International Journal of Recent Technology and Engineering (IJRTE)
ISSN: 2277-3878 (Online), Volume-12 Issue-5, January 2024

Security-oriented Face Detection Technology Utilizing Deep Learning Techniques Along with the CASIA Datasets

Iqra Yamin, Yang Gaoming, Marcel BAKALA, Muhammad Asad Yamin, Usama Masood

Abstract*: Recently, face recognition technology has become increasingly important for safety purposes. Masks are now required in most countries and are increasingly used. Public health professionals advise people to conceal their facial features outdoors to reduce COVID-19 transmission by 65%. Detecting people without masks on their faces is crucial. This has become widely used as face recognition outperforms PINs, passwords, fingerprints, and other safety verification methods. Sunglasses, scarves, caps, and makeup have made facial identification harder in recent decades. Thus, such masks impact facial recognition performance. Face masks also make traditional technology for facial recognition ineffective for face authorization, security checks, school monitoring, and cellphone and laptop opening. Thus, we proposed Masked Facial Recognition (MFR) to recognize veiled and exposed-face people so they don't need to remove their masks to verify themselves. This deep computing model was trained with Inception Res Network V1. CASIA is responsible for preparing pictures and using LFW to validate models. Dlib creates masked datasets utilizing vision algorithms. About 96% accuracy was achieved using our three models that were trained. Thus, covered and uncovered recognition of faces and detection techniques in security and safety verification might easily be used. These systems can be used in various settings, such as airports, train stations, and other public places, to enhance security and prevent crime. Overall, deep learning within face recognition technology has significant potential for improving safety and security in various settings.

***Keywords*: CASIA Dataset, Dlib, face recognition, Masked Facial Recognition.**

Manuscript received on 08 November 2023 | Revised Manuscript received on 17 November 2023 | Manuscript Accepted on 15 January 2024 | Manuscript published on 30 January 2024.

*****Correspondence Author(s)*

Iqra Yamin*, Department of Computer Science and Engineering, Anhui University of Science and Technology, Huainan (Anhui), China. E-mail: Iqrak6113@gmail.com, ORCID ID: 0009-0009-8126-1934

Yang Gaoming, Department of Computer Science and Engineering, Anhui University of Science and Technology, Huainan (Anhui), China. E-mail: gmyang@aust.edu.cn, ORCID ID: 0000-0002-7666-1038

Marcel BAKALA, Department of Computer Science and Engineering, Anhui University of Science and Technology, Huainan (Anhui), China. E-mail: bklmarcel@111.com, ORCID ID: 0009-0004-1597-2897

Muhammad Asad Yamin, Department of Computational Science and Engineering, University of Rostock Germany. E-mail: sam_1219@qq.com, ORCID ID: 0009-0006-8559-0645

Usama Masood, Department of Mechanical Engineering, Anhui University of Science and Technology, Huainan (Anhui), China. E-mail: amnaiqbal.1298@gmail.com, ORCID ID: 0009-0004-1710-0208

I. INTRODUCTION

As a result of COVID-19, face mask usage is expanding quickly; everyone inside and outside of buildings must consistently wear facial coverings to mitigate the transmission of the illness. To prioritize the well-being and security of all individuals, it is imperative to accurately discern the presence of individuals using facial masks. Face detection refers to the computational process of finding and localizing a face within an image or a pre-defined image within a database. On the other hand, face recognition involves the automated identification of an individual based on collected footage or photos [1]. The ongoing study on this subject holds substantial significance as it has growing relevance in several industries, such as ATMs, illegal identification, controlling entry, webinars, passport and license provision, and outdoor monitoring.

This security system is becoming more sophisticated, causing significant changes in our daily lives. As a result, the Safety system contains a critical regulation to protect humans. Regarding applications in practice, cover face recognition is one concerning the study fields. It could be used for detection in criminal investigations [2]. Due to the prevalence of facial masks in this particular case, a specific area is protected by surveillance footage for the site's safety [3]. It watches persons who did not wear masks to sensitive regions and compares their shots to database images to define the individual's exposure to the location [4][25][25][27][28]. Many techniques for recognizing masked faces, including varied perspectives [5], are known. This work focuses on developing an algorithm for cover facial recognition utilizing various CNN learners. Mask face detection is faster than alternatives for safety concerns, considering numerous faces can be studied or identified during the precise period. Working with CNN provides more preciseness in detecting the mask face in a particular region, and it is more effective whenever the individual's face enters a specific area of a live camera, so it is swifter over others. Despite the reality that facial identification studies served as a foundation for studies about masked facial recognition, it could be more effective. Reason detection focuses on distinguishing characteristics that prove it represents a face, while the mask obscures half or a more significant portion of the face features. As an outcome, this is developed further to study research models that may determine veiled characteristics.

Additionally, using face masks renders standard facial recognition technology useless in various situations, including face authorization, security checks, monitoring attendance at places of employment and education, and unlocking phones and laptops. Moreover, the diverse algorithms have exhibited limited ability to extend their achievements from unveiled facial recognition to veiled facial recognition. One advantage of facial recognition without a mask is that deep learning models can utilize several facial landmarks and characteristics to identify an individual accurately. The nose or lips hide when a person's face is veiled. Therefore, it is more difficult to recognize people based solely on their eyes and, occasionally, their forehead [6]. As a result, both masked and unmasked faces could be identified using the proposed approach.

II. RELATED WORK

Extensive research was undertaken on the subject of face mask-detecting systems. In their study, Das et al. [7] propose an approach to uncover this answer by integrating Scikit-Learn, Keras, TensorFlow, OpenCV, and other foundational machine-learning techniques. The proposed methodology effectively identifies the facial features depicted in the image and assesses whether a mask is present. Furthermore, it can identify both a human face and the presence of a cover in dynamic scenarios during the execution of an observational assignment. Accordingly, the approach demonstrates a maximum accuracy of 95.77% and 94.58% over a pair of separate datasets. The Sequentially Convolutional Neural Network (CNN) is employed to analyze optimal parameter values for successfully detecting mask proximity while mitigating overfitting.

A deep learning-based technique for identifying mask recordings is presented in [8] by Aniruddha Srinivas Joshi et al. The approach described in this study utilizes the MTCNN face identification algorithm to showcase its capability in accurately discerning faces and identifying their corresponding facial characteristics inside a given video clip.

Three different types of masked face identification datasets are provided by Adnane Cabania et al. [9], along with a technique for altering photos. This method combines datasets containing correctly and incorrectly masked faces.

A face detection strategy identifies a visage from a photograph containing multiple attributes. Facial recognition, facial monitoring, and position computation are all necessary for a successful face detection inquiry [10]. The goal is to identify a human face from a single photograph. Recognizing faces is difficult since facial features vary in size, shape, color, and more. The procedure is more challenging when dealing with obscure photos, such as those in which the subject is outside the camera's field of view. According to the authors of [11], occlusive identification of faces presents two significant challenges:

1) Absence of massive datasets with veiled and unmasked traits,

2) The lack of gestures inside the closed region

The locally linear embedding (LLE) method and dictionaries learned on a very large set of masked faces can be used to combine unimportant facial features to bring back many lost expressions and make facial signals less important. Convolutional neural networks (CNNs) impose stringent constraints on the dimensions with input images, as scholars specialized in computer vision [12] observed. The photographs are typically reconfigured before being added to the network to remove the obstacle.

The main difficulty of this exercise is determining whether or not the human face is wearing a mask in the photograph. The proposed method must identify a moving look and a show to perform surveillance tasks.

III. METHODOLOGY

A. Obtaining and Creating Data

Figure 3.1 depicts the whole strategy for data gathering and preparation. The following sections will go into the specifics of data collection and processing.

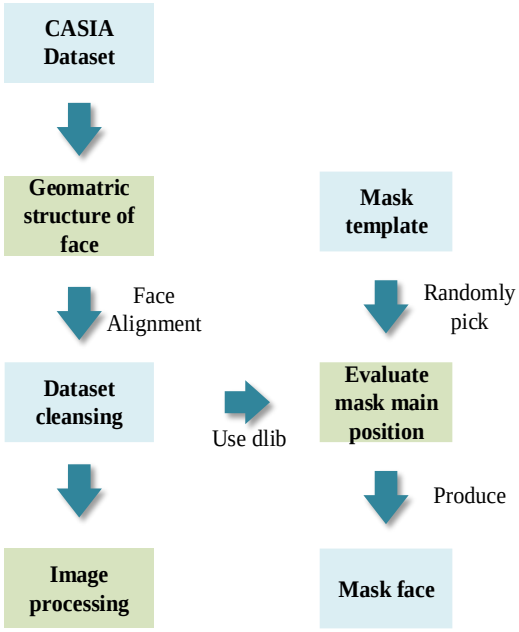

Figure 3.1 Process for Preparing Data

To facilitate the algorithm's training process, the first phase in this investigation endeavor is the acquisition of a dataset of facial photographs. The CASIA dataset [13] was employed in this study, consisting of 10,585 categories. Each category contains photographs, ranging from less than (10) to over 100 images depicting an identical subject. The training images employed for data preparation are limited to PNG, VMP, and JPG formats. In comparison, the CASIA dataset encompasses images with supplementary attributes such as human hair, neck region, and shoulders. Our specific requirement was confined to the facial area for every photograph. The initial stage in extracting the facial part of the photographs is employing image alignment techniques. Multiple datasets were generated for training (3) unique designs, each with varying quantities of training photographs. The specific details of these datasets and their corresponding models may be seen in Table 3.1.

Table 3.1 All three CASIA Training Databases Utilized in this Study Have Been Succinctly Summariz

Dataset	No. of Classes	For each class No. of Images	Improved	No. of Masked images for each class	For each class No. of Unmasked images	Entire Images
Dataset 1	10585	5	4	10	10	211700
Dataset 2	10585	10	4	20	20	423400
Dataset 3	10585	15	4	30	30	635100

B. Performing Face Alignment

The face alignment technique involves cropping the face in an image, leading to a cropped photo representing its facial characteristics. Face recognition is used as part of the process. Identification of faces is the process of finding a look in a picture that has been taken or in a set of photos that have already been taken. In this case, the SSD (Single Shot Detector) face detection method [14] is employed for image alignment. The SSD method was chosen for visual alignment because it works better and is easier to use than other face recognition algorithms like MTCNN, OpenCV, and Dlib. The first step involved identifying each face in the CASIA class, then cropping and storing the recognized facial component in the appropriate class folder. The outer edges of the resized picture are uniformly adjusted to a value of 20 pixels in every one of the four main axes. This deliberate adjustment ensures that the resulting cropped face encompasses all relevant facial features while minimizing the presence of extraneous background elements. Comparably, we modified to ensure all images conformed to the pattern [112, 112, 3]. At this point, the initial two figures (112) represent the width and the height dimensions, respectively, while the final one (3) denotes the number of channels. The training consists of graphics produced in the Red, Green, and Blue (RGB) colour model. Implementing picture alignment techniques has enhanced the effectiveness and preciseness of the learning model by lowering the size of a picture. Figure 3.2 displays a collection of sample photographs wherein the facial features have been correctly positioned.

Figure 3.2 Pictures that Were Shot After the Faces Were Aligned

C. Clearing the Data

The CASIA categories may contain incorrectly identified photographs originating via other types. If the scenario that the category includes facial images that are not representative of the class itself, our objective is to exclude photographs that compromise the precision of the learning models. Figure 3.3 depicts the procedure. We cannot manually remove incorrectly labeled photographs from every folder since it would take too long. As a result, FaceNet pre-trained weights are used [15]. The reference photos were chosen one at a time after the target photos, which were chosen initially. Thus, to acquire 128-dimensional structure embeddings across the target and reference pictures, we use the FaceNet framework. Calculating an average figure involves determining the distance measured by Euclid between the target picture and the reference picture. The expression below demonstrates the mathematical approach used to calculate the distance Euclid defines. We eliminated a target image from the class if the average length was more significant than the cutoff point of 0.8. As a result, the target image does not fall under this category. The average Euclidean distance between two faces should always be close to zero if they are comparable. To eliminate the outlier photos, we established a threshold value of 0.8.

Euclidean distance (d)

$$= \sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2 + (x_3 - y_3)^2 + \ldots\ldots + (x_{128} - y_{128})^2}$$

The coordinates representing the initial positions of the item targeted, along with the picture references, are designated at x and y, respectively. The computation for the Euclidean distance commences with the initial value and progresses sequentially until the 128th value, as the embedding encompasses 128 directions.

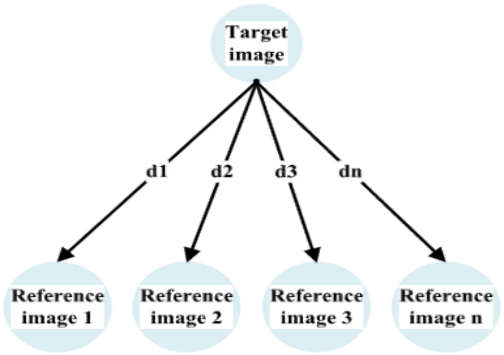

Figure 3.3 Method for Determining the Separation in the Middle of the Target Plus Reference Pictures

D. Generate the Masked Faces

The generation of masked face datasets is facilitated using a computer vision technology called Dlib. It can determine the mask's crucial placement on the face by referencing known landmarks. It features 68 facial markers, 48-68 of which represent the mouth. Creating veiled face shots involves substituting some areas of interest (ROI) within a person's face with a randomly selected mask template from a collection of Sixteen mask photographs. To ensure proper alignment with the mask onto a facial image, we modified the face mask design to correspond with the dimensions of the region of interest (ROI) encompassing the mouth. Furthermore, because PNG files possess four channels, with the fourth channel specifically designated for transparency, keeping all mask layouts in the PNG file style is customary.

At any given moment, a mask is selected randomly from a pool of (16) mask designs that have been utilized. Figure 3.4(b) shows an example of an image used as a mask template. We used this technique to transform the CASIA dataset into a masked face dataset to train our Masked visage recognition system further. This methodology facilitates the transformation of an existing facial dataset into a dataset containing masked faces, as it presents difficulties in capturing photographs of the same individual regardless of whether they are wearing a mask on their face. Figure 3.4 (a) illustrates the process of building a face masking dataset. The present study employed a strategy to facilitate the conversion of the CASIA dataset as a mask face dataset, hence contributing to the training of our Masked visage recognition model. This approach enables the transformation of an existing facial dataset into a dataset containing masked faces. This is due to the inherent difficulty in capturing photographs of the same individual without or with a mask covering their face. The process of generating a dataset of masking visages is illustrated in Figure 3.4 (a).

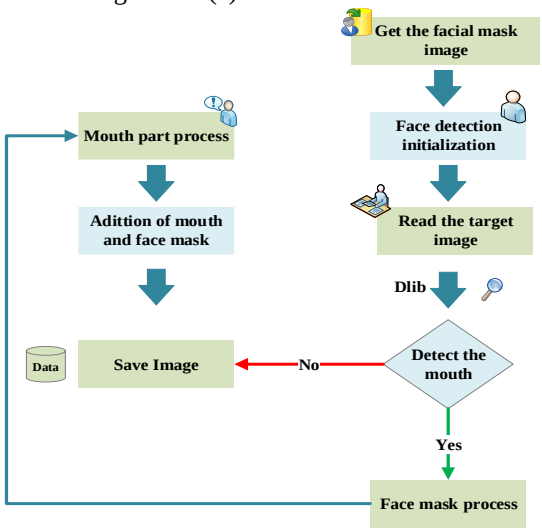

Figure 3.4 (a) The Process of Generating A Masked Face

Figure 3.4 (B) Images for The Mask Template

E. Build a Balance Dataset

There are numerous classes in the CASIA dataset, with fewer than ten to more than 100 identical photos in each category. Data imbalance harms the trained model's accuracy, a severe issue [16]. We chose various images for each class, as indicated in Table 3.1, to tackle this problem. To create more photographs of the same person with multiple expressions, we used the image processing technique to randomly select 5, 10, and 15 images from each class. We added four different augmentations to each chosen picture. Thus, creating four additional images—two with masks and two without from just a single image is possible. This approach incorporates several techniques, such as arbitrary cropping, randomized brightness adjustments, random masks, arbitrary blur, arbitrary flip, and arbitrary angles. The activities were performed using Dlib [17] and OpenCV [18]. Therefore, employing this approach facilitates the generation of an equitable quantity of photographs for every category. To reduce the training process of our model, we used a set of equitably distributed photos. Table 3.1 presents an overview of the three separate datasets generated to train each model, each utilizing different training photograph sizes. The quantity of veiled and revealed photographs used for every training model has been presented. The training samples' augmented images are displayed in Figure 3.5. Similarly, this part provides the source code for image processing and argumentation.

Figure 3.5 Sample Augmented Images from Randomly Chosen Photographs

F. Training the Model

The training dataset in the first stage of Figure 3.6 consists of masked and unmasked photos. Training models involve utilizing the complete set of facial characteristics in unmasked faces. Still, in the case of masked faces, the features extracted are limited to the forehead regions, eyes, and brows. The facial features of individuals wearing masks were concealed, specifically their mouths, noses, and cheeks, allowing only the exposed regions to be utilized to obtain components. MFR (Masked 35 Facial Recognition) is trained with the Inception ResNet V1 [19] architecture and a specific dataset. Using multiple CNN layers within Inception, ResNet V1 makes it easier to do extensive calculations while keeping all visual details. Additional information regarding our AI model can be found in the subsequent section. Training loops encompass many epoch iterations to enhance the model's performance, wherein the epoch duration is specified, and the training cycle is executed iteratively. During the model's training process, the precision and loss coefficients are computed for every epoch by utilizing testing images and the Cross-Entropy function. The training model weights were iteratively improved at every epoch, along with the algorithm's precision and elimination function computation. The epoch is defined as a constant value.

Notably, an increase in training epoch duration corresponded to a concurrent improvement in precision.

Additionally, the ratio of the total number of accurate estimates to the overall number of calculations is a measure of accuracy. In addition, Cross-Entropy is employed to calculate the average degradation of the learning model by quantifying the disparity between the anticipated probability and the observed outcomes. The training of the model consistently expected improved accuracy and reduced loss values. The process of modeling training is depicted in Figure 3.6, illustrating the sequential phases involved. The following equations demonstrate the methodology for calculating the loss function and its accuracy.

$$\text{Accuracy} = \frac{\text{Number of the correct pridiction}}{\text{Total number of the prediction}}$$

$$\sum_{C}^{M} = 1 y_{i,c} \log(p_{i,c})$$

The label is represented by y, and the prediction is denoted by p.

Face-matching accuracy is indicated by a one-to-one correspondence between the forecast and the response. We used this value to determine the trained model's accuracy for each epoch. We do not use the prediction probability our trained model provides for face matching, although we do obtain it. For face matching, we substituted embeddings for predictions.

Our trained model created embeddings for any input photos, which were then used to determine the Euclidean distance. A facial trait is embedded when converted into a series of numbers that are then used to define the feature. Despite the availability of embeddings with lengths ranging from 64 to 512, we have decided to utilize 128-dimensional embeddings to represent facial characteristics. The increased size of the embed will necessitate more computational time, while a reduced size may result in the exclusion of some facial parts. As a result, 128-d embedded data are used in our research for displaying the face feature. It is anticipated that by prolonging the training loop until the desired epoch value is achieved, there will be an improvement in precision and a decrease in the loss function. Real-time veiled facial recognition is performed using the fixed model utilized during the training phase and stored in a designated local directory. A PB (Protocol Buffer) document will keep the model trained, consisting of embeddings and foresight.

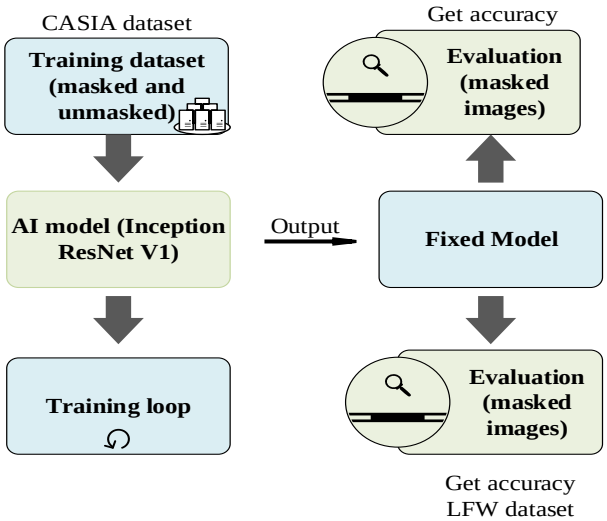

Figure 3.6 The Model Training Procedure for MFR

G. Enhancement of Model

We enhanced our training model in each epoch by supplying the same images with a distinctive appearance. Consequently, using photo processing and computer vision methods, arbitrary filters, arbitrary cropping, arbitrary blurring, randomized angles, arbitrary flipping, and randomized brightness are applied. Also, we sent every picture to the training dataset, containing over 10,000 categories with different ages, ethnic groups, and sexes, to ensure the model will learn from a range of face pictures. Furthermore, a picture alignment was performed across the training and testing images, enabling us to compress every picture while retaining only the facial portion. As a consequence, the training model worked better and with greater accuracy. Furthermore, because our model cannot learn from inaccurate photos, eliminating the incorrectly identified image from every class served to increase the system's precision. Examining [20] taught us that selecting an unequal number of photographs from each category can negatively impact the training model's accuracy. Consequently, we ensured that an equal number of pictures were chosen from every class across all three prepared models. The model's training time was lowered since we used a smaller batch size (96,192 images per loop). Thus, all three models averaged 30 hours for maximum accuracy. We selected tiny train photographs having 112*112 height and breadth and reduced the filter dimensions by 50% to help reduce the sizes of the training models. Compact training models are of utmost importance as they significantly enhance the speed of recognition accuracy and require a shorter period for inference. The tiny variation of the model exhibits compatibility with larger server-side models and may be effectively utilized on mobile devices as well.

H. The Inception ResNet V1 Architecture

The artificial intelligence (AI) algorithm learns with input photographs produced beforehand. Weight layers, softmax average pooling, reduction, stem, fully connected (FC), and CNN (Convolutional Neural Network) layers have all parts of Inception ResNet V1. The Inception ResNet framework incorporates many variables and techniques to enhance its performance. The factors above encompass the implementation of filter measurements, precisely 32, 64, 80, 192, along with 256, a kernel dimension of 3*3, the utilization of the ReLU activation function, batches with sizes of 32, 96, along with 192, an initial training rate about 0.0005, the application of the Adam optimization technique, strides, a predetermined number of epochs, model measurements, and the distribution of images from a diverse range of 10,585 categories. They are improving training efficiency by increasing batch size. Due to the limitation of processing capacity, it is necessary to divide the training images into batches and input them sequentially throughout every iteration. The number of iterations inside each epoch is determined by dividing the number of training pictures by the number of sets. Table 3.2 displays the values we used during training for the three models.

Table 3.2, Parameters used During Training for Our Three Different Models

Parameters	Values		
	Model. I.	Model. II.	Model. III.
Model Shape	[112,112,3]	[112,112,3]	[112,112,3]
Feature number	128-d	128-d	128-d
Learning rate	0.005	0.005	0.005
Batch size	192	96	96
Maximum epoch	38	50	61
Loss	Cross Entropy	Cross Entropy	Cross Entropy
Masked images	317,550	211,700	105,850
Unmasked images	317,550	211,700	105,850
Optimizer	Adam	Adam	Adam

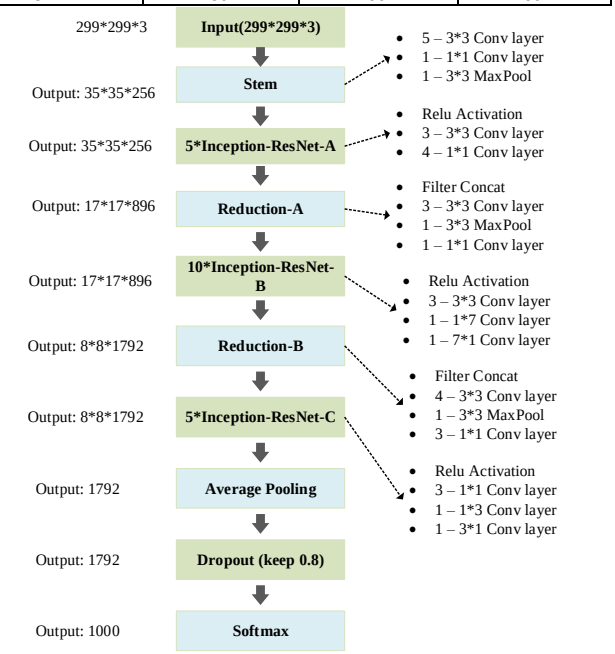

Figure 3.7 Inception ResNet V1 Model Architecture

I. *Embedding generation techniques*

Using a Euclidean distance among targets along with reference photos is combined with characteristic embedding, which quantifies numerical values of facial characteristics to ascertain a person's individuality. Furthermore, although they can generate embedding with 64, 128, 256, or 512 dimensions, they opt for utilizing 128-dimensional embedding to represent facial characteristics. Although the increased dimensions of the embedding need additional computational time, the reduced sizes may result in some facial parts' exclusion. For representing facial features in our studies, 128-dimensional embeddings are utilized. After normalization, our model receives embedding from Dropout, which is 128-d in size. Figure 3.8 depicts the steps involved in obtaining the embedding.

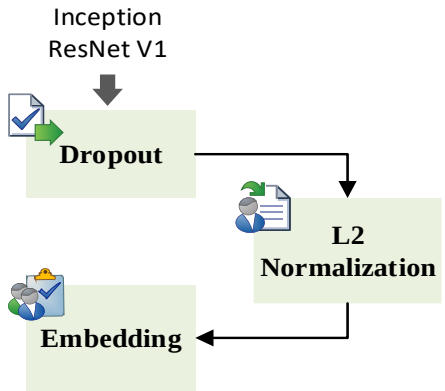

Figure 3.8: Method for Obtaining the Embedding

The following is the source code to create embedding

```
prelogits, _ = inception_resnet_v1(tf_input, tf_keep_prob, phase_train=tf_phase_train,
bottleneck_layer_size=embed_length, weight_decay=0.0, reuse=None)
prelogits = tf.identity(prelogits,name='prelogits')
embeddings = tf.nn.l2_normalize(prelogits, 1, 1e-10, name='embeddings')
```

J. Face Matching via Embedding

Initially, it is vital to ascertain how the input photographs and face information are incorporated. Subsequently, employing all of the facial database embeddings, compute the Euclidean distances among the provided embedding. The image entered is referred to as the target picture, while the photograph retrieved from the face databases is a reference picture, as depicted in Figure 3.9. Similarly, as shown in Figure 3.9, we used face embedding to determine the Euclidean distance between each target and reference image. If the minimum distance computed from a group of distances is less than some fixed threshold number (in this case, 0.8), then that distance is the solution.The threshold value is compared with the shortest Euclidean distance to determine whether the two faces match. The equation below illustrates how to compute the Euclidