Obscure Face Recognition Using Deep Neural Networks

by

MD SHAHADAT HOSSAIN AKASH
19101440
SHADMAN BIN SHARIFE
22241139
BONDON DATTA
19101143

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

Department of Computer Science and Engineering
Brac University
June 2023

Declaration

It is hereby declared that

1. The thesis submitted is 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:

MD Shahadat Hossain Akash

19101440

Shadman Bin Sharife

22241139

Bondon Datta

19101143

Approval

The thesis/project titled "Obscure Face Recognition Using Deep Neural Networks" submitted by

1. Md Shahadat Hossain Akash (19101440)
2. Shadman Bin Sharife (22241139)
3. Bondon Datta (19101143)

Of Spring, 2023 has been accepted as satisfactory in partial fulfillment of the requirement for the degree of B.Sc. in Computer Science on June 04, 2023.

Examining Committee:

Supervisor:
(Member)

Aminul Huq

Lecturer
Department of Computer Science and Engineering
Brac University

Co-Supervisor:
(Member)

Mr. Md. Tanzim Reza

Lecturer
Department of Computer Science and Engineering
Brac University

Thesis Coordinator:
(Member)

Dr. Md. Golam Rabiul Alam

Professor
Department of Computer Science and Engineering
Brac University

Head of Department:
(Chair)

Sadia Hamid Kazi

Chairperson and Associate Professor
Department of Computer Science and Engineering
Brac University

Abstract

One of the most effective biometric tools for securing diverse systems is facial recognition technology. Traditional security mechanisms like PINs, passwords, and fingerprints have shown to be less effective and dependable than this technology. Systems for surveillance, finance, and security have all made substantial use of facial recognition technology. But facial recognition technology is currently facing a huge challenge from the COVID-19 pandemic. Due to face occlusion brought on by the widespread use of masks, it is now challenging to precisely identify people. This problem has motivated a number of academics to develop facial recognition technology that is more accurate by focusing on hidden facial features. In this paper, we provide a sophisticated method for identifying faces even when there are facial alterations or masks are used. The CASIA, VGG2, LFW Databases are used to help us build our technique, which entails figuring out which facial traits are still discernible even when the face is partially hidden. Convolutional neural networks (CNNs), which are the foundation of our method and are used to extract pertinent facial information, are based on deep learning techniques. We contrasted our findings with those of other cutting-edge facial recognition systems to assess our strategy. By obtaining more accuracy and speed, our system outperformed other systems. We got 95.85% accuracy in face recognition with the model InceptionResNetV1 by using ArcFace loss function. Also got 94.28% accuracy on the same model by using Triplet loss function. We worked on the never before worked model for face recognition which is SE-ResNeXt-101. We also got 93.41% accuracy on that model with ArcFace loss function. That indicates that our research underlines the significance of creating environmental change-resistant facial recognition

Keywords: Facial recognition; Face detection; Convolution neural networks; Deep Neural Network; InceptionResNetV1; ArcFace loss function; SE-ResNeXt-101; Triplet loss function

Acknowledgement

First of all, we would like to express our sincere appreciation to Allah for allowing us to finish our thesis without any significant setbacks. Second, we want to sincerely thank our supervisor, Aminul Huq sir, for all of his help and advice during the course of our work. When we needed aid, he was always there to provide it. We also like to express our gratitude to our co-supervisor, Md. Tanzim Reza, for his technical know-how and unflinching support.
Finally, we want to thank our parents from the bottom of our hearts. We would not have been able to achieve this academic milestone without their unfailing support and prayers. We owe their affection and support for getting us to this point—we are about to graduate.

List of Figures

Nomenclature

The next list describes several symbols & abbreviation that will be later used within the body of the document

AI Artificial Intelligence

ANN Artificial Neural Network

$CCTV$ Closed-Circuit Television

CNN Convolutional Neural Network

DNN Deep Neural Networks

ID Identity Document

Ml Machine Learning

MLP Multi-Layer Perceptron

PIN Personal Identification Number

Table of Contents

Chapter 1

Introduction

1.1 Introduction

Face recognition is a biometric system that has advanced rapidly in recent years, thanks to advances in deep neural networks. This cutting-edge technology recognizes and verifies an individual's identity by analyzing their facial features and matching them against a database of faces.
Deep neural networks, which serve as the modern facial recognition system's backbone, are intricately linked with facial recognition technology. An artificial intelligence called deep neural networks is able to learn and recognize patterns in data, including face traits. These networks are made up of multiple layers of artificial neurons that collaborate to analyze and recognize complex patterns in facial data. Deep neural networks are especially well-suited to facial recognition because they can learn from massive amounts of data and adapt to changing lighting conditions, angles, and facial expressions. They are trained on a large dataset of facial images and can accurately identify a person by identifying specific features such as the distance between the eyes or the shape of the nose.
In circumstances when individuals are wearing masks owing to public health concerns such as COVID-19, facial recognition technology has become more significant. Traditional facial recognition algorithms may be unable to reliably identify persons when masks cover a large section of the face. Nonetheless, recent advances in facial recognition technology have enabled the identification of facial characteristics such as the eyes and brows, which can still offer enough information to accurately identify individuals. This is especially crucial in places where security and safety are paramount, such as airports, hospitals, and government facilities. The capacity to reliably identify persons in a disguised face circumstance can aid in preventing security breaches and ensuring public safety. Moreover, facial recognition technology may be used to track adherence to mask regulations and aid in the containment of contagious illnesses. In general, the employment of facial recognition technology in scenarios involving masked faces has evolved into a crucial instrument for upholding safety and security in the contemporary period.
Due to its capacity to distinguish between persons based on their facial expressions and traits, face recognition technology has emerged as a crucial tool in a number of sectors. Its biggest use is in security situations where it can keep an eye on public areas, spot possible dangers, and deter crime. To detect known terrorists and criminals and stop them from entering the nation, it is frequently employed in airports

and other transportation hubs. Access control, personalizing, and healthcare are further applications of the technology. It can recognize clients and offer personalized experiences, detect medical issues, and aid law enforcement in more effectively solving crimes.
Despite its many advantages, there are also worries about how facial recognition technology might be used improperly and how that might affect privacy, which emphasizes the significance of using this technology responsibly and ethically. For instance, there are worries about the use of facial recognition technology in public areas because it may cause people to lose their privacy. The accuracy of facial recognition technology is another issue, especially for people
Our study mainly investigates the use of deep neural networks in facial recognition and tackles issues of accuracy and privacy. In comparison to other articles, we used several models with unique loss functions to reach superior accuracy findings. Our primary goal was to overcome the shortcomings of existing face recognition technology and deliver a more dependable and safe solution.

1.2 Problem Statement

In the past two years, coronavirus has taken almost 29114 people's lives in our country. A virus that damages a lot but the precaution needed to be taken is small. Wearing a mask helps in reducing the chance of getting this deadly virus. But due to security measures, ID verification, mobile phone unlocking, and many more reasons, we can see that people need to or tend to remove their mask for facial verification which ultimately increases the risk of getting infected with the deadly coronavirus. In mobile facial unlocking systems, the device most of the time cannot detect an obscure face or a masked face which also results in the removal of facial masks. When we talk about the harmfulness of taking off the mask, we can also see that many crimes have increased where the criminals use a mask as a cover or other substance to cover their face to not get caught in the CCTV camera. As a result, recognition of the criminal cannot be identified, this leads to a great opportunity for the criminals to move around freely. A report says that from the time span of Jan 1 – Oct 31, the criminal incidents in LA had risen to 1164 which is a 15% increase where the criminal incidents were done by wearing facial masks [5]. Some other problems that can be found prominent are:

1. Since conventional face recognition algorithms rely on clear facial pictures, they might not be able to identify faces that are partially blocked by objects or other impediments.

2. Robust face recognition systems that can function in difficult situations are becoming more and more necessary as face recognition is used more often in a variety of applications, including security, surveillance, and identification.

3. Low identification accuracy, high computing complexity, susceptibility to noise, and fluctuations in lighting may be problems with the current methodologies for obscured face recognition, such as template matching and feature-based algorithms.

4. Deep neural networks have demonstrated promising results in a number of computer vision applications, such as face recognition, but little research has been done on how well they function in occluded environments.

5. Research on the effects of various types and degrees of obscuration on face recognition accuracy is lacking, as is the creation of deep neural network-based algorithms that may successfully address such difficulties.

6. Comprehensive assessment measures and data sets are required to make it easier to create and contrast obscured face recognition systems.

1.3 Research Objectives

More often people are exposed in public by removing their mask for facial recognition purposes. We kept our main goal of developing a medium that will help in the sector of facial recognition in the security sector, surveillance sector, schools, and colleges. The main aim of this research is to identify objects/faces regardless of their position or orientation. It is sufficient to distinguish an item by identifying the manifold to which it belongs [20]. This goal will help in the reduction of criminal activities by getting the criminals with the help of facial recognition. The accuracy rate and the efficiency of our model will help in the betterment of the security sectors and technology sectors. Along them, the other objectives of this paper are:

1. Constructing a deep neural network technique for detecting faces that are partially obscured or blocked by objects or other obstructions.

2. Assessing the algorithm's performance and contrasting it with current face recognition methods.

3. Examining the effects of various obscuration kinds and levels on face recognition accuracy and determining the most difficult situations.

