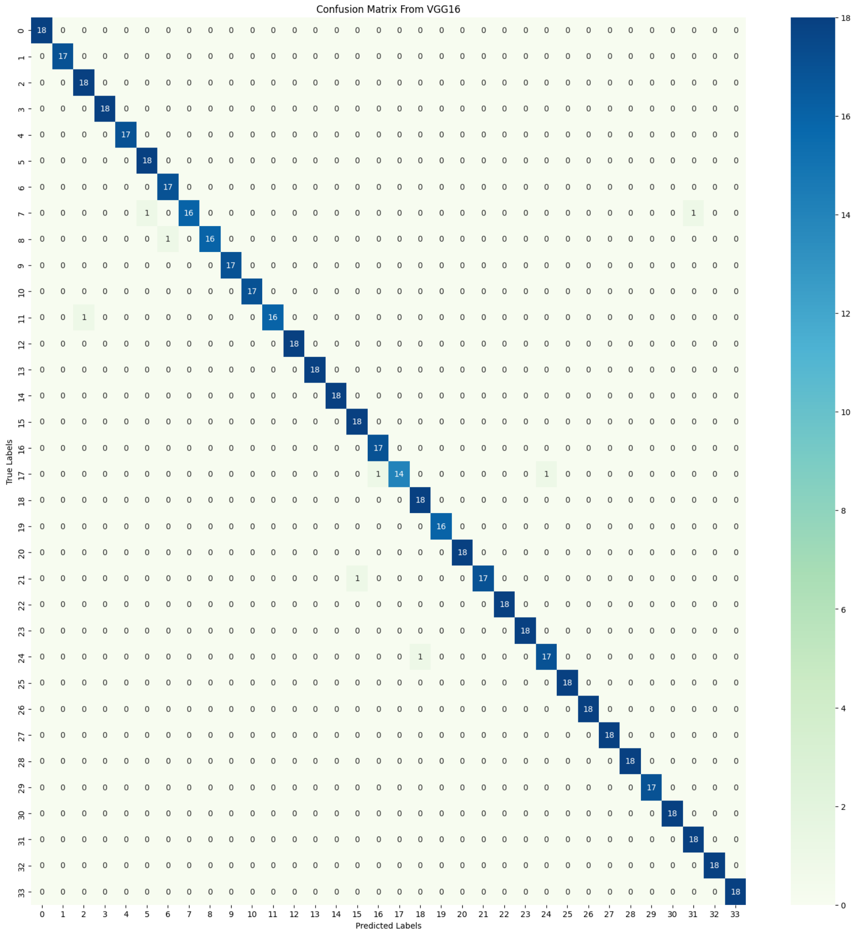


Figure 5.3: Confusion Matrix of VGG16 architecture

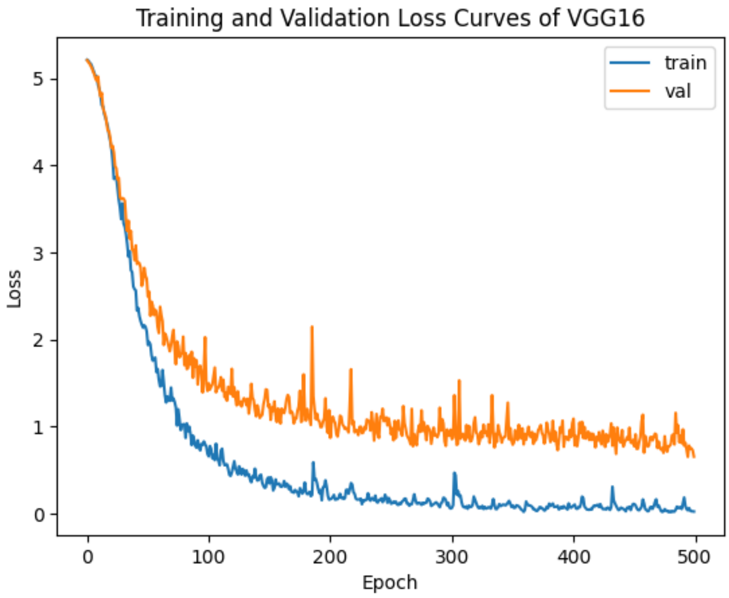


Figure 5.4: Loss curve(s) of VGG16 architecture

From figure 5.4, it is evident that the loss was decreasing gradually from 5.5 to 0.05, from 5.5 to 1.125 for training and validation respectively for 500 epoch.

5.1.5 Performance analysis with the learning curves ResNet50

During the ResNet50 model training period, the curves were generated which is accuracy and loss curves. After that we ran another CNN model with a batch size of 32 ,learning rate set to 0.00001. We loaded the weights from the model trained on ImageNet dataset using Keras library and we ran the model for 500 epochs. Figure 5.5 shows the training and validation accuracies of the model and Figure 5.7 shows the training and validation losses of the model during the training process. We used Keras Checkpoint to save the model with the validation accuracy around 0.85 and training accuracy around 1.00.The accuracy curve of Resnet50 is attached below

