A Face Recognition Method Using Deep Learning to Identify Mask and Unmask Objects

Saroj Mishra*, Hassan Reza
School of Electrical Engineering
and Computer Science
University of North Dakota
Grand Forks, ND 58202, USA
Email: *saroj.mishra@und.edu, hassan.reza@und.edu

***Abstract*—At the present, the use of face masks is growing day by day and it is mandated in most places across the world. People are encouraged to cover their faces when in public areas to avoid the spread of infection which can minimize the transmission of Covid-19 by 65 percent (according to the public health officials). So, it is important to detect people not wearing face masks. Additionally, face recognition has been applied to a wide area for security verification purposes since its performance, accuracy, and reliability [15] are better than any other traditional techniques like fingerprints, passwords, PINs, and so on. In recent years, facial recognition is becoming a challenging task because of various occlusions or masks like the existence of sunglasses, scarves, hats, and the use of make-up or disguise ingredients. So, the face recognition accuracy rate is affected by these types of masks. Moreover, the use of face masks has made conventional facial recognition technology ineffective in many scenarios, such as face authentication, security check, tracking school, and unlocking phones and laptops. As a result, we proposed a solution, Masked Facial Recognition (MFR) which can identify masked and unmasked people so individuals wearing a face mask do not need to take it out to authenticate themselves. We used the Deep Learning model, Inception ResNet V1 to train our model. The CASIA dataset [17] is applied for training images and the LFW (Labeled Faces in the Wild) dataset [18] is used for model evaluation purposes. The masked datasets are created using a Computer Vision-based approach (Dlib). We received an accuracy of over 96 percent for our three different trained models. As a result, the purposed work could be utilized effortlessly for both masked and unmasked face recognition and detection systems that are designed for safety and security verification purposes without any challenges.**

***Index Terms*—Face Recognition, Masked Facial Recognition, Verification, Security, Accuracy, CASSIA Dataset, LFW Dataset, Deep Learning, Dlib, Computer Vision**

I. Introduction

The use of face masks is growing rapidly with Covid-19. People are required to wear a face mask all the time when they are outside or at large indoor gatherings to minimize the spread of infection. So, it is important to detect those people with face masks for health safety reasons. Face recognition is the process of automatically identifying an individual from captured images or videos [4] and face detection is the process of identifying the face from the captured image or the specified image from the database. Additionally, facial recognition has been extensively applied for security verification purposes since its performance, accuracy, and reliability [15] are better than any other traditional techniques like fingerprints, passwords, tokens, and so on. Nowadays, it has become an arduous task due to various occlusions or masks such as scarves, sunglasses, hats, makeup, and other different types of disguise elements and they are causing a significant impact on the accuracy of facial recognition systems (FRS). Moreover, the use of face masks has made conventional facial recognition technology ineffective in many scenarios, such as face authentication, security check, tracking school, and office attendance, and unlocking phones and laptops.

Furthermore, the different algorithms that succeed on unmasked faces have been unable to generalize such successes on masked faces. One of the advantages associated with detecting an unmasked face is that the deep learning models would use the whole facial features/landmarks to identify someone. However, with a masked face, the nose and mouth are occluded. So, the problem of identifying individuals with just the eyes and sometimes, the forehead is more challenging [1]. Our purposed solution includes a masked facial recognition system so it can recognize the masked and unmasked people. As a result, people in enclosed spaces who need to verify their identity on mobile phones, laptops, or other devices, do not need to take off their face masks as the purposed solution can can recognize masked faces easily.

The main goal of our research paper is to perform real-time Masked Facial Recognition (MFR). The work is divided into three main parts to achieve the goal. The first part is data collection and preparation. We take the CASIA dataset [17] for training face images. The obtained images were not ready to use for training, so we performed various cleaning, alignment, and removal operation to make them ready for model training purposes. The CASIA dataset does not include masked faces, so we used the augmented method to generate masked faces for our dataset. Furthermore, the second part involved training the face recognition model that would be used for MFR. The training is done using the deep learning (Inception ResNet V1) model. We applied different hyperparameter functions for training and the model is evaluated using the LFW (Labeled Faces in the Wild) dataset [18]. The accuracy and loss functions were calculated in every epoch to validate the model. Also, we evaluated the three different trained models with five, ten, and fifteen training images per class. Lastly, the real-time

2022 IEEE World AI IoT Congress (AIIoT) | 978-1-6654-8453-4/22/$31.00 ©2022 IEEE | DOI: 10.1109/AIIOT54504.2022.9817324

978-1-6654-8453-4/22/$31.00 ©2022 IEEE

MFR is carried out using the previously trained model. We verified the robustness of the purposed solution for masked and unmasked facial recognition under various conditions like gender, skin tone, age, types of masks, etc. As a result, this work could be used for different purposes including security and safety of the people.

A. Facial Recognition System

Facial recognition is the process of automatically identifying an individual from captured images or videos [4]. It is a significant key area of research today. Its applications are becoming more important in various fields like ATM machines, criminal identification, access restriction, video conferencing, issuing drivers' licenses and passports, and monitoring the public areas. The imaging conditions, feature occlusion, inter-person similarity, and variance of faces are making the task of face recognition more challenging. Face recognition algorithms deal with a vast number of images, which leads to millions of operations, so it needs to have a specialized model for real-time implementation [3].