4. Investigating how various pre-processing methods, such as feature extraction and picture augmentation, might increase recognition accuracy in difficult-to-see situations.

5. Examining transfer learning approaches' ability to improve the performance of the created algorithm, especially when training data is scarce.

6. Carefully examining the generated algorithm's resistance to threats and exposure to hostile cases

.

Chapter 2

Literature Review

2.1 Brief History of Deep Neural Networks

Deep neural networks (DNN) have been around for several decades, but have recently gained significant popularity in the field of machine learning due to advancements in computing power and data availability.
In the middle of the twentieth century, researchers began experimenting with artificial neural networks (ANN) as a way to model the human brain and create artificial intelligence. However, these early networks were limited in their complexity and ability to learn from large amounts of data.
At the end of the twentieth century, researchers began developing deeper networks with multiple layers, known as multi-layer perceptrons (MLPs). These networks were able to achieve better performance on certain tasks but were still limited in their ability to learn from large datasets.
In the early years of the twenty-first century, researchers began using unsupervised learning techniques, such as stacked auto-encoders, to train deep neural networks with multiple layers. These networks were able to learn more complex representations of data and achieve improved performance on a variety of tasks.
Convolutional neural networks (CNNs) and recurrent neural networks are examples of even deeper neural networks that have been developed in the contemporary era as a result of improvements in computer power and the accessibility of vast quantities of data (RNNs). These networks have been employed in many different applications, ranging from speech and picture recognition to self-driving automobiles and natural language processing.
A recent research found that the number of papers on DNNs published in prestigious conferences and publications has grown rapidly, with an annual growth rate of almost 60% between 2012 and 2018. DNNs have also been utilized to perform at the cutting edge on a number of tasks, including image classification, speech recognition, and natural language processing.

2.2 Artificial Intelligence and Machine Learning against Facial Recognition

Facial recognition technology makes extensive use of machine learning (ML) and artificial intelligence (AI) algorithms to locate and recognize faces. These algorithms

examine facial photos or videos, find distinctive facial traits, then match those features to well-known people or identities using pattern recognition techniques. By allowing systems to learn from fresh data and situations, AI and ML are essential to facial recognition. To uncover common visual traits and patterns that may be used to identify people, machine learning algorithms, for instance, can be trained on enormous datasets of faces.
The accuracy of facial recognition has also been improved by using deep learning, a kind of machine learning. Deep neural networks can examine several layers of data, which enables them to recognize face characteristics in more complicated patterns and with higher accuracy.More robust face recognition systems that can manage changes in lighting, position, and other aspects that may affect identification accuracy may be created using AI and ML.
The use of face recognition technology is not without controversy, though, especially in light of privacy issues and inherent biases. In addition to raising ethical concerns about the technology's usage and its effects on society, the use of AI and ML in face recognition systems also brings up legal and regulatory issues. Ultimately, AI and ML are essential parts of face recognition technology, but their use must be carefully thought out and weighed with ethical issues, worries about privacy, and concerns about biases.

2.3 Related Works

In the past many multiple types of research have been done prior to facial recognition with the help of different algorithms and data sets.
For the first part of facial recognition the paper H. Goyal et. al. [14] used the MobileNetV2 which is pre-trained as having a global pooling block to help in the detection of the face. Here it creates a multidimensional component from a given image. The Kaggle data set is used to train and assess the model of the paper. The acquired data set contains roughly 4,000 images and gives a 98% performance accuracy rate. In comparison to DenseNet-121, MobileNet-V2, VGG-19, and Inception-V3, the suggested model is computationally efficient and exact. This model comes in handy as a digital scanning tool in schools, hospitals, banks, airlines, and a variety of other public and commercial settings.[3]
Z. Mang et. al. [19] talk about the use of ResNet34 and its use as a feature for facial extracting. Mask transfer, AMaskNet, is a low-cost approach for mitigating the detrimental impact of mask errors on face recognition. RMFRD, COX, and Public-IvS are the three masked face recognition datasets that the proposed technique regularly outperforms. Meanwhile, qualitative analysis trials with CAM show that AMaskNet's contribution is more suited to masked face recognition.[6]
For the help of real-time face recognition use of ResNet-10 is prominent and plays a backbone architecture for the proposed methodology of the paper by PreetiNagrathet. al. It helps in the construction of the baseline to have a complete study of the impact of the masked faces. The SSDMNV2 technique uses the Single Shot Multibox Detector as a detector for faces and the MobilenetV2 architecture for an structure to build the classifier, which is relatively light and can be used in embedded devices like the NVIDIA Jetson Nano and Raspberry Pi for real-time mask detection. As a result of the methodology used in this study, the ultimate result is a 92.64% score for accuracy and a 93.4% F1 score.[9]

For further improvement we can see, Mehendale et. al. state the use of OpenCV for the comparison of the images and helping in the execution of the proposed methodology described in the paper. They suggested a system that generates 128-d encoding for face identification from static photos, live video streams, and static video files using the deep metric learning approach and our proprietary FaceMaskNet-21 deep learning network. They attained an 88.92 percent testing accuracy with an execution time of less than 10 milliseconds.[13] The system's capacity to do real-time masked face recognition makes it suited for recognizing persons in CCTV. [12]
C. Li et. al. [4] state about the procedure of the detection of face through the DarkNet53 (YoloV3).
Sadid et. al. [11] used the InceptionV3 model that was run along the data base to give a performance accuracy for face detection than most others. For the Comparison of theResNet50, MobilenetV2 and InceptionV3 the use of SVM and Softmax is used in order to get performance accuracy. Another classifier used in the proposed methodology is the Multilayer Perceptron Classifier (MLP) for the classification of faces. For the optimizer of an image, the Adam optimizer is used. [1]
H. N. Vu et. al. [18] talk about the LPB textual feature extraction. The paper proposes a method that uses RetinaFace, a joint extra-supervised and self-supervised multi-task learning face detector that can deal with various scales of faces, as a fast yet effective encoder to recognize the masked face by combining deep learning and Local Binary Pattern (LBP) features. The local structure of a picture is encoded using the texture descriptor LBP. The LBP features are then utilized to produce a feature vector for each masked face image The proposed method pipeline in the figure:2.1 [18] shows the methodology for the paper discussed. The CNN classification

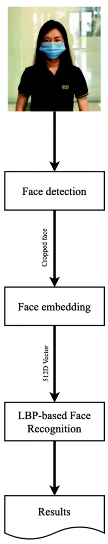

Figure 2.1: Proposed method pipeline

module then receives the feature vectors and learns a mapping between the feature vectors and the identities of the dataset's individuals. Convolutional and fully connected layers that are trained using the backpropagation technique make up the CNN classification module. We are able to see a feature histogram and localization of landmark points in image in figure:2.2 & figure:2.3 [18]
The study's findings demonstrated that the suggested method had a high level of success in masked face recognition. For the LFW-Masked dataset, the model had

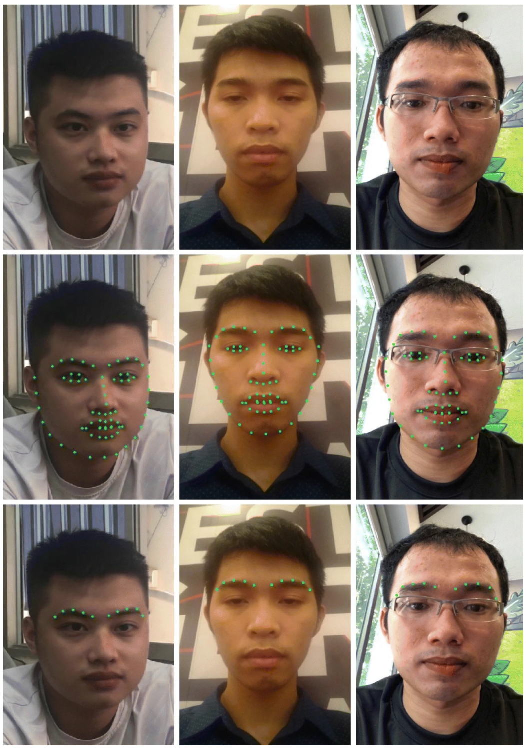

Figure 2.2: Localization of landmark points

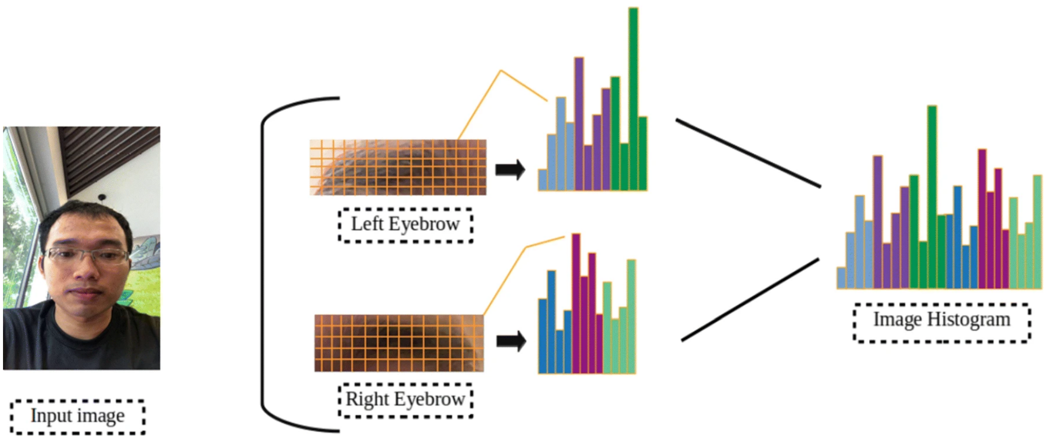

Figure 2.3: Feature histogram created by the LBP pipeline

an accuracy of 97.36 percent, while on the MFRD dataset, it had an accuracy of 92.36 percent. The Roc curve is shown in figure:2.4 [18]