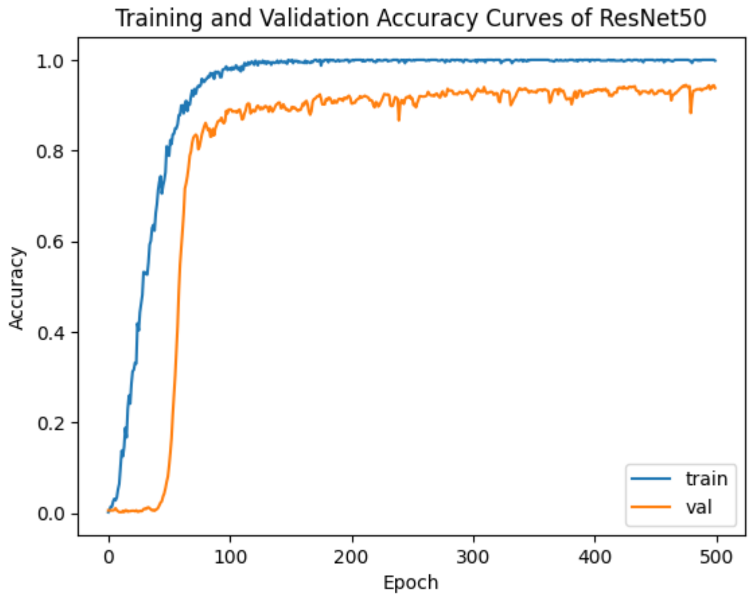


Figure 5.5: Accuracy curve(s) of Resnet50 architecture

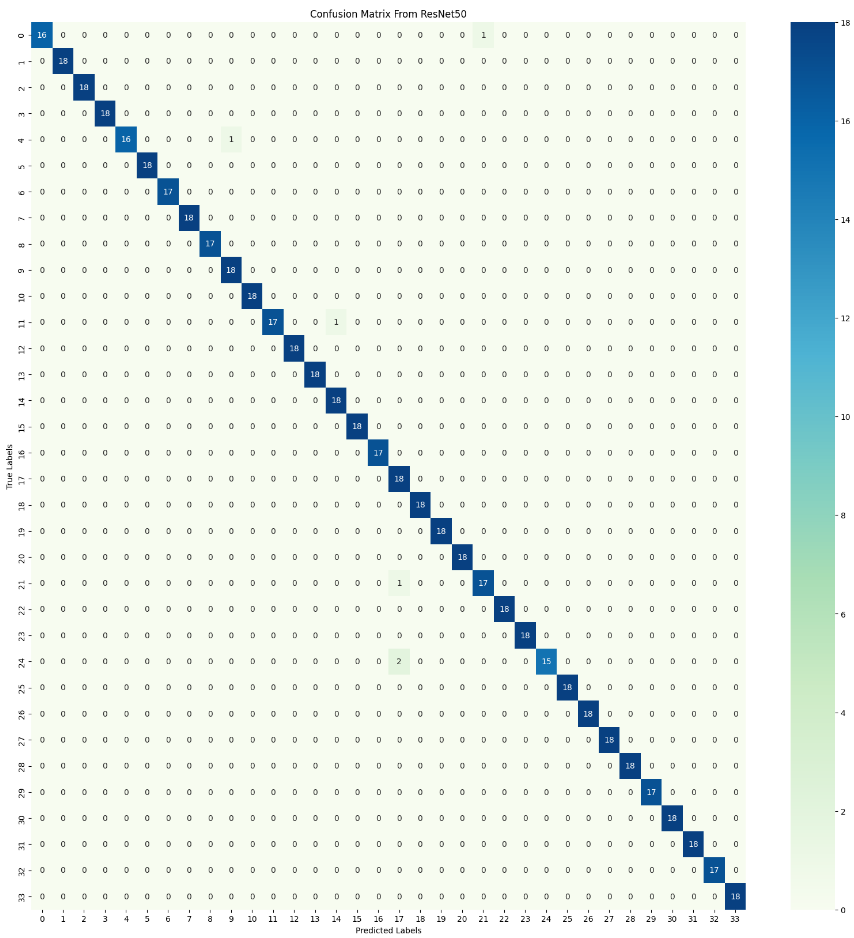


Figure 5.6: Confusion Matrix of ResNet50 architecture

From the above confusion matrix of ResNet50 we can analysis that for almost each class True positive is 18 and in some classes it is 17,16 and 15.True negative is quite satisfied in each class.False positives and false negatives are 1 and 2 in some classes.

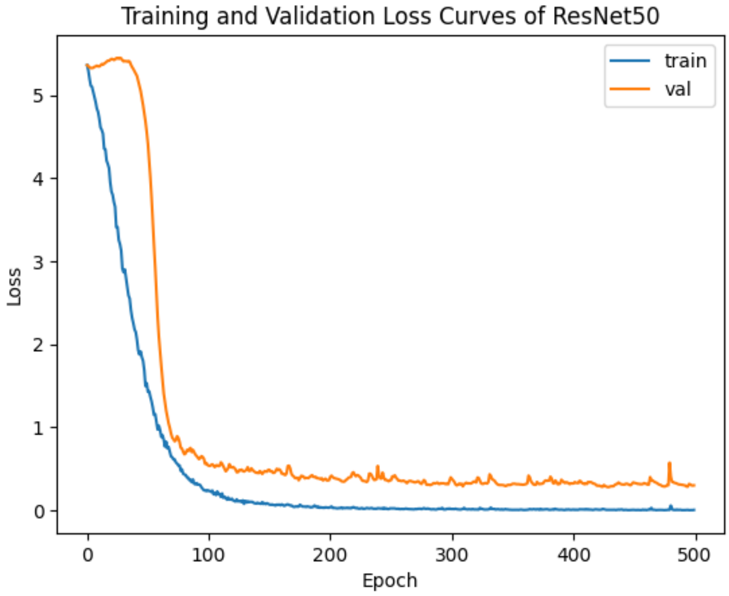


Figure 5.7: loss curve(s) of Resnet50 architecture

After analysisng the above loss curve of ResNet50 we could say that training and validation has reached around 0.00 and 0.40.

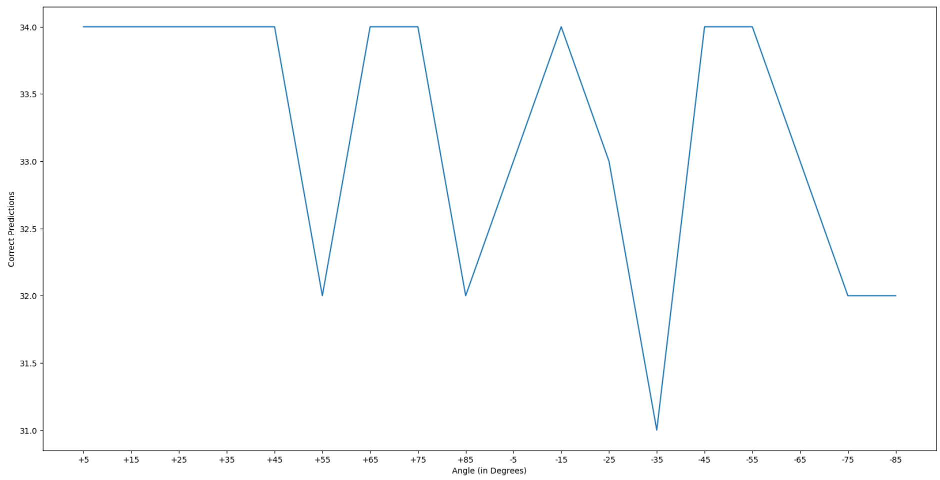


Figure 5.8: Correct predictions of Resnet50 on 5 degree testing

From this ResNet50 angular prediction curve we can see that correct predictions were high from +5 ° to +45 °, +65 ° to +75 ° and -45 ° to -55 ° which is 34.6 and faces in -75 ° to -85 °have the lowest prediction which is 32. In other degrees it is either in increasing or decreasing order.

5.1.6 Performance analysis with the learning curves VGG19

During the VGG19 model training period, the curves were generated which is accuracy and loss curves. After that we ran another CNN model with a batch size of 32 ,learning rate set to 0.00001. We loaded the weights from the model trained on

ImageNet dataset using Keras library and we ran the model for 500 epochs. Figure 5.9 shows the training and validation accuracies of the model and Figure 5.10 shows the training and validation loss of the model during the training process. We used Keras Checkpoint to save the model with the accuracy of training and validation has reached around 0.98 and 0.80.This model has been overfitted.The accuracy curve has been attached below