Moreover, various algorithms have been proposed for face recognition in the last few decades, with varying degrees of success. These algorithms analyze images and extract information such as shape, size, and location of facial features. So, the algorithms with the highest accuracy typically require intensive computation [4]. The facial recognition process usually has four interrelated steps shown in figure 1 [6].

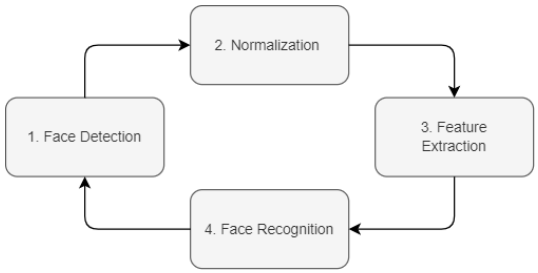

Fig. 1. Facial recognition process

In figure 1, the primary function of the first step is to detect the face from the captured image or the specified image from the database. This face detection process checks whether the image has a face image or not, after detecting the face, this output is further given to the preprocessing step. The second step performs the normalization of face recognition images. It removed the unwanted noise, blur, varying lighting conditions, and shadowing effects using different normalization techniques. Then we receive a fine smooth face image used for the feature extraction process. Additionally, features of the faces would be extracted as an embedding. It is applied to information packaging, dimension reduction, salience extraction, and noise-cleaning. After this step, the face patch usually turns into a vector with a fixed dimension or a set of fiduciary points and their corresponding locations [21]. Once the feature extraction is done, the last step analyzes the representation of each face which is used to recognize face identities for automatic face recognition. For any input images, the face is detected, and the embeddings are computed using facial features/landmarks. So, these embeddings are applied to calculate the Euclidean distance between the input image and the known database images. When the system detects input images, the fourth step first performs normalization, and feature extraction then computer the Euclidean distance for face matching. If the distance is below the threshold value, then it provides the name as an identity [6], otherwise, an unknown message is given.

B. Deep Learning

Deep learning is the pile of Convolutional Neural Network (CNN) layers and CNN is one of the most effective neural networks that has shown its superiority in a wide range of applications, including image classification, recognition, retrieval, and object detection. In addition, Neural Network (NN) is a sub-field and a key area of machine learning which are biological brain-inspired function approximators and have been successfully applied to various issues such as classification, regression, control, learning (online and offline), and robotics [2]. Furthermore, the neural network is an enormously powerful and robust classification technique that can be used for predicting not only the known data but also the unknown data. It works well for both linear and non-linear various datasets [19].

The NN has been used in multiple areas such as object detection, speech recognition, face recognition, fingerprint recognition, forecasting, and so on [5]. A standard feed-forward neural network is made up of multiple input layers of neurons, some of them hidden, and an output layer of neurons. A neuron is a basic part of a neural network. It processes signals by accepting them as an input and then outputs a signal using a function. Moreover, the NN receives information on the environment as a normal signal from its input layer and then outputs a signal through the output layer of neurons [2]. The general workflow of Neural Networks is shown in figure 2. In our research, we applied the CNN layer to train the MFR (Masked Face Recognition) model. It is challenging to obtain all the facial characteristics from a single layer so multiple CNN layers are utilized to extract various patterns of the face images. As a result, deep learning is significantly important to learn all the details of facial attributes.

In figure 2, the input layer receives information in the form of a numeric expression and transfers it to the hidden layers, which calculate the weighted sum and weights. The information is displayed as activation values, where each layer has given a number, the higher the number greater the activation. Additionally, this information is then transferred throughout the network. Based on the strength of the connection which are weights, inhibition, and transfer functions, the activation value is transmitted from layer to layer. Individual layers sum up the activation values it collects; then transform the value based on its transfer function [5]. Similarly, the Activation value goes via hidden layers through the network until it makes the output layer. The output layer then reflects the meaning. The neural

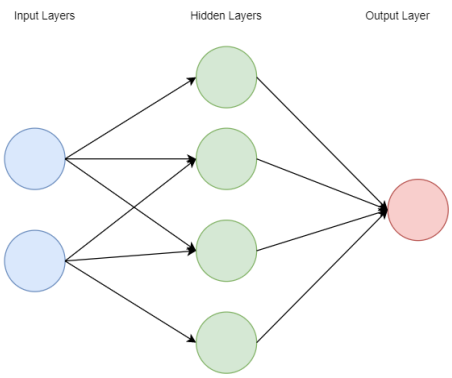

Fig. 2. The workflow diagram of NN

network could have many inputs, hidden, and output layers. There are several types of neural networks (NN) and some of them are recurrent neural network (RNN), convolutional neural network (CNN), and deep convolutional Network (DCN). The NN is applied in various fields such as computer vision, time series prediction, pattern recognition, robot control, anomaly detection, object detection, and so on.

In this section, we have presented the theoretical background of the research. It included our approach, motivation, and the work that has been done in this paper. Additionally, it discussed the Face Recognition System (FRS), Deep Learning (DP), Convolutions Neural Network (CNN), and their block diagrams with some details. We start in section two, by introducing the works that have been done in masked and unmasked face recognition. Section three outlines the different approaches we have applied to our work. There are data collection and preparation, model training, and real-time Masked Facial Recognition (MFR). Similarly, Section four describes the results of our work, experiment setup, performance evaluation, and comparisons of our solution against different methods. Lastly, section five concludes with the conclusion which is followed by future work in section six.