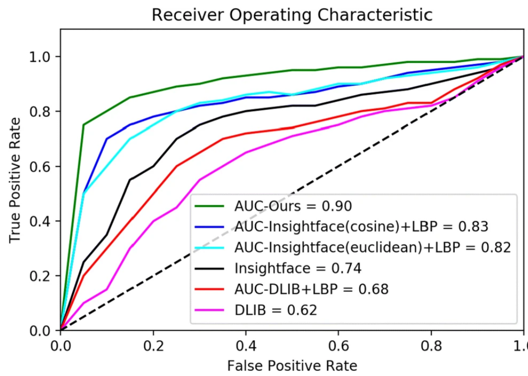

Figure 2.4: ROC curves

They blend characteristics learned from RetinaFace with local binary pattern features extracted from the masked face's eye, forehead, and eyebrow regions to provide a single framework for detecting masked faces. The accuracy of two datasets, COMASK20 and Essex, is 87% f1-score and 98% f1-score, respectively. [17]
P. Ghadekar et. al. [8] talk about the VGG16 model and compares the paper-proposed model with the architecture of VGG16 for the better accuracy rate of the facial extraction. In the case of deployments and structuring, a novel Keras-based model is provided that outperforms the MobileNet-V2 and VGG-16 standard designs in terms of accomplishment. The MTCNN method is used to challenge effective face localization. In terms of accuracy, support values, precision, recall, and F1-score metrics, the suggested model is compared to conventional VGG16 and MobileNetV2 architectures. [16]
In a paper by Hariri and Walid [15] they talk about the BOP (Bag of Features) which is a paradigm that helps in presenting the images as an unordered set of local fea-tures. It is used as the pooling layer for the training convolution layer. The three deep Convolutional Neural Networks (CNN) which were trained on millions of images are ResNet-50, VGG-16, and AlexNet which are utilized to extricate deep facial features from the region of interest (mainly the eyes and forehead regions). The requirement for efficiency and the simplicity of reducing occlusions are the driving forces behind the suggested method. The RMFRD and SMFRD datasets' masked portions are eliminated using this technique. The overview of the proposed model of the paper is shown in figure:2.5 [15]

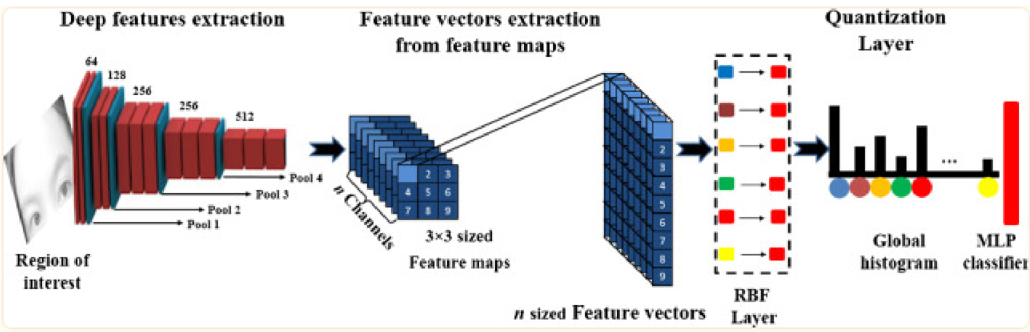

Figure 2.5: Overview of the proposed method

For preprocessing, Dlib-ml is utilized to perform a 2D rotation of the face as shown in figure:2.6 [15], normalize the image to 240 × 240 pixels, and partition the image into 100 fixed-sized blocks, with only blocks 1-50 being considered as shown in figure:2.7 [15]

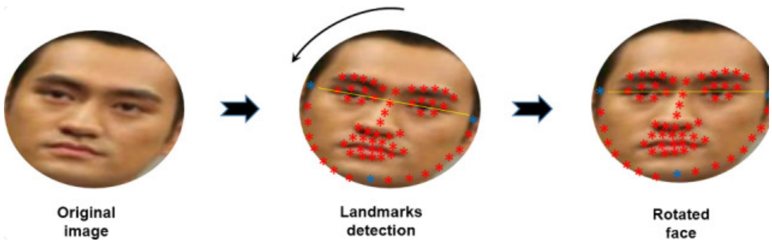

Figure 2.6: 2D Face rotation

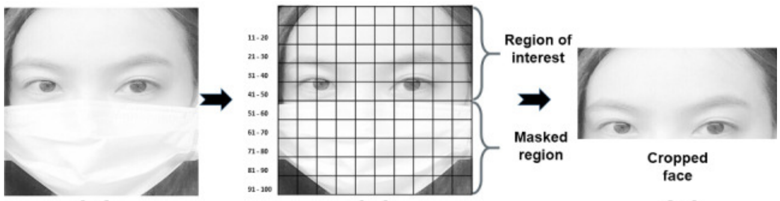

Figure 2.7: Masked face, Sampling the masked face image,Cropping filter

For feature extraction, three pre-trained models, namely AlexNet, VGG16, and ResNet50, are employed for quantization.
A Multilayer Perceptron is used in the classification method (MLP). Experimental results on the Real-World-Masked- Face-Dataset demonstrate high recognition performance when compared to other cutting-edge methods. [2]
The paper of E. Ryumina et. al. [10] shows us the accuracy rate of these different models compared with each other with the help of a dataset. On the MAFA and RMFD datasets, the proposed hybrid technique improves the Unweighted Average Recalls (UARs) of identification of the presence/absence of a protective mask on a human's face by 0.96 % and 1.32 %, respectively, when compared to standard CNNs. The comparison is able to give a clear image of the algorithms and their role in image recognition having a masked face or an obscure face.

Chapter 3

Methodology

First, we scaled all of the faces-depicting images from the CASIA, VGG2, and LFW datasets to a 128:128 ratio. The images are 128 pixels wide and 128 pixels tall in size. We performed augmentation and normalization to enhance the performance of the models. While augmentation includes randomly modifying the training data to boost its diversity, normalization entails scaling the input data to have a mean of 0 and a standard deviation of 1. Then, using deep convolutional layers, we applied filters to the images in order to extract numerous facial patterns and attributes. We introduced completely connected 1-layer layers for prediction, where each neuron is connected to every other neuron in the layer below. It was employed in the very last levels of our models. 512-dimensional embedding, a vector representation of an input image created by a neural network, was the next step. We utilized this vector, which has 512 dimensions, to assess how similar the dataset's photos are to one another. Then, in order to better distinguish between various classes of faces in the embedding space, we applied two loss functions which are the ArcFace loss function and Triplet loss function.

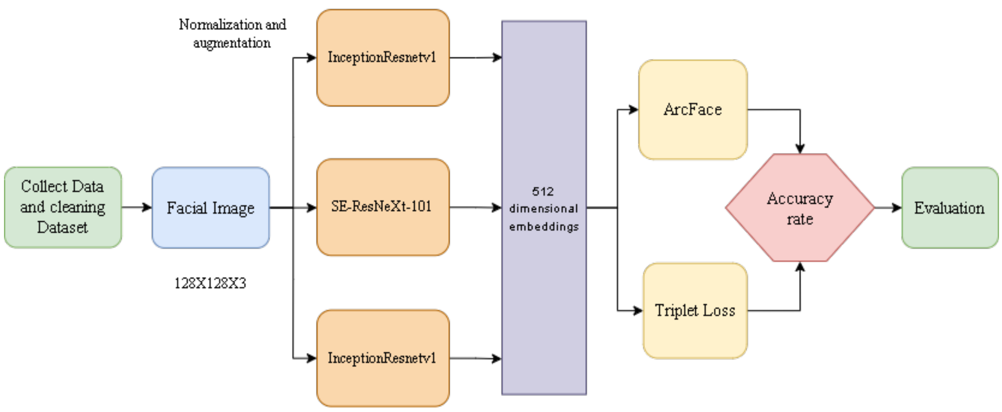

Figure 3.1: Work Methodology

Chapter 4

Description of Data and Models

4.1 Description of Data

4.1.1 Pre-Processing

For our working procedure, we have used 3 datasets. The three datasets were CASIA, VGG2, and LFW of which two datasets were used for training and one of the datasets was used for the testing part. CASIA and VGG2 were used for training and LFW was used for the testing purpose. The CASIA, VGG2, and LFW datasets had identities and images of 10,585 and 492,832, 8,631 and 2,024,897, and 5,749 and 64,811 respectively before using MaskTheFace tool. As the faces had no masks we used MaskTheFace tool for masking the faces for getting a better result and better embedding. The MaskTheFace tool was able to generate different types of mask, for example, surgical mask, kn95 and cloth for the training datasets. And after using this tool the identities were left of about 12k identities When MaskTheFace tool was applied on the CASIA and VGG2 datasets we found some error images and a good amount of masked face images. After that we manually created a csv file consisting of almost 1 million images. After using MaskTheFace tool for LFW we've got some error images and a good amount of masked images. Among them we've manually created a csv file consisting of almost 64K and 10% of that we've used for testing is 6482 images and the rest of that is used for validation 58310 images

Figure 4.1: CASIA Dataset using MaskTheFace

Figure 4.2: CASIA Dataset without mask

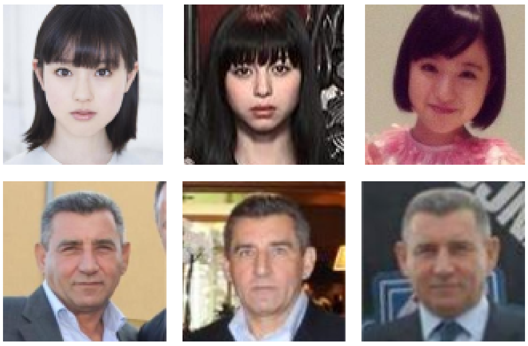

Figure 4.3: VGG2 Dataset without mask

Figure 4.4: VGG2 Dataset using MaskTheFace

Figure 4.5: LFW Dataset without mask

Figure 4.6: LFW Dataset using MaskTheFace

4.1.2 Data Augmentation

Mainly, Data augmentation is the process to increase the diversity of a dataset by employing different transformations in our datasets with the objective to improve generalization and performance of the model by displaying it to large extent of variations. We have augmented all the data of our used datasets. For training, validation, and testing we need to do data augmentation. So, we rescaled all the face images of all the three datasets. We have resized all the images to a 128:128 ratio and performed a horizontal flip with a probability of 0.5. Finally, we have normalized the dataset images. We performed these tasks for both the training, validation and the testing dataset.

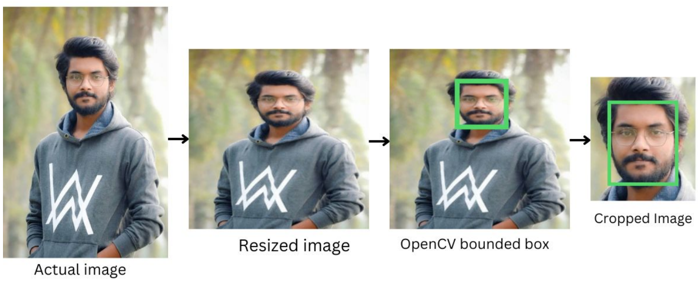

Figure 4.7: Image Augmentation

4.2 Description of Models

4.2.1 InceptionResNetV1

Since we have a large amount of facial images in our dataset, we used transfer learning. Without using transfer learning, it will take a lot of time to train. Transfer learning also helps the loss converge faster.
FaceNet-Pytorch provides pretrained InceptionResNetV1. After loading the InceptionResNetV1, which was pre-trained, as the classifier, we used PyTorch to keep training this pretrained model with our masked face dataset. The output size of the 1 layer fully connected layer was not needed to be modified as in InceptionResNetV1's last linear layer's output size is 512.

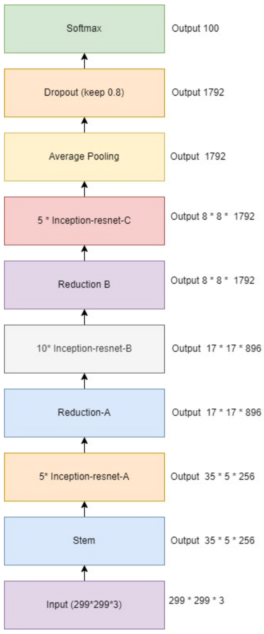

Figure 4.8: InceptionResNet V1

4.2.2 Triplet Loss

In addition to the three triplet pairs—easy, semi-hard, and hard—the toughest triplet is used to train the model. This implies that the face images of the same person with the greatest difference from the anchor picture and the individual with the greatest similarity to the anchor image will be linked. We would use our function to determine the most difficult triplet pairings in order to carry out offline triplet mining, and we would use an online triplet loss package to carry out online triplet mining. We sample our dataset using Rwightman's PK samplers before applying online triplet mining since the dataset has to be sampled in accordance with its goal.The sampler assists in the mapping of the original dataset into batches by specifying the number of classes and the number of images that should be included in each batch for each class. This can assist prevent the failure condition in pair selection when a class only contains one image in a batch.

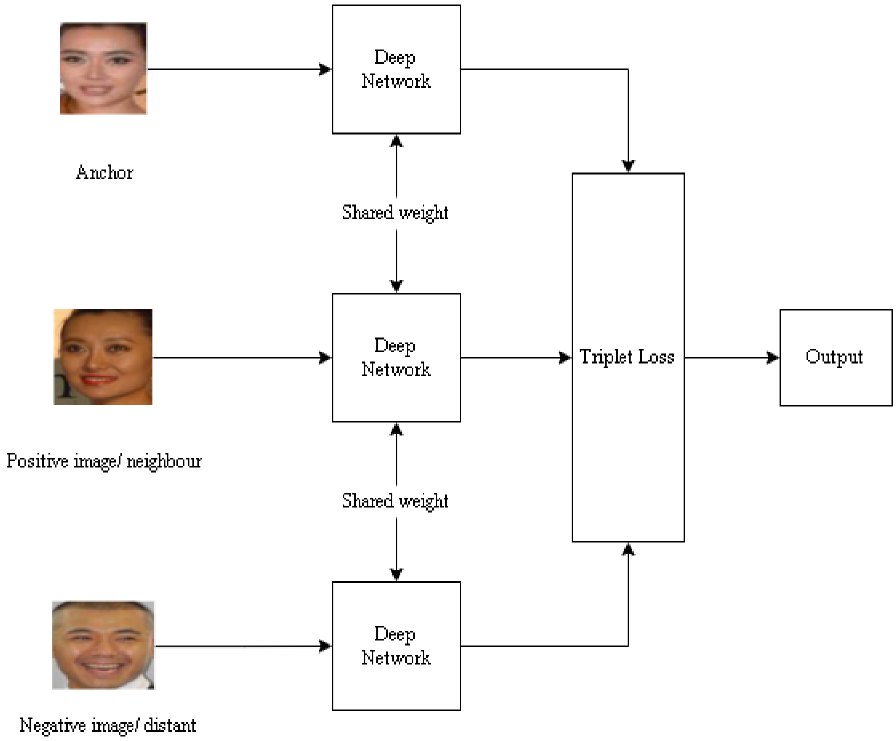

Figure 4.9: Triplet Loss

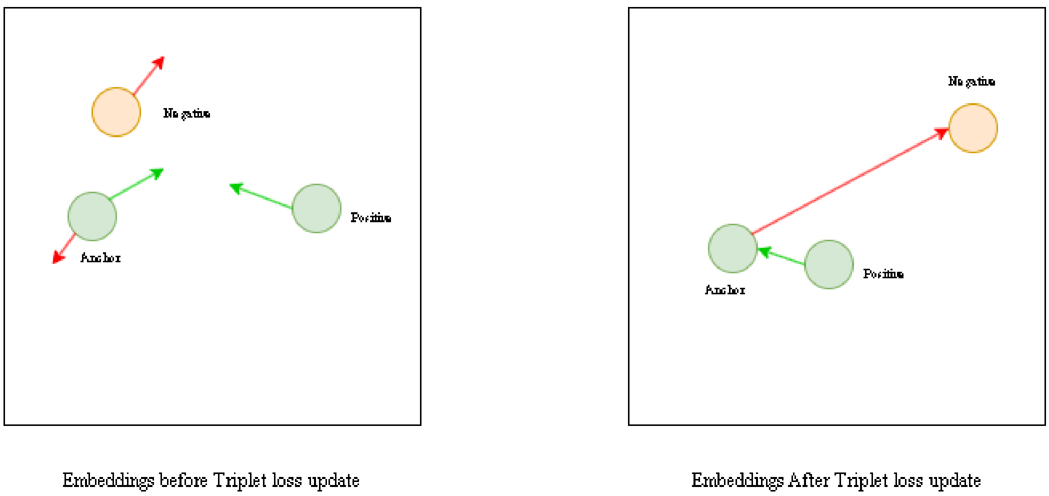

Figure 4.10: Triplet Loss update

4.2.3 SE-ResNeXt-101

There was no SE-ResNeXt-101 model that had been trained using face picture datasets. The bulk of the pre-trained SE-ResNeXt-101 models were trained on the massive object recognition dataset ImageNet. A pre-trained object recognition model cannot be used for transfer learning in face recognition. Using the model's feature extraction, only object-specific properties may be retrieved. Using this model is at least as awful as starting from scratch and training a model with weight initialization. We train a SE-ResNeXt-101 from the ground up, with no assistance from transfer learning. The Python package timm contains a number of pre-implemented image models. SE-ResNeXt-101 is one of the pre-implemented models in timm. SE-ResNeXt-101 was imported from timm using it's function, and it was trained using the same approach and hyper-parameters.

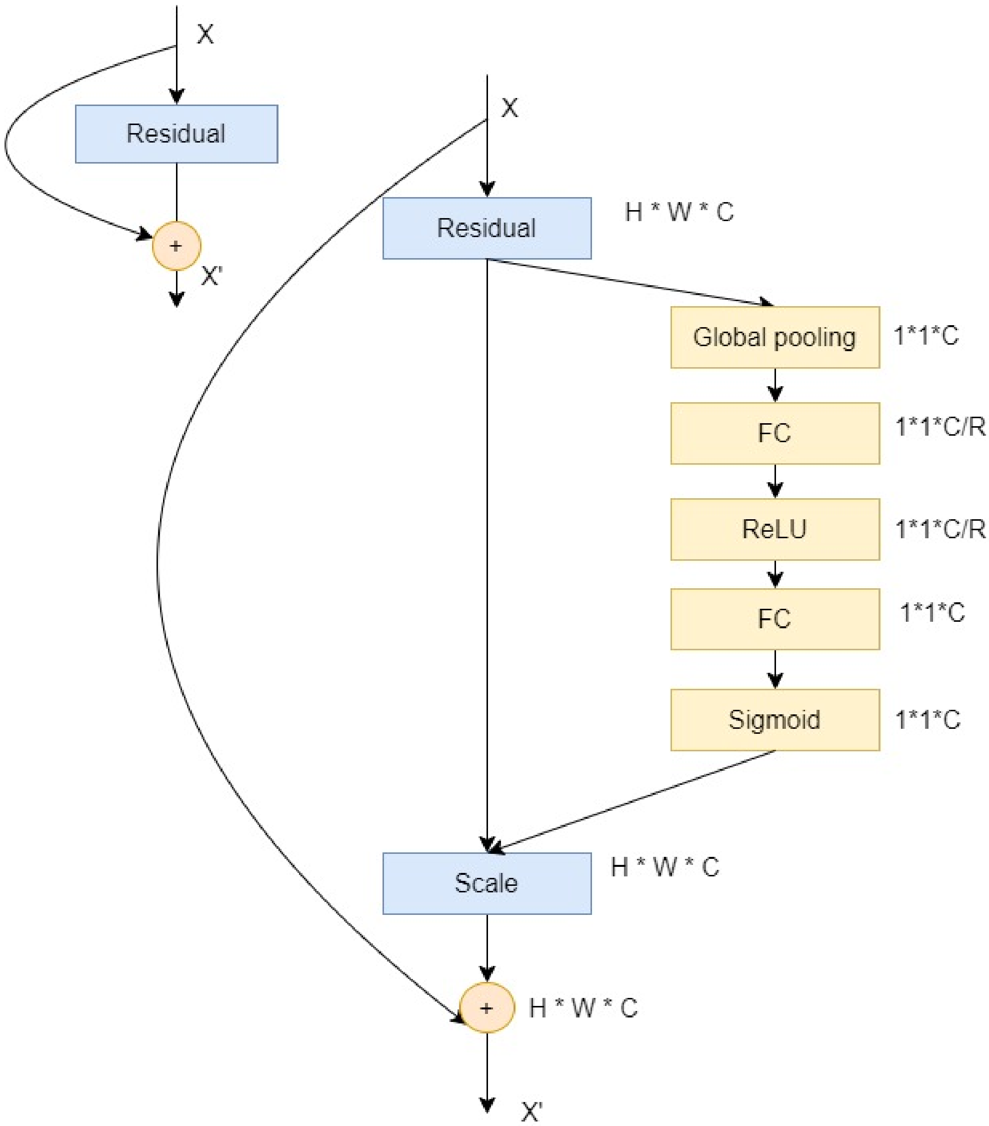

Figure 4.11: SE-ResNeXt-101

4.2.4 ArcFace Loss

While ArcFace is more difficult to build than triplet loss, it is easier to tune the hyperparameters and does not require mini-batch sampling during training. In the past year, facial recognition function loss has become more prevalent than triplet loss. We optimized using this implementation of ArcFace and focus loss. In order to determine which performed better for our masked facial recognition model, we also evaluated ArcFace and triplet loss performance.
The ArcFace loss function that is often used in face recognition tasks, and it can be defined as follows:

$$L = -\frac{1}{N}\sum_{i=1}^{N}\left[\log \frac{e^{s(\cos(\theta_{y_i})-m)}}{e^{s(\cos(\theta_{y_i})-m)} + \sum_{j\neq y_i} e^{s\cos(\theta_j)}}\right] \tag{4.1}$$

here,
N = number of samples
s = scaling factor
m = margin
θyi= angle between the i-th input embedding and its corresponding class center (yi)
θj = angle between the i-th input embedding and the j-th class center $(where j \neq yi)$

Figure 4.12: ArcFace loss function

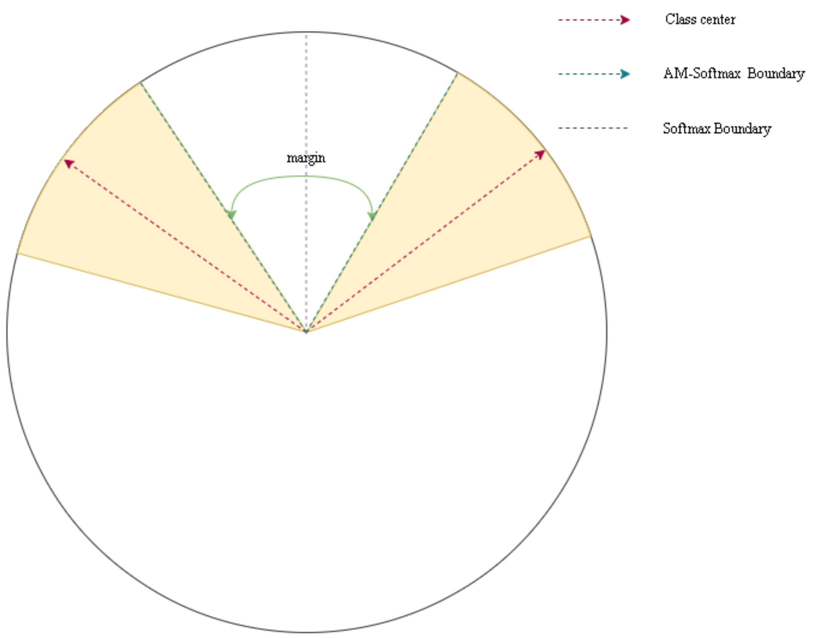

Figure 4.13: ArcFace

4.3 Training and Testing

4.3.1 Training

In the training phase, we used a batch size of 128 images in each epoch that iterates over the whole dataset and reads them one by one. From the beginning of our training, when we input the first image in our model, which outputs an embedding, we see if our model is able to recognize the image as we pass it along to the loss function. In the loss function, we first ensure that we are calculating the loss for similar images by comparing their UIDs. If this is the case, then we proceed to compare the embedding and get the loss. The loss function was highest in the beginning since it had no previous information and would not be able to recognize the image, so we saved the loss. Sequentially, when we read the next image, we read them again and see if our model recognizes them and passes them to the loss function. It compares with the previous image's embeddings to get the loss and add it cumulatively, and now the train loss is slightly lower since it encounters some similarities. Moreover, we cumulatively add the loss, update the threshold, and keep updating parameters for every other batch too. Additionally, after the end of batch training, we want to find a distance threshold, so we use a dataset that we divide into two parts that contain the same person image pair and a different person image pair, take the minimum value from it, repeat this for every epoch, and update our distance threshold. And gradually we get a decreasing graph of train loss, which signifies our training is working properly. Furthermore, when we finish training our model for the full epoch, we get a distance threshold, but to validate, we again use the same and different image pair dataset to update and get a near-perfect distance threshold, which will be used for training purposes. One thing to remember here is that we don't update any losses when we calculate the loss between the last image of a similar image and the first image of a different image since they are supposed to be different and our model is supposed to find no similarities between them.

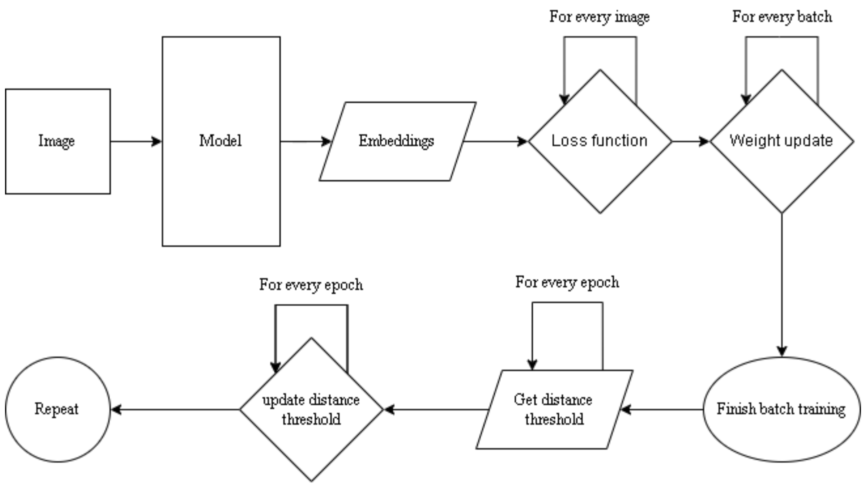

Figure 4.14: Training Workflow

4.3.2 Testing

For testing, we used the augmented LFW dataset with 6482 images that had exactly 50 percent of similar users' image pairs and another 50 percent of different users' image pairs. We input an image pair in our trained model (true label) that is being read one by one along with the distance threshold, and our model outputs a prediction (predicted label) if the image pair is the same person or not.

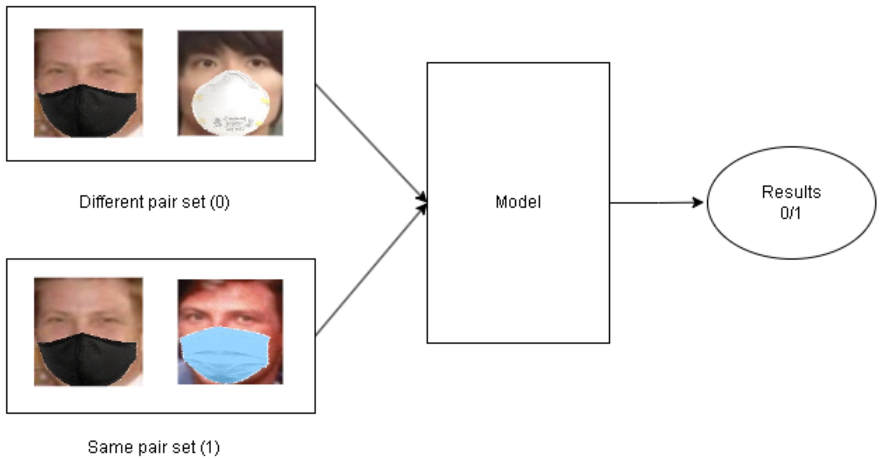

Figure 4.15: Testing Workflow

Chapter 5

Results and Comparative Analysis

5.1 Results

Via the training, validation, and testing phases, we compare their outcomes. InceptionResNetV1 and ArcFace have produced the best model, which will also be used as the ultimate model in the access control system. The best outcome we obtain using various model and loss function combinations is displayed in the following table. t-SNE study also shows how our model is capable of obtaining similar embedding for photographs from the same person.

Model	Architect	Loss function	Pre-trained	Acuuracy
Model1	InceptionResNetV1	Arcface	Yes	95.85%
Model2	InceptionResNetV1	Triplet Loss	Yes	94.28%
Model3	SE-ResNeXt-101	Arcface	No	93.41%

Figure 5.1: Accuracy Rate of the Models

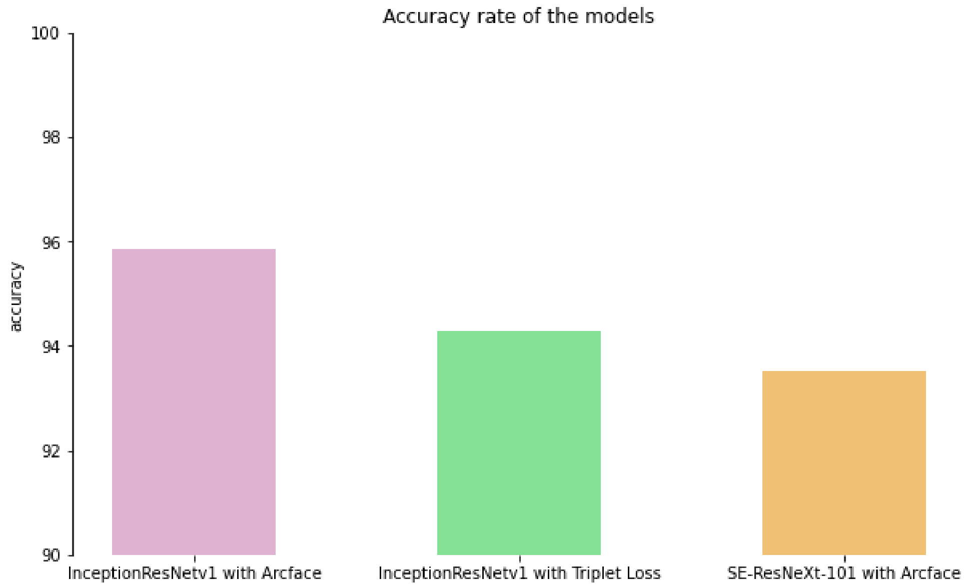

Figure 5.2: Accuracy Rate graph

5.2 InceptionResNetV1 with Triplet Loss

Because there are so many various permutations of triplet pairings that might be picked, we begin with the harshest condition, where we have chosen the hardest negative and hardest positive for each image pair. The performance of the pre-trained model and the training result using a small dataset are nearly comparable. It also takes a long time to re-pair the full dataset. Using a small dataset of less than 10000 data by choosing only the pair for each training process is still justifiable, but it would take tremendously long time if we re-pairing a large dataset consists of almost 2.5 million data at the beginning of each epoch, if we consider the time to re-pair for each epoch it will be even longer and outweigh the actual training time. We attain an accuracy of 94.28% in our testing set of the top model with hard triplet training. The test accuracy at 94.28% and considering the lowest IOU at 0.054, we can notice that IOU that we select as the criterion for validation is almost the same as the real test result. The dataset is abandoned because it is too small, causing the loss on it to be unstable, and since the IOU value may already correctly reflect how well our model is performing.

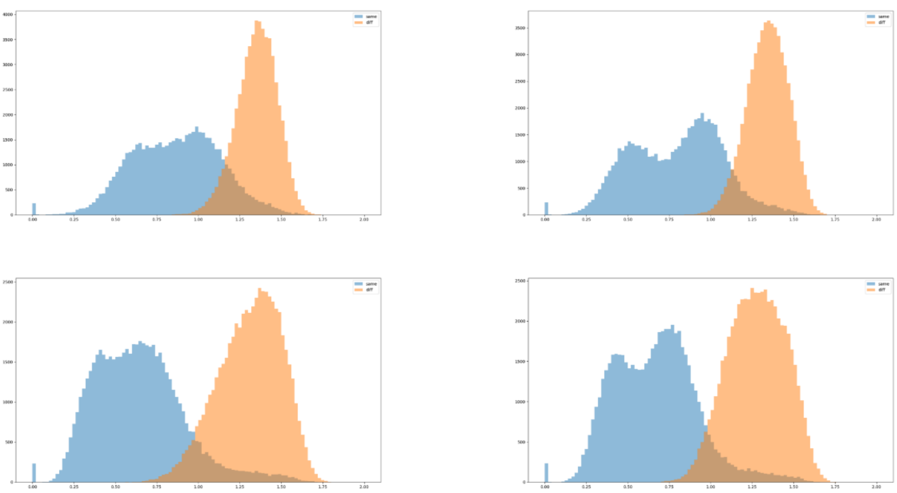

Figure 5.3: Triplet loss after 5, 12, 20, 28 epoch

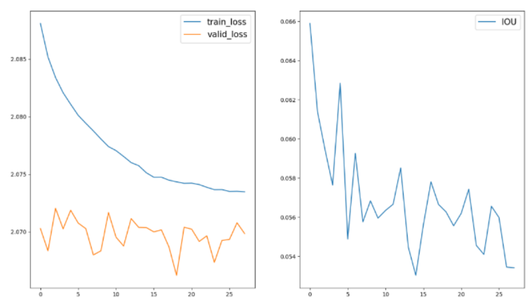

Figure 5.4: Training loss and IOU(Intersection over union) of the best model during training.

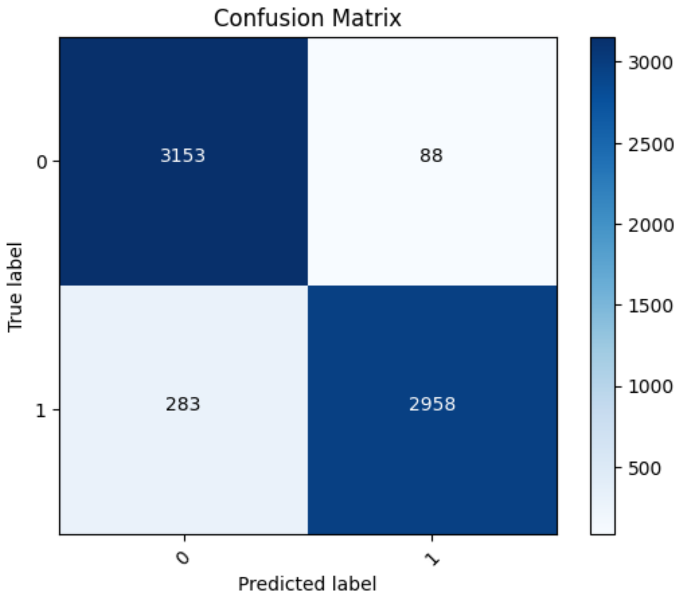

Figure 5.5: Test Results

5.3 InceptionResNetV1 with ArcFace Loss

We attempt to optimize the pre-trained InceptionResNetV1 utilizing ArcFace in addition to triplet loss. ArcFace's test accuracy is marginally higher than the triplet loss outcome using the identical model architecture, InceptionResNetV1.

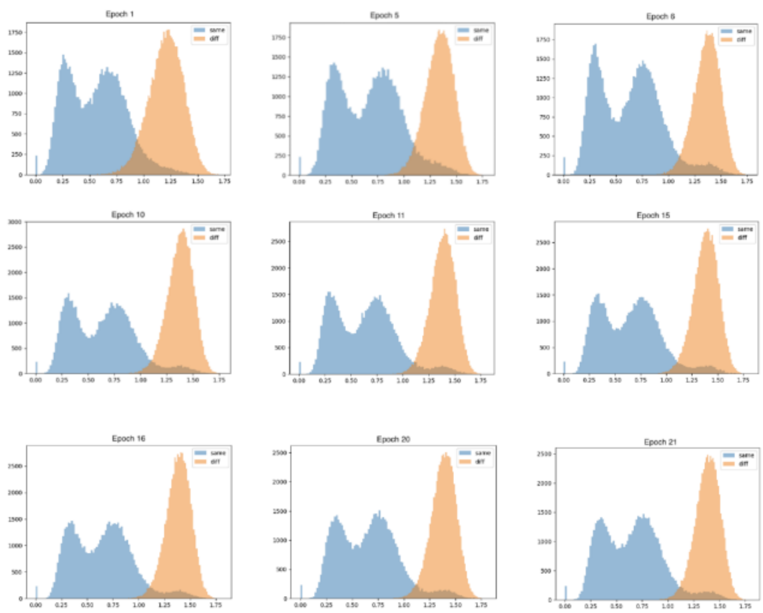

Figure 5.6: Validation graph in epoch 1, 5, 6, 10, 11, 15, 16, 20, 21

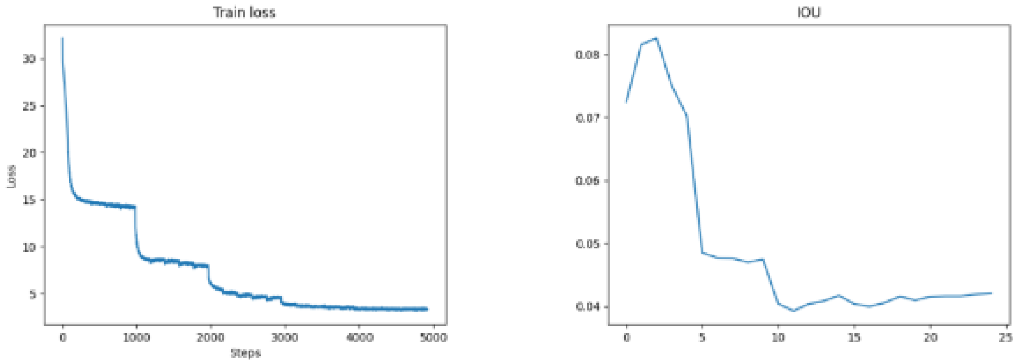

Figure 5.7: Training loss and IOU(Intersection over union) of InceptionResNetV1 with ArcFace Loss

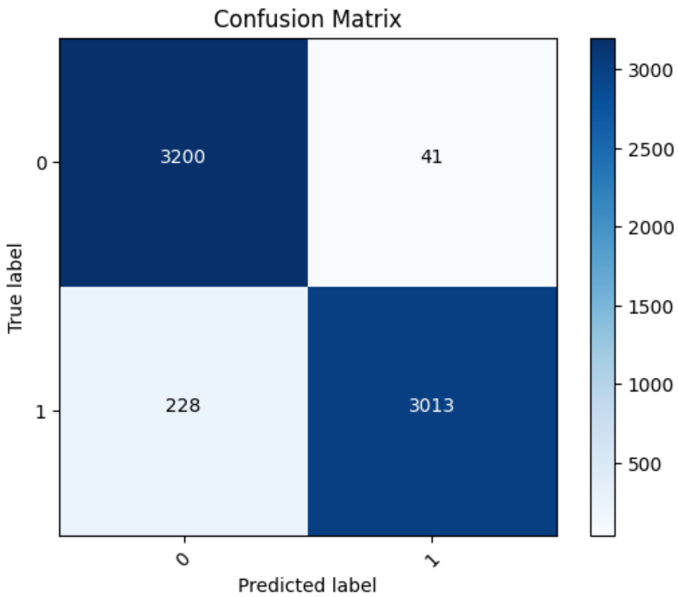

Figure 5.8: Test Result

5.4 SE-ResNeXt-101 with ArcFace Loss

Although the findings of InceptionResNetV1 are generally positive, we started running further tests to investigate how much transfer learning and the model's architecture impact the test outcomes. As a consequence, we start from scratch training SE-ResNeXt-101 with ArcFace, a new model. The result of this model is the worst of the three possibilities. This, we believe, is because SE-ResNeXt-101 was trained on a face picture dataset without the use of transfer learning. As a result, it is less capable of extracting characteristics than others.
Nevertheless, SE-ResNeXt-101's performance might also be due to its design, which varies from that of InceptionResNetV1 . The SE-ResNeXt-101 model is deeper and has a higher number of parameters, which might make training without transfer learning more challenging. To increase the performance of facial recognition models, further experimentation with alternative architectures and training methodologies may be required.

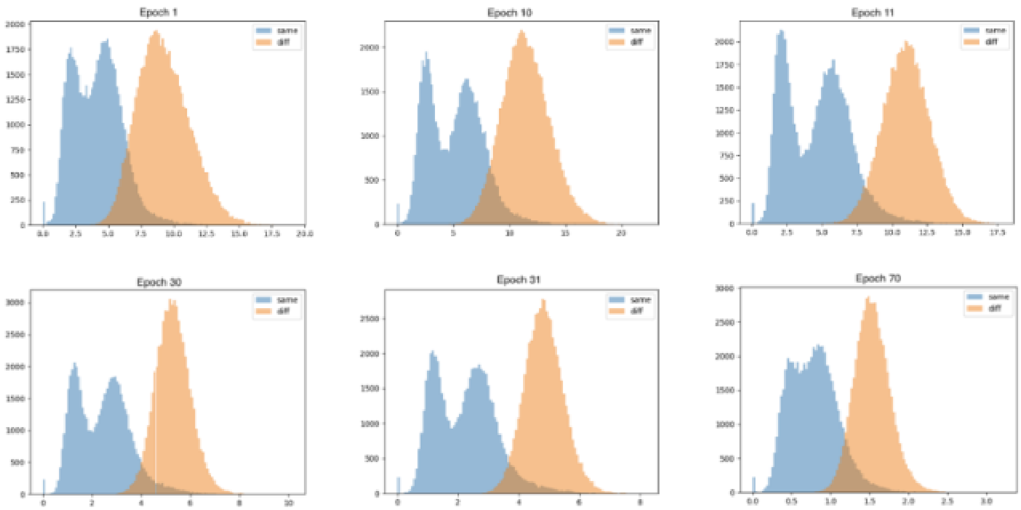

Figure 5.9: Validation graph in epoch 1, 10, 11, 30, 31, 70

After the testing is done we are able to find the test results of precision , recall and F1-score as shown in the figure 5.12 .The accuracy of a model's prediction of a positive class is measured by precision, and we correctly recognized the two faces with a high precision of 96%. In terms of the percentage of true positive classifications among all real positives in the data, we likewise attained a recall of 90%. To provide a fair assessment of the model's performance, the F1-score, a statistic that combines accuracy and recall into a single figure, was chosen as the final metric. The harmonic mean of recall and accuracy is used to get the F1-score.These assessment measures are critical in determining how well facial recognition algorithms function. High accuracy and recall scores imply a model that is extremely precise and effective in identifying positive classes in data. In conclusion, our model's excellent accuracy and recall show that it is capable of successfully detecting faces, giving it a potential option for a variety of real-world applications. However, more testing and refinement of the model can assure its usefulness and dependability in many settings.

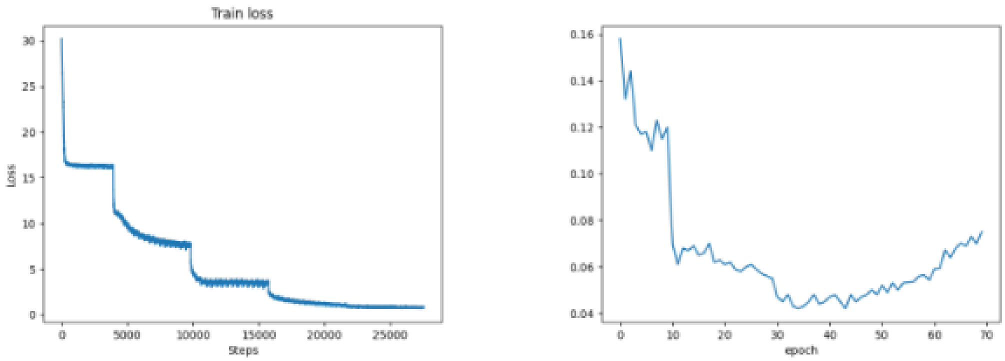

Figure 5.10: Training loss and IOU(Intersection over union) of SE-ResNeXt-101 with ArcFace Loss

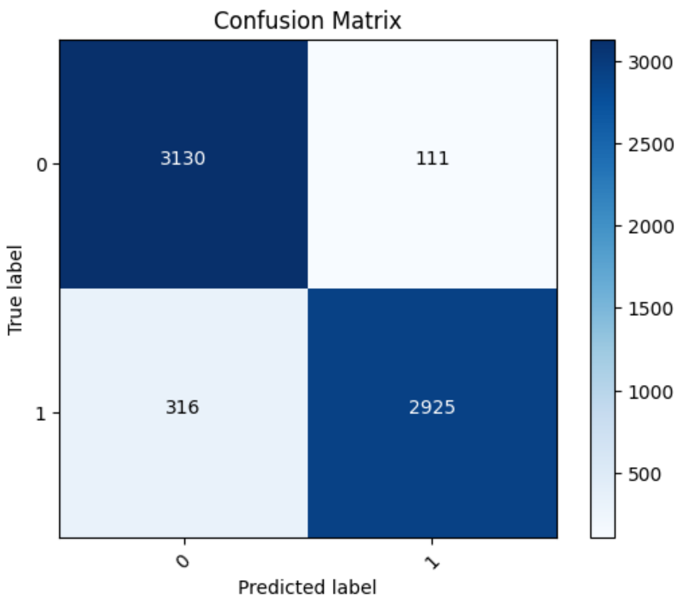

Figure 5.11: Test Output

	Precision	Recall	F1-score
0	0.91	0.97	0.94
1	0.96	0.90	0.93
accuracy			0.93
macro avg	0.94	0.93	0.93
weighted avg	0.94	0.93	0.93

Figure 5.12: Test Result of precision, recall and F1-score

5.5 Comparison Analysis

In our comparison paper [7], they used Inception-ResNet-v1 as their primary model architecture. Inception-ResNet-v1 is composed of two primary components which are a reduction module and an Inception-ResNet. It is for the sole purpose of extracting features and shrinking the size of the feature map, as a parallel structure is used in the reduction module. The Inception-ResNet module does away with the necessity for feature map size transformation by using residual connections to replace the pooling operation. The Inception-ResNet module is being replaced with the Attention model to improve the network's performance. CBAM attention system is implemented in the Attention model of Inception-ResNet in contrast of using only the Inception-ResNet module, allowing the model to focus on the most important aspects of the image. Furthermore, they integrated ArcFace as their loss function by replacing non-grouping learning techniques used in triplet loss function. Here the Inception-ResNet-V1 with reduction module is shown in figure:5.13 [7]

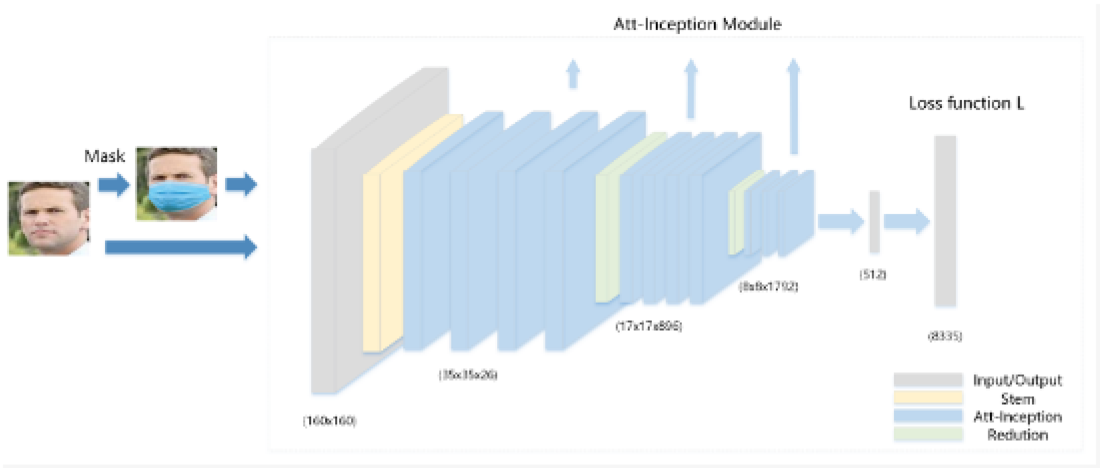

Figure 5.13: Inception-ResNet-v1 with reduction module

They employed 85,000 images that were obtained after pre-processing the RMFD dataset and worked on their model with it. Their experiments yielded 92.28% accuracy on RMFD utilizing the Arcface approach. As a result, we achieved an accuracy of 93.41% in our work using SE-ResneXt-101 as our primary model and Arcface as our loss function 1.13% higher than their proposed model with different dataset.

Chapter 6

Conclusion

In our research, we proposed and implemented an innovative strategy for recognizing obscure masked faces by utilizing cutting-edge techniques. Deep neural networks are difficult to use for face detection, but with careful analysis and extraction, we can achieve our goals. Specifically, we integrated the SE-ResNeXt-101 algorithm for fast and accurate recognition of masked faces. Our proposed system utilizes the CASIA, VGG2, and LFW datasets that will help our model to train on obscure face recognition of a person that will successfully detect if the person matches with his obscure masked face version. Through our research and studies, we achieved an accuracy of 93.41% using conventional RGB images.
Our suggested approach represents a significant advance in the field of face recognition since it can recognize human faces even when they are obscured by masks. Although these computations are computationally demanding, our access to resources is a restriction. Additionally, our findings suggest that the suggested approach has the potential to enhance face recognition's effectiveness and provide benefits. We firmly think that our suggested method can be integrated into the current face recognition process and will accomplish its goals for the greater benefit. Yet, in order to increase our accuracy and strive for state-of-the-art performance, we will combine our unique model with other loss functions.

Bibliography

[1] M. Anik and K. Islam, “An efficient approach of face detection and recognition from digital images for modern security and office hour attendance system,” Ph.D. dissertation, BRAC University, 2015.

[2] M. Haque, T. Samee, *et al.*, “Facial recognition using dynamic image processing,” Ph.D. dissertation, BRAC University, 2017.

[3] B. Shams, F. I. Nasim, *et al.*, “Analysis on face recognition based on five different viewpoint of face images using mtcnn and facenet,” Ph.D. dissertation, Brac University, 2019.

[4] C. Li, R. Wang, J. Li, and L. Fei, “Face detection based on yolov3,” in *Recent Trends in Intelligent Computing, Communication and Devices*, Springer, 2020, pp. 277–284.

[5] E. Ward, *A spate of mask-wearing criminals. crosstown*, 2020.

[6] C. Z. Basha, B. L. Pravallika, and E. B. Shankar, “An efficient face mask detector with pytorch and deep learning,” *EAI Endorsed Transactions on Pervasive Health and Technology*, vol. 7, no. 25, e4–e4, 2021.

[7] H. Deng, Z. Feng, G. Qian, X. Lv, H. Li, and G. Li, “Mfcosface: A masked-face recognition algorithm based on large margin cosine loss,” *Applied sciences*, vol. 11, no. 16, p. 7310, 2021.

[8] P. Ghadekar1, G. Singh, J. Datta, A. K. Gupta, and D. S. Anand, “Covid-19 face mask detection using deep convolutional neural networks and computer vision,” *ndian Journal of Science and Technology*, vol. 14, no. 38, pp. 2899–2915, 2021.

[9] P. Nagrath, R. Jain, A. Madan, R. Arora, P. Kataria, and J. Hemanth, “Ssdmnv2: A real time dnn-based face mask detection system using single shot multibox detector and mobilenetv2,” *Sustainable cities and society*, vol. 66, p. 102 692, 2021.

[10] E. Ryumina, D. Ryumin, D. Ivanko, and A. Karpov, “A novel method for protective face mask detection using convolutional neural networks and image histograms.,” *International Archives of the Photogrammetry, Remote Sensing & Spatial Information Sciences*, 2021.

[11] T. Sadid, M. S. Hossain, S. K. Sengupta, and S. M. Khondaker, “Approaching transfer learning to identify correctly worn facial masks,” Ph.D. dissertation, Brac University, 2021.

[12] M. Z. Asghar, F. R. Albogamy, M. S. Al-Rakhami, *et al.*, “Facial mask detection using depthwise separable convolutional neural network model during covid-19 pandemic,” *Frontiers in Public Health*, vol. 10, 2022.

[13] R. Golwalkar and N. Mehendale, “Masked-face recognition using deep metric learning and facemasknet-21,” *Applied Intelligence*, pp. 1–12, 2022.

[14] H. Goyal, K. Sidana, C. Singh, A. Jain, and S. Jindal, “A real time face mask detection system using convolutional neural network,” *Multimedia Tools and Applications*, vol. 81, no. 11, pp. 14 999–15 015, 2022.

[15] W. Hariri, “Efficient masked face recognition method during the covid-19 pandemic,” *Signal, image and video processing*, vol. 16, no. 3, pp. 605–612, 2022.

[16] D. Hussain, M. Ismail, I. Hussain, R. Alroobaea, S. Hussain, and S. S. Ullah, “Face mask detection using deep convolutional neural network and mobilenetv2-based transfer learning,” *Wireless Communications and Mobile Computing*, vol. 2022, 2022.

[17] G. Kaur, R. Sinha, P. K. Tiwari, *et al.*, “Face mask recognition system using cnn model,” *Neuroscience Informatics*, vol. 2, no. 3, p. 100 035, 2022.

[18] H. N. Vu, M. H. Nguyen, and C. Pham, “Masked face recognition with convolutional neural networks and local binary patterns,” *Applied Intelligence*, vol. 52, no. 5, pp. 5497–5512, 2022.

[19] M. Zhang, R. Liu, D. Deguchi, and H. Murase, “Masked face recognition with mask transfer and self-attention under the covid-19 pandemic,” *IEEE Access*, vol. 10, pp. 20 527–20 538, 2022.

[20] *Face recognition : University of dayton, ohio*, (Accessed on 01/12/2023).