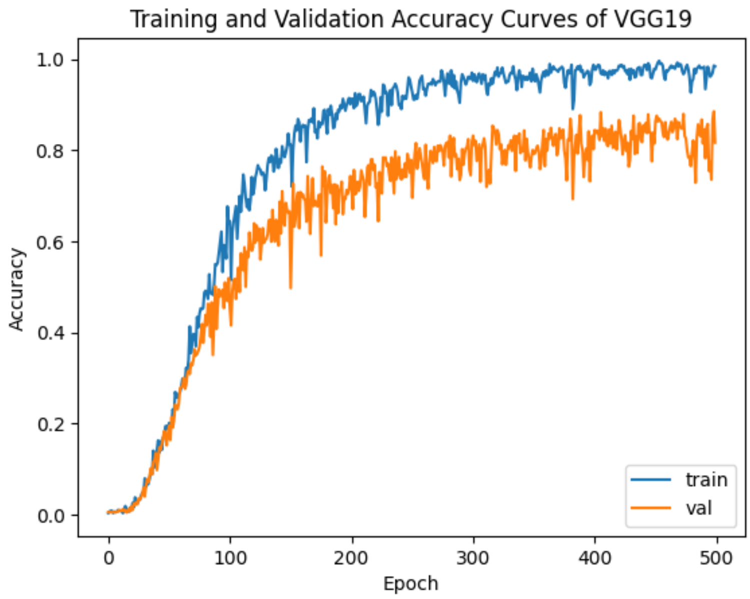


Figure 5.9: Accuracy curve(s) of VGG19 architecture

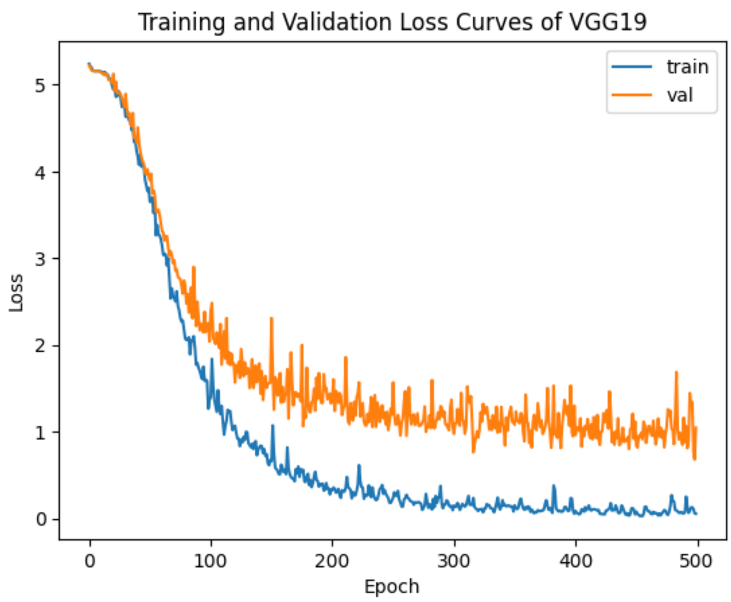


Figure 5.10: loss curve(s) of VGG19 architecture

From the above loss curve of VGG19 it can seen that training and validation has reached around 0.012 and 1.111 . Overfitting took place in this model.

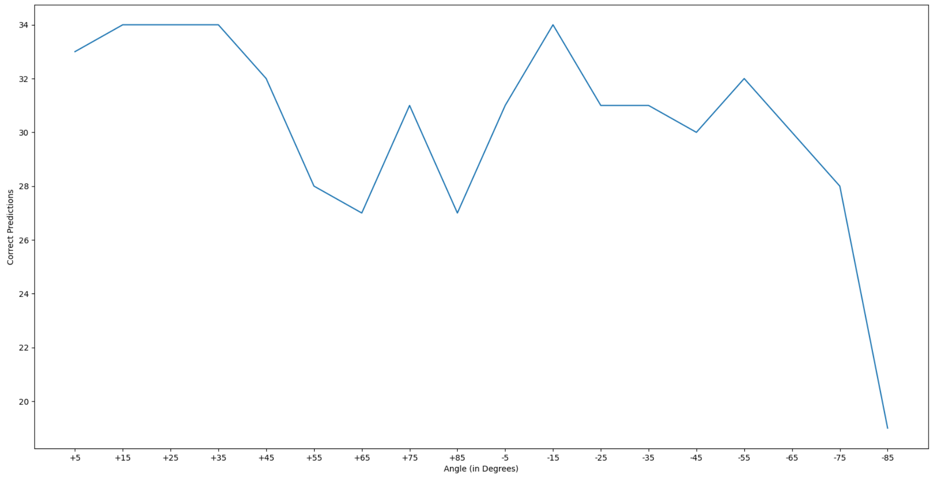


Figure 5.11: Correct predictions of VGG19 on 5 degree testing

From the attached angular prediction range of VGG19 we can see there is lots of fluctuation in this model. The prediction for this model is not steady. The graph is either decreasing or increasing order. The predictions were high from +15 ° to +35 °and -15 ° which is 34 and faces in -85 ° have the lowest prediction which is 20.

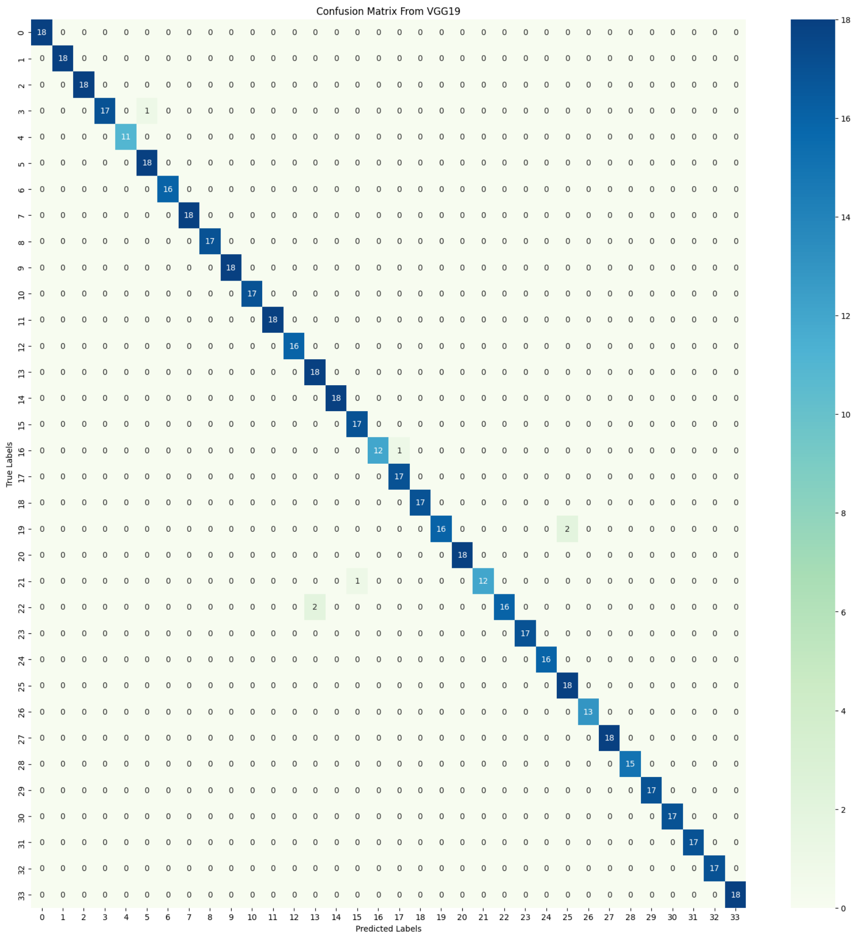


Figure 5.12: Confusion Matrix of VGG19 architecture

In case of confusion matrix of VGG19 in the above for almost each class True positive is 18 and in some classes it is 17,16,12 and 11.True negative is quite satisfied in each class.False positives and false negatives are 2 and 1 in some classes.