II. Related Work

This section discusses the related work and identifies what aspects of previous work will be applied. These studies focus on a few selected research that contributes to the use of facial recognition with and without face masks and how it can be implemented using different approaches. This section's primary goal is to provide a review of the related works and present an overview of the current research that contributes to the use of MFR (Masked Facial Recognition). Moreover, it describes some of the challenges in the development of facial recognition systems, their benefits, and the approaches that are applied to solve them.

A. Face Recognition Developing Stages

The earliest pioneers of facial acknowledgment were Woody Bledsoe, Helen Chan Wolf, and Charles Bisson. In 1964 and 1965, Bledsoe began working with computers with Wolf and Bisson to identify the human face. Due to the financing of the project coming from an unnamed intelligence agency, much of their work was never made public. However, later it turned out that their initial work applied the manual marking of different facial landmarks on the faces, such as eyes centers, nose, and mouth. These were then statistically turned by a computer to compensate for pose variation. Then the distance between the landmarks was calculated and compared between images to determine identity automatically [7].

These earliest steps into Facial Recognition in a manner consistent with the Bledsoe, Wolf, and Bisson were severely hampered by the technology of the era, but it remains an important first step in proving that Facial Recognition was a practical biometric. Carrying on from Bledsoe's original work, the baton was picked up in the 1970s by Goldstein, Harmon, and Lesk who expanded the work to include 21 specific subjective markers including hair color and lip thickness to automate the recognition. The National Institute of Standards and Technology (NIST) started Face Recognition Vendor Tests (FRVT) in the early 2000s. Building on FERET (Face Recognition Technology), FRVTs (Face Recognition Vendor Tests) were designed to provide independent government evaluations of commercially available facial recognition systems and prototype technologies. These assessments were designed to provide law enforcement agencies and the U.S. government with the information needed to determine the best ways to deploy facial recognition technology.

Back in 2010, Facebook began implementing facial recognition features that helped identify people whose faces may feature in Facebook photos that users update daily. The feature was immediately controversial with the news media, triggering a slew of privacy-related articles. However, Facebook users did not seem to mind. Having no clear adverse effect on the Web site's use or popularity, more than 350 million pictures are uploaded and tagged using face recognition every day. Facial Recognition technology has advanced rapidly from 2010 onwards and September 12, 2017, was another significant breakthrough for the integration of facial recognition into our day-to-day lives. This was the date that Apple launched the iPhone X, which was the first iPhone users could unlock with Face ID – Apple's marketing term for facial recognition. So, in this way, the development of facial recognition took place from past to present.

B. Masked and Unmasked Facial Recognition Systems

In 2015 Schroff et al. [15] proposed the FaceNet model, a unified embedding for face recognition and clustering. Despite significant recent progress in face recognition, implementation of face verification and recognition effectively poses serious challenges to current approaches. In this paper, the researcher presented a FaceNet model, which learns directly from face images to a close Euclidean distance where distances directly compare to a measure of facial matching. Once this space is created, it can easily implement tasks such as face identification, validation, and clustering using standard techniques as feature vectors with FaceNet embeddings. Their method applied a deep convolution network trained to optimize the

facial embedding, while the previous methods of deep learning used an intermediate layer of a bottleneck. A new triplet mining approach generated the non-matching and matching face patches to train. The main advantage of their approach was greater representation efficiency where they used only 128 bytes/face and obtained state-of-the-art facial recognition performance. In the massively used Labeled Faces in the Wild (LFW) dataset, their system reached the accuracy of 99.63 percent, a new record high. On the YouTube Faces database, it received 95.12 percent. Furthermore, their system shortened the error rate by 30 percent on both datasets as compared to the top published result of other papers [24]. Also, the researcher presented the notion of harmonic embeddings and a harmonic triplet loss, which represented various versions of face embeddings that were compatible with each other and authorized for direct comparison.

In 2021 Ullah et al. [1] presented a novel DeepMaskNet model for face mask detection and masked facial recognition. Testing people who are not wearing face masks manually in public places is a challenging task. Moreover, using face masks makes traditional face recognition techniques ineffective, typically designed for unveiled faces. Therefore, the researcher introduced a reliable system that can detect people who do not wear face masks and recognize different people while wearing face masks. In this paper, they proposed a novel DeepMasknet framework capable of both face mask detection and masked facial recognition. Moreover, presently there is an absence of a unified and diverse dataset that can be used to evaluate both face mask detection and masked facial recognition. For this purpose, they also developed a largescale and diverse unified mask detection and masked facial recognition (MDMFR) dataset to measure the performance of both the face mask detection and masked facial recognition methods. The proposed work has two main phases. The first phase includes the data collection and dataset preparation, while the second phase presents a novel Deepmasknet model construction for face mask detection and masked facial recognition. They got an accuracy of 100 percent for face detection and 99.33 percent for masked facial recognition. Researchers said that their experimental results on multiple datasets including the cross-dataset setting showed the superiority of their DeepMasknet framework over the contemporary models.

In 2015 Simonyan et al. [11] proposed very deep convolutional networks for large-scale image recognition. In this work, they studied the effect of the depth of the convolutional network on its accuracy in a large-scale image recognition environment. Their main contribution was a thorough evaluation of networks of increasing depth using an architecture with exceedingly small (3×3) convolution filters, which showed that a significant improvement on the prior-art configurations can be achieved by pushing the depth to 16–19 weight layers. These results were the basis of their 2014 ImageNet Challenge submission, where the research team secured first and second place in localization and classification tracks, respectively. They also showed that their representations generalize well to other datasets, where they achieve state-of-the-art results. To promote further research on the use of deep visual representations in computer vision, the researchers made two of the most powerful ConvNet models public.

In 2020 Mundial et al. [14] presented a paper on facial recognition problems in the covid-19 pandemic. The researchers proposed a methodology that can improve the existing facial recognition technology capabilities with masked faces. They used a supervised learning method to recognize masked faces together with in-depth neural network-based facial features. A dataset of masked faces was collected to train the Support Vector Machine classifier on a state-of-the-art Facial Recognition Feature vector. Their proposed methodology gave recognition accuracy of up to 97 percent with masked faces. They mentioned that this model performed better than existing devices not trained to handle masked faces.

In 2019 Ejaz et al. [16] presented the implementation of principal component (PCA) analysis on masked and non-masked face recognition. In this paper, a statistical procedure was selected that is applied in the recognition of the non-masked face and applied in the masked facial recognition technique. This method achieved an accuracy of masked face image recognition on average of 72 percent whereas non-masked face was on average 95 percent. PCA gave a poor recognition rate for masked face images rather than non-masked faces. It was found that extracting facial features from a masked face is less than a non-masked face because of missing features from masked faces. As a result, the researcher concluded that the PCA Analysis is better for normal face recognition but not for masked face recognition.

In 2020 Anwar et al. [12] presented masked face recognition for secure authentication. With the recent worldwide COVID-19, face masks have become an important part of our lives. People are encouraged to cover their faces when in public areas to avoid the spread of infection which can reduce the transmission of Covid-19. Face recognition system is commonly used for security verification purposes and the use of face masks has made conventional facial recognition technology ineffective in many scenarios, such as face authentication, security check, community visit check-in, tracking school, office attendance, and unlocking phones and laptops. Because of Covid-19, people in closed spaces must wear face masks to verify their identity on their mobile phones or laptops. Many organizations use facial recognition as a means of authentication and have already developed the necessary datasets in-house to be able to deploy such a system. Unfortunately, masked faces make it difficult to be detected and recognized, thereby threatening to make the in-house datasets invalid and making such facial recognition systems inoperable. As a result, the researcher addressed a methodology to use the current facial datasets by augmenting them with tools that enable masked faces to be recognized with low false-positive rates and high overall accuracy, without requiring the user dataset to be recreated by taking new pictures for authentication. They presented an open-source tool, MaskTheFace to mask faces effectively creating a large dataset of masked faces. The dataset generated with this tool is then used towards training an effective

facial recognition system with target accuracy for masked faces. They received an increase of around 38 percent in the true positive rate for the Facenet system. Additionally, the researcher tested the accuracy of the re-trained system on a custom real-world dataset MFR2 and report similar accuracy.

In 2021 Mandal et al. [13] proposed masked face recognition using ResNet-50. Over the last twenty years, there have seen several outbreaks of different coronavirus diseases across the world. These outbreaks often led to respiratory tract diseases and have proved to be fatal sometimes. Currently, we are facing an elusive health problem with the arrival of the COVID-19 disease of the coronavirus family. Airborne transmission is one of the modes of transmission of COVID-19 and it transfers when humans breathe, speak, sing, cough, or sneeze in droplets released by an infected person. As a result, public health officials have prescribed the use of face masks that can reduce disease transmission by 65 percent [13]. Facial recognition systems are used for security verification purposes and the use of face masks presents a difficult challenge since these systems were typically trained with human faces without masks but now due to the onset of the Covid-19 pandemic, they are forced to identify faces with masks. Therefore, the researcher studied the same problem by developing a deep learning model capable of accurately identifying face masks. In this paper, the authors trained a ResNet-50-based architecture that performs well at recognizing masked faces. The results of this study could be seamlessly integrated into existing facial recognition programs designed to detect faces for safety verification purposes.

III. Methodology

This section describes the three main parts: the first is data collection and preparation, the second is training the MFR (Masked Facial Recognition) model, and the last one is real-time Masked Facial Recognition. All three steps are explained in detail below.

A. Data Collection and Preparation

The first step of this research paper is to collect the dataset so that we can prepare the face images to train the model. For that, we used the CASIA dataset [17] which has 10585 classes, and each class has many images of the same person. We only required the facial part of each class image, however, the CASIA dataset has images with other attributes like hair, and body parts also. Therefore, the first step that we need to take is image alignment. Image alignment is a process of cropping the face from images that can represent the facial features. This is done by using the SSD (Single Shot Detector) [10] face detection model. We could use different face detection methods like MTCNN (Multi-Task Cascaded Convolutional Neural Networks), Dlib, and OpenCV but SSD is faster and easy to implement so we preferred that. At first, we detected the face of each image from each class and cropped only the face part and saved it to each class folder. Image alignment allowed us to reduce the size of each image which helped to improve the model training time and accuracy.

Moreover, the dataset class might include mislabeled images from other classes. For example, the class has faces that do not belong to that class, so our aim is to remove those images which affect the accuracy of the training model. This operation is done by using the Dlib [8], which finds out the facial landmarks of each face image. Furthermore, 128-dimensions of the facial embedding matrix of facial features are generated, and those numbers are used to describe the facial characteristics [15]. Then we calculated the average Euclidean distance between the target image and the reference image. And we removed the image from that class whose average distance surpassed the threshold value since we could say that image does not belong to this class. Additionally, if two faces are similar, then their average Euclidean distance always should be near zero. We set the threshold value of 0.8 to remove the outlier images. The overall data preparation process is shown in figure 3.

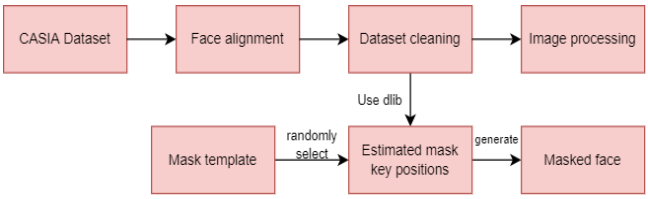

Fig. 3. Data preparation process

Lastly, the CASIA dataset has many classes, and each class has 15 to more than 100 images of the same person. Data imbalance is a big issue so to solve this problem we selected an equal number of images for each class. We randomly took five images from each class at a time and created more images of the same individual with different looks using the image processing method. It utilizes random crops, random noise, random mask, random angle, random flip, and random brightness methods. All these operations were performed by using OpenCV [9] and Dlib. So, this method helps to create a balanced number of images for each class. As a result, we used these balanced images to train our model. Moreover, the masked datasets are created using a computer vision-based approach (Dlib). Dlib is useful to locate the mask key position on the face using facial landmarks. So, these key positions of the face are replaced by a random mask temple to generate the masked face dataset. We applied sixteen mask templates, and one random mask is selected at a time. As a result, this approach helped to convert the CASIA dataset to a masked face dataset. It is difficult to collect the same person's images with a face mask and without a face mask, so this approach helps to convert any existing face dataset to a masked face dataset. As a result, all three steps were applied to prepare the training dataset. The process to create a masked face dataset is shown in figure 4.

B. Model Training

Our model training environment includes TensorFlow 2.1, Python 3.7, OpenCV 4.5, Matplotlib 3.5, Dlib 19.23, and

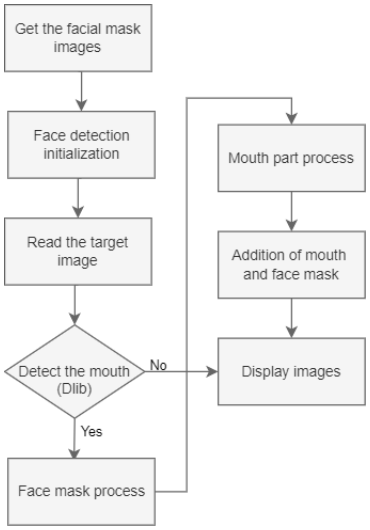

Fig. 4. Process to create masked face dataset

NumPy 1.21 of versions. The previously prepared CASIA dataset (with and without face mask) was used for training the model. The LFW [18] dataset was tested to evaluate the model. In each epoch, we calculated the accuracy and loss function of the training model with the testing dataset. It is noted that the accuracy was increased with the increase of epoch size in training. Accuracy is calculated by using a total number of correct predictions divided by a total number of predictions. Moreover, the average loss of the training model is calculated using Cross-Entropy and which is the difference between output probabilities and answers. While training the model, it was always expected to get higher accuracy and lower loss value. We used LFW face images that the model never learned to evaluate its real ability. The process involved in the modeling training is shown in figure 5.

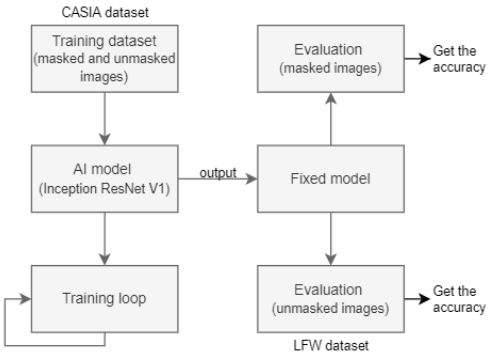

Fig. 5. Model training process for MFR

In the first step of figure 5, the training dataset includes both masked and unmasked images. Inception ResNet V1 [20] with training dataset is applied to train the MFR (Masked Facial Recognition) model. Inception ResNet V1 has many CNN (Convolutional Neural Network) layers to perform massive calculations and store all the image features. Moreover, training loops include the different epoch operations, where we set the epoch size and perform the training loop repeatedly to optimize the model. While training the model, in each epoch the accuracy and loss function are calculated using testing images and the Cross-Entropy function, respectively. The training model will save the embedding in the PB (Protocol Buffer) file, and it includes prediction and embeddings. Embedding is a facial feature that is transformed into a sequence of numbers, and these are used to describe facial characteristics. 128-dimension embeddings are used to represent the facial feature. It is expected to run the training loop until it reaches the set epoch value, and we would expect to get higher accuracy and less loss function. After the training epoch is done, the fixed model is saved in a local folder, and we utilize that fixed model to perform real-time Masked Facial Recognition.

1) Architecture of AI Model: The normal classification model only outputs the probabilities of trained class numbers, and it is impossible to train different faces all over the world. The very deep convolutional networks (Inception model) have been applied for facial recognition systems in the past and it has shown better performance and low computational cost [21]. The combination of Inception architecture [22] and residual network [23] (Inception ResNet V1) provides better recognition performance since training with residual connections accelerates the training of Inception networks. Therefore, Inception ResNet V1 architecture is proposed to use, and it provides the embeddings to perform face matching. Embedding is the numeric values of facial features which are used to recognize the person by calculating the Euclidean distance between the target and source image. The Inception ResNet V1 has weight layers, CNN (Convolutional Neural Network) layers, Average pooling, Stem, Reduction, fully connection (FC), and SoftMax. The parameters that are applied to Inception ResNet architecture are filter size, kernel size, the activation function (ReLU), batch size, learning rate, optimization method (Adam), strides, model size, and balance image size per class.

Inception ResNet [20] is a combination of the Inception and ResNet models where it has ten different steps to train the model. As figure 6 shows in the first step, the input is passed through the stem. The stem has five 3*3 convolutions of different filter sizes 32, 64, 80, 192, and 256, respectively. It has one 3*3 MaxPool of stride 2. Also, it includes 1*1 Convolution. The second step is 5 times Inception ResNet-A. It contains three 3*3 convolutions of filter size 32, and four 1*1 convolutions of filter size 32, and 256. Moreover, the third step is Reduction-A which includes the convolutions of 1*1, 3*3, and MaxPool of size 3*3. Similarly, it is followed by 10 times of Inception ResNet-B, Reduction-B, 5 times Inception ResNet-C, Average Pooling, Dropout, and the last one is Softmax. Average Pooling computes the average value of each feature map and returns that value. The embedding is received from Dropout once the normalization is done on it. We get the prediction value from Softmax whose output ranges from 0 to 1. Softmax can make the bigger number become a larger ratio and make the smaller numbers even smaller. As a

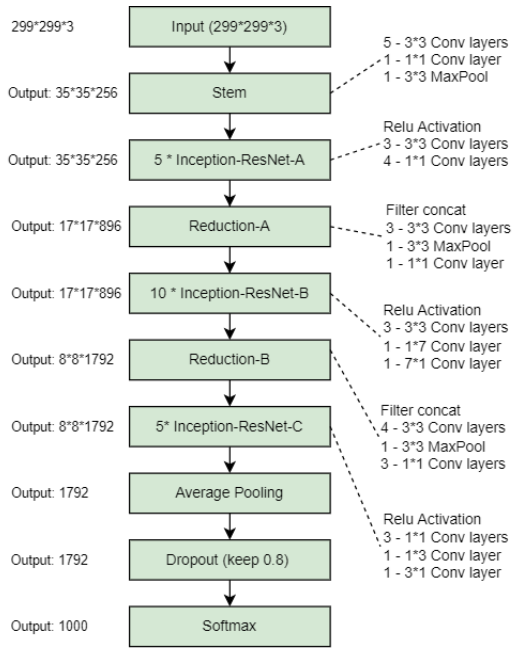

Fig. 6. Architecture of Inception ResNet V1 model [20]

result, our trained model includes embeddings and predictions.

C. Real-Time Masked Facial Recognition

We take the input images from real-time video streaming using a laptop camera. The face and face mask of each input frame is detected by the SSD (Single Shot Detector) model. Then the previously trained model, the Fixed model is used to find out the identity of input images. Additionally, the Euclidean distances are calculated by using embeddings of the input face and the face database. Among all the distances, we find out the smallest distance, and if the smallest distance is even smaller than the threshold value (0.8), then that would be an answer. Moreover, the distance should be around zero if two images belong to the same person. Therefore, the smallest Euclidean distance is compared with the threshold value to make the final facial recognition decision. If the distance is less than the threshold value (face matched), then the name of the person will be printed on the input image frame, otherwise, an unknown message is printed. Similarly, if the person is wearing a face mask, this system shows the "Mask" message, else "No Mask" message is displayed. The overall process of real-time Masked Facial Recognition (MFR) is shown in figure 7.

IV. Results and Discussion

This section describes the output of the experiments carried out on MFR (Masked Facial Recognition). It presents an in-depth explanation of various experiments meant to assess the effectiveness of our solution. Also, it highlights additional information about our experimental setup, prepared dataset, comparisons of our approach against other methods, and performance evaluation for different setups of environments.

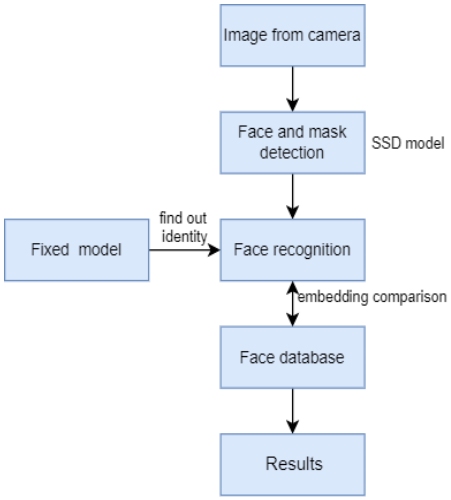

Fig. 7. Masked facial recognition process

A. Experiment Setup

Our experiments were carried out on a computer with an Intel Core i7 vPro processor, 16 GB of RAM, 500 GB SSD (Single Shot Detector) Hard disk, Windows 10 OS (operating system), and Nvidia GPU (Graphics processing units) card. Additionally, we used Python as a programming language, tools such as OpenCV, TensorFlow, CUDA, NumPy, and Matplotlib for image processing and model training, and Jupyter and PyCharm were utilized as an IDE (Integrated Development Environment). Additionally, CASIA [17] datasets and LFW (Labeled Faces in the Wild) [18] datasets were applied for training and testing, respectively. Both datasets are open-source images that are easily accessible online. Furthermore, TensorFlow is significantly important to perform the massive math calculations for building the model, and GPU helps to do the operation much faster. GPU has thousands of cores which can finish many calculations faster than CPU which normally has 8 cores. TensorFlow offers many models and relative functions and helps to communicate with GPU to do the task. Whereas OpenCV has rich image processing functions like reading, saving, resizing, displaying, cropping, transforming, and changing the color format of the images so it was applied for data preparation, image processing, and model training for Masked Facial Recognition (MFR).

B. Dataset and Performance Evaluation

We had to have both masked and unmasked face images to train our model. CASIA datasets are applied for unmasked images after image preprocessing, and we generated artificial masked images from the CASIA dataset using the mask augmentation method (Dlib). Both types of prepared face images are shown in figure 8. Moreover, we implemented the image processing method to make the same face images with different looks that utilized random crops, random noise, random angle, random flip, and random brightness methods.

Additionally, we trained the model with the imbalanced and balanced dataset, and it is studied that the significant

Fig. 8. Masked and unmasked training images

improvement in the accuracy is with the balanced dataset. As a result, we trained the three different models with five, ten, and fifteen training images from each class at a time, respectively. The parameters we used and the performance evaluation of the LFW (Labeled Faces in the Wild) dataset for three different trained models are presented in figure 9.

Item	1st Trained Model	2nd Trained Model	3rd Trained Model
Image type	masked & unmasked	masked & unmasked	masked & unmasked
Model shape	[N, 112, 112, 3]	[N, 112, 112, 3]	[N, 112, 112, 3]
Selected image number	15	10	5
Model	Inception ResNet V1	Inception ResNet V1	Inception ResNet V1
Loss function	Cross Entropy	Cross Entropy	Cross Entropy
Feacture number	128	128	128
Learning rate	0.0005	0.0005	0.0005
Epochs	38	50	61
Batch size	192	96	96
Augmentation times	4	4	4
GPU memory (GB)	8.6	4.6	4.6
Avg. epoch time (min.	64	45	24
Best accuracy	0.969	0.966	0.969
Epoch at best accuracy	34	41	53
Time best accuracy(hr)	37	31	22
GPU	Nvidia	Nvidia	Nvidia

Fig. 9. Performance evaluation for three different models

Figure 9 shows that the first trained model selected only fifteen images from each class and this model received the best accuracy of 96.9 percent in the 34th epoch. Similarly, the second trained model received only ten images for each class, and the third trained model selected only five images per class and obtained the best accuracy of 96.6 percent and 96.9 percent respectively. Moreover, the Inception ResNet V1 deep learning model is applied for all three training models. Each input face is augmented four times to generate more training images and Nvidia GPU (Graphics Processing Units) made the training process much faster since all the trained models received the best accuracy within 37 hours (about 1 and a half days) of training. It is observed that even if we used a smaller number of training images, we have achieved significantly better testing accuracy for the LFW dataset. Additionally, if we select a smaller number of training images, it is also easy to train the AI (Artificial Intelligence) model with normal GPU cards such as GTX 1660.

C. Comparison of Current Approach Against Other Methods

Figure 10 shows the comparison of our work with other different methods. Based on these results, we studied that our MFR (Masked Facial Recognition) method significantly outperforms the five other models. Additionally, we achieved an accuracy of 1.9 percent higher than the second-best performing model (Attention-based) and around 49 percent more than the worst performing model (ResNet-50) for masked and unmasked facial recognition. Our purposed model (Inception ResNet V1) is the combination of Inception architecture [22] and residual network [23] and it provides better recognition performance since training with residual connections accelerates the training of Inception networks.

Work Ref.	Model	Method	Dataset	Accuracy (best)
Purposed	Inception ResNet V1	MFR	CASIA, LFW	96.90%
[26]	FaceMaskNet-21	Deep metric learning	Collected dataset	88.92%
[1]	DeepMaskNet	CNN	MDMFR	93.33%
[13]	ResNet-50	Domain Adpation	Real World Masked Face Dataset	47.91%
[16]	PCA	Nearest Neighbor (NN)	ORL	73.75%
[25]	Attention-based	Face-eye-based	MFDD, RMFRD	95.00%

Fig. 10. Comparison of current approach against other methods

D. Results

We have created a system that can recognize both masked and unmasked face images. In our research, we implemented a new augmented way to create a masked face dataset and performed the image alignment and data cleaning using Dlib, OpenCV, and SSD (Single Shot Detector) model. Similarly, we applied the balanced face images from each class and received significantly better results than the imbalanced images trained model. LFW and face mask datasets were tested for evaluation, and they showed superiority over any contemporary models [1, 13]. Our evaluation model included the LFW dataset which is not used to learn the training model. Furthermore, we verified the robustness of our purposed model for masked and unmasked facial recognition under various conditions like gender, skin tone, age, types of masks, etc. As a result, we achieved the MFR (Masked Facial Recognition) of over 96 percent accuracy for our three different trained models. The results of our experiment are shown in figure 11.

V. Conclusion

In conclusion, this research paper has presented a solution to identify the masked and unmasked faces accurately. The proposed approach provided over 96 percent accuracy for MFR (Masked Facial Recognition). Furthermore, the masked face dataset was created using a computer vision technique. CASIA datasets were used to train the model after performing image preparation and the LFW (Labeled Faces in the Wild) dataset was tested to evaluate the performance of the model. Also, the performance of three different models has been studied for

Fig. 11. Masked facial recognition results

MFR. Additionally, we verified the robustness of our purposed model for masked and unmasked facial recognition under various conditions like gender, skin tone, age, types of masks, etc. As a result, the purposed solution could be seamlessly integrated for both masked and unmasked face recognition and detection systems that are designed for safety and security verification purposes without any challenges.

VI. Future Work

In the future, we intend to use the real-time mask face dataset since some of our generated masked images do not perfectly fit the rotated faces, so using the real masked images could increase the recognition accuracy of the system. Also, it is expected to increase the number of balanced images for each class to train the model for better quality and diversity (we applied a maximum of 15 faces per class in our experiment). Additionally, we would try to build a small facial recognition model which could improve the overall recognition rate of the system.

References

[1] Ullah, Naeem, et al. "A novel DeepMaskNet model for face mask detection and masked facial recognition." Journal of King Saud University-Computer and Information Sciences (2022).

[2] Mason, Karl, Jim Duggan, and Enda Howley. "A multi-objective neural network trained with differential evolution for dynamic economic emission dispatch." International Journal of Electrical Power and Energy Systems 100 (2018): 201-221.

[3] Kumar, A. Pavan, V. Kamakoti, and Sukhendu Das. "An Architecture for Real Time Face Recognition Using WMPCA." ICVGIP. 2004.

[4] Soyata, Tolga, et al. "Cloud-vision: Real-time face recognition using a mobile-cloudlet-cloud acceleration architecture." 2012 IEEE symposium on computers and communications (ISCC). IEEE, 2012.

[5] Kasar, Manisha M., Debnath Bhattacharyya, and T. H. Kim. "Face recognition using neural network: a review." International Journal of Security and Its Applications 10.3 (2016): 81-100.

[6] Agagu, T. T., and B. A. Akinnuwesi. "Automated students' attendance taking in tertiary institution using hybridized facial recognition algorithm." Journal of Computer Science and Its Application 19.2 (2012): 1-13.

[7] de Leeuw, Karl Maria Michael, and Jan Bergstra, eds. The history of information security: a comprehensive handbook. Elsevier, 2007.

[8] Dlib, https://github.com/davisking/dlib

[9] OpenCV, https://github.com/opencv/opencv

[10] Liu, Wei, et al. "Ssd: Single shot multibox detector." European conference on computer vision. Springer, Cham, 2016.

[11] Simonyan, Karen, and Andrew Zisserman. "Very deep convolutional networks for large-scale image recognition." arXiv preprint arXiv:1409.1556 (2014).

[12] Anwar, Aqeel, and Arijit Raychowdhury. "Masked face recognition for secure authentication." arXiv preprint arXiv:2008.11104 (2020).

[13] Mandal, Bishwas, Adaeze Okeukwu, and Yihong Theis. "Masked face recognition using resnet-50." arXiv preprint arXiv:2104.08997 (2021).

[14] Mundial, Imran Qayyum, et al. "Towards facial recognition problem in COVID-19 pandemic." 2020 4rd International Conference on Electrical, Telecommunication and Computer Engineering (ELTICOM). IEEE, 2020.

[15] Schroff, Florian, Dmitry Kalenichenko, and James Philbin. "Facenet: A unified embedding for face recognition and clustering." Proceedings of the IEEE conference on computer vision and pattern recognition. 2015.

[16] Ejaz, Md Sabbir, et al. "Implementation of principal component analysis on masked and non-masked face recognition." 2019 1st international conference on advances in science, engineering and robotics technology (ICASERT). IEEE, 2019.

[17] CASIA dataset, https://github.com/SamYuen101234/MaskedFaceRecognition

[18] LFW dataset, http://vis-www.cs.umass.edu/lfw/

[19] Alzu'bi, Ahmad, et al. "Masked Face Recognition Using Deep Learning: A Review." Electronics 10.21 (2021): 2666.

[20] Szegedy, Christian, et al. "Inception-v4, inception-resnet and the impact of residual connections on learning." Thirty-first AAAI conference on artificial intelligence. 2017.

[21] Rath, Subrat Kumar, and Siddharth Swarup Rautaray. "A survey on face detection and recognition techniques in different application domain." International Journal of Modern Education and Computer Science 6.8 (2014): 34.

[22] Szegedy, Christian, et al. "Rethinking the inception architecture for computer vision." Proceedings of the IEEE conference on computer vision and pattern recognition. 2016.

[23] He, Kaiming, et al. "Deep residual learning for image recognition." Proceedings of the IEEE conference on computer vision and pattern recognition. 2016.

[24] Sun, Yi, Xiaogang Wang, and Xiaoou Tang. "Deeply learned face representations are sparse, selective, and robust." Proceedings of the IEEE conference on computer vision and pattern recognition. 2015.

[25] Wang, Zhongyuan, et al. "Masked face recognition dataset and application." arXiv preprint arXiv:2003.09093 (2020).

[26] Golwalkar, Rucha, and Ninad Mehendale. "Masked-face recognition using deep metric learning and FaceMaskNet-21." Applied Intelligence (2022): 1-12.