5.1.7 Performance analysis with the learning curves InceptionNetV3

After that we ran another CNN model with a batch size of 32 ,learning rate set to 0.00001. We loaded the weights from the model trained on ImageNet dataset using Keras library and we ran the model for 500 epochs. Figure 5.13 shows the training and validation accurcies of the model and Figure 5.14 shows the training and validation losse of the model during the training process. We used Keras Checkpoint to save the model with the validation accuracy of 0.99 and training accuracy of 0.94.

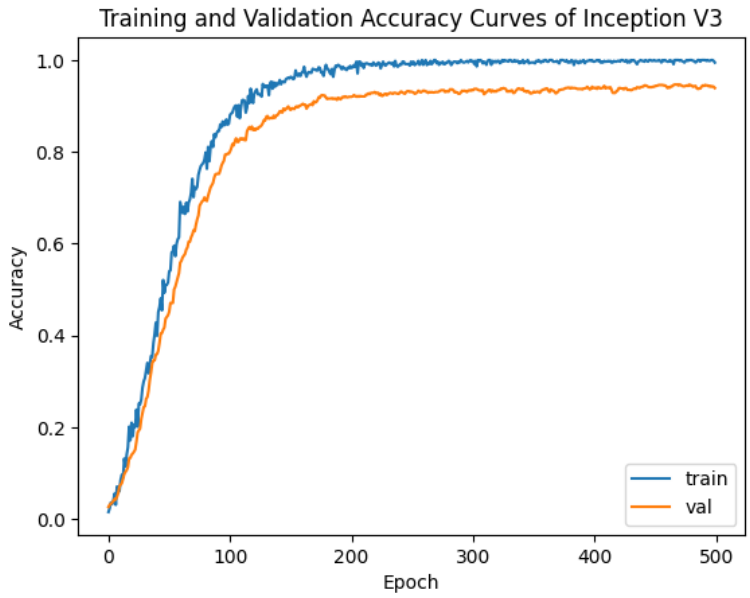


Figure 5.13: Accuracy curve(s) of IneceptionV3 architecture

From this below loss curve of InceptionV3 for training and validation has reached around 0.01 and 0.04 .

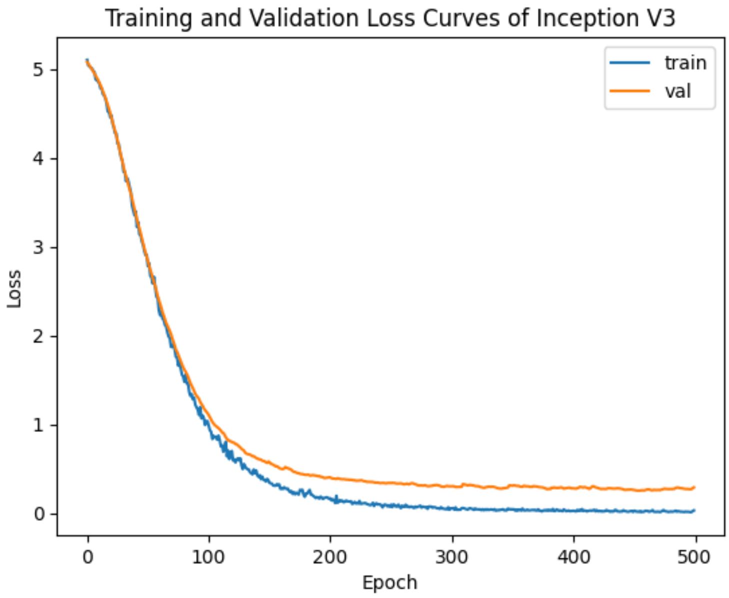


Figure 5.14: loss curve(s) of InceptionNetV3 architecture

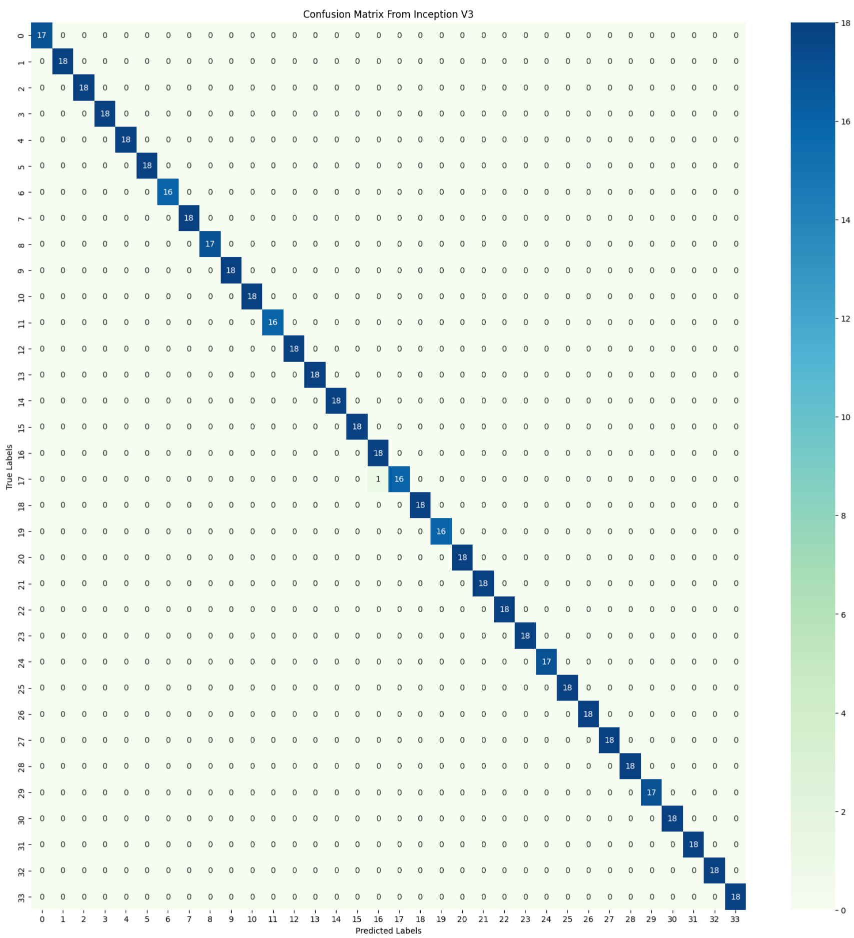


Figure 5.15: Confusion Matrix of InceptionNetV3 architecture

From this degree prediction figure we can see that correct predictions were high from +5 ° to +35 °,+65 ° to +75 ° and -25 ° to -55 ° which is 34 and faces in -65 ° to -75

° have the lowest prediction which is 31.In other degrees it is either decreasing or increasing.

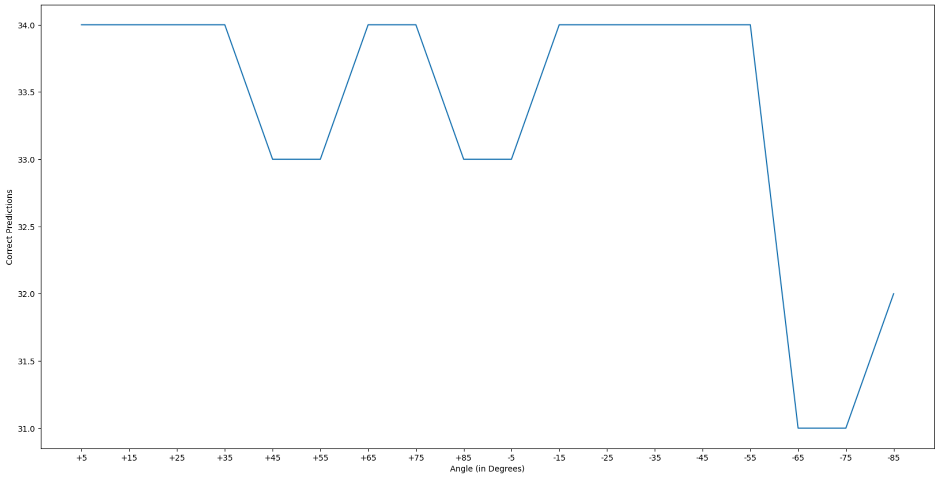


Figure 5.16: Correct predictions of InceptionNetV3 on 5 degree testing

From the above confusion matrix of InceptionV3 for almost each class True positive is 18 and in some classes it is 16 or 17.True negative is quite satisfied in each class.False positives and false negatives are 1 in some classes

5.1.8 Result Comparison

In our research, we have implemented VGG16, ResNet50, InceptionNetV3 and VGG19 architecture on our dataset. After implementation we have received an accuracy of 97%, , precision of 99%, recall of 97% and f1-score of 98% for VGG16 architecture .ResNet50 architecture has obtained accuracy of 98 %,precision of 99 %, recall of 98% and f1-score of 98 % .InceptionV3 architecture has obtained accuracy of 98 %,precision of 100 %, recall of 98% and f1-score of 99 % .VGG19 architecture has obtained accuracy of 92 %,precision of 99 %, recall of 92% and f1-score of 95 % . From the table 5.1, we can determine that ResNet50 and InceptionNetV3 is the

Architecture	Accuracy (%)	Precision (%)	Recall (%)	F1 score (%)
VGG16	97	99	97	98
ResNet50	98	99	98	98
InceptionNetV3	98	100	98	99
VGG19	92	99	92	95

Table 5.1: Comparison table between VGG16, ResNet50, InceptionNeV3 and VGG19 models

best performing architectures among the architectures we have implemented so far.

5.1.9 Ensemble

Esnemble: It refere to a process where it combine two or multiple models.It uses many techniques and algorithms.It takes best result from each model and give more better result. We ensemble our best two models ResNet50 and InceptionV3.After ensembling those two models we can see from this below attached confusion matrix that ensemble does not give us better result .Rather , it gives accuracy as InceptionV3 and ResNet50.So,in this research ensemble doesn not give us benefit.So, we can claim that InceptionV3 is best suited in our system.

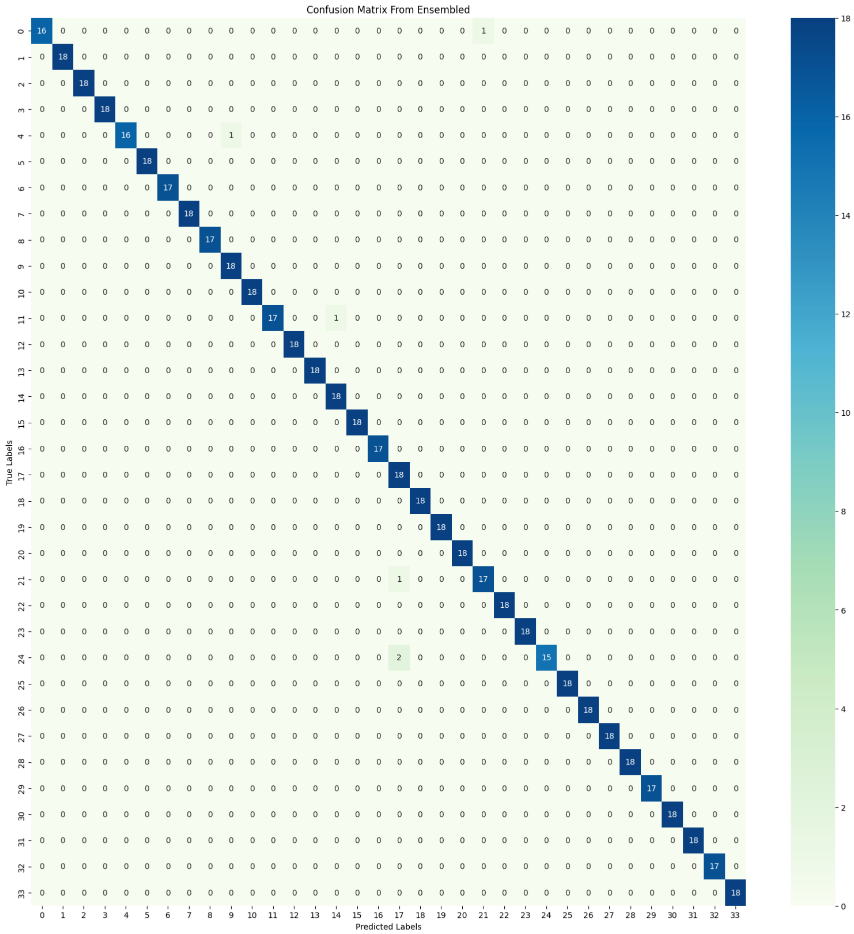


Figure 5.17: Confusion matrix after ensembling ResNet50 and InceptionNetV3

Chapter 6

Conclusion and Future Work

In the research process of face recognition, we train machines to recognize people using sample images. This system is composed of three steps: image accusation, pre processing, and recognition. In the accusation we captured multiple angles of individual images. Our goal was to train the machine to identify a single person from multiple angles. We captured a total of 19 images for a single face at 10 degree intervals and additionally 37 images were taken for some persons at 5 degree intervals to test our system. VGG16,VGG19, ResNet50, InceptionNetV3 models are trained through our data set. We collected 174 people's pictures for a total of 3914. After comparing, we can see that ResNet50 and InceptionNetV3 both give the best accuracy for this research. Because ResNet50 and InceptionNetV3 give 98% accuracy of all the models. So this proposed system can work in dynamic motion accurately. The main advantages of this system is reduced time for detection of a person. Also this study fulfills our motive. In the future, this system face recognition will be highly effective as there is a higher risk of security and a person can easily be detected without any hassle. We aim to create an effective method for detecting crystal clear pictures and a person can easily be detected without any hassle

Bibliography

[1] N. F. Troje and H. H. Bülthoff, "Face recognition under varying poses: The role of texture and shape," *Vision research*, vol. 36, no. 12, pp. 1761–1771, 1996.

[2] J. Fan, S. Upadhye, and A. Worster, "Understanding receiver operating characteristic (roc) curves," *Canadian Journal of Emergency Medicine*, vol. 8, no. 1, pp. 19–20, 2006.

[3] J. Shi, A. Samal, and D. Marx, "How effective are landmarks and their geometry for face recognition?" *Computer vision and image understanding*, vol. 102, no. 2, pp. 117–133, 2006.

[4] B. Kamgar-Parsi, W. Lawson, and B. Kamgar-Parsi, "Toward development of a face recognition system for watchlist surveillance," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 33, no. 10, pp. 1925–1937, 2011.

[5] D. N. Parmar and B. B. Mehta, "Face recognition methods & applications," *arXiv preprint arXiv:1403.0485*, 2014.

[6] S. Gao, Y. Zhang, K. Jia, J. Lu, and Y. Zhang, "Single sample face recognition via learning deep supervised autoencoders," *IEEE transactions on information forensics and security*, vol. 10, no. 10, pp. 2108–2118, 2015.

[7] G. Hu, Y. Yang, D. Yi, *et al.*, "When face recognition meets with deep learning: An evaluation of convolutional neural networks for face recognition," in *Proceedings of the IEEE international conference on computer vision workshops*, 2015, pp. 142–150.

[8] R. Naik and K. Lad, "Face recognition from multi angled images," *International Journal Of Engineering Research & Technology (Ijert) Issn*, pp. 2278–0181, 2015.

[9] Y.-Y. Zheng and J. Yao, "Multi-angle face detection based on dp-adaboost," *International Journal of Automation and Computing*, vol. 12, no. 4, pp. 421–431, 2015.

[10] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, "Rethinking the inception architecture for computer vision," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 2818–2826.

[11] E. I. Abbas, M. E. Safi, and K. S. Rijab, "Face recognition rate using different classifier methods based on pca," in *2017 International Conference on Current Research in Computer Science and Information Technology (ICCIT)*, IEEE, 2017, pp. 37–40.

[12] Z. Mahmood, N. Muhammad, N. Bibi, and T. Ali, "A review on state-of-the-art face recognition approaches," *Fractals*, vol. 25, no. 02, p. 1 750 025, 2017.

[13] S. Pouyanfar, S.-C. Chen, and M.-L. Shyu, "An efficient deep residual-inception network for multimedia classification," in *2017 IEEE International Conference on Multimedia and Expo (ICME)*, IEEE, 2017, pp. 373–378.

[14] D. Brown, "Multi-angled face segmentation and identification using limited data," in *2019 15th International Conference on Signal-Image Technology & Internet-Based Systems (SITIS)*, IEEE, 2019, pp. 98–105.

[15] B. Gecer, S. Ploumpis, I. Kotsia, and S. Zafeiriou, "Ganfit: Generative adversarial network fitting for high fidelity 3d face reconstruction," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2019, pp. 1155–1164.

[16] A. B. Perdana and A. Prahara, "Face recognition using light-convolutional neural networks based on modified vgg16 model," in *2019 International Conference of Computer Science and Information Technology (ICoSNIKOM)*, IEEE, 2019, pp. 1–4.

[17] R. Gradilla, "Multi-task cascaded convolutional networks (mtcnn) for face detection and facial landmark alignment," *link]. Acessado em*, vol. 13, 2020.

[18] Y.-C. Hsieh, C.-L. Chin, C.-S. Wei, I.-M. Chen, P.-Y. Yeh, and R.-J. Tseng, "Combining vgg16, mask r-cnn and inception v3 to identify the benign and malignant of breast microcalcification clusters," in *2020 International Conference on Fuzzy Theory and Its Applications (iFUZZY)*, IEEE, 2020, pp. 1–4.

[19] Y. Kortli, M. Jridi, A. Al Falou, and M. Atri, "Face recognition systems: A survey," *Sensors*, vol. 20, no. 2, p. 342, 2020.

[20] P. Shivaprasad, *Understanding confusion matrix, precision-recall, and f1-score*, Oct. 2020. [Online]. Available: https://towardsdatascience.com/understanding-confusion-matrix-precision-recall-and-f1-score-8061c9270011.

[21] H. Wang, Y. Wang, J. Liu, and G. Ying, "A novel sample-enhanced dataset based on mff for large-angle face recognition," 2020.

[22] Y. Du and Q. Wang, "Multi-angle face detection based on improved rfcn algorithm using multi-scale training," in *2021 6th International Conference on Intelligent Computing and Signal Processing (ICSP)*, IEEE, 2021, pp. 319–322.

[23] N. Harikrishnan, *Confusion matrix, accuracy, precision, recall, f1 score*, 2021.

[24] S. Phawinee, J.-F. Cai, Z.-Y. Guo, H.-Z. Zheng, and G.-C. Chen, "Face recognition in an intelligent door lock with resnet model based on deep learning," *Journal of Intelligent & Fuzzy Systems*, vol. 40, no. 4, pp. 8021–8031, 2021.

[25] M. Tamilselvi, S. Karthikeyan, and G. Ramkumar, "Face recognition based on spatio angular using visual geometric group-19 convolutional neural network," *Annals of the Romanian Society for Cell Biology*, pp. 2131–2138, 2021.

[26] M. Tan and Q. Le, "Efficientnetv2: Smaller models and faster training," in *International Conference on Machine Learning*, PMLR, 2021, pp. 10 096–10 106.

[27] Z. Zhang, H. Zhang, H. Liu, S. Xin, N. Xiao, and L. Zhang, "Frontal face generation based multi-angle face identification system," in *2021 International Conference on Computer, Control and Robotics (ICCCR)*, IEEE, 2021, pp. 329–334.

[28] M. Gu, X. Liu, and J. Feng, "Classroom face detection algorithm based on improved mtcnn," *Signal, Image and Video Processing*, vol. 16, no. 5, pp. 1355–1362, 2022.

[29] S. Mohmmad and B. Rama, "Extended informative local binary patterns (eilbp): A model for image feature extraction," in *Communication, Software and Networks: Proceedings of INDIA 2022*, Springer, 2022, pp. 381–392.

[30] S. Aiswarya, K. Himanush, and S. H. Yedla, "Multi-angle face recognition using deep learning,"

[31] I. Rajendra, P. Nikhil, D. Sumanth, M. Sai Charan, and G. Manohar, "Biometric security for browser privacy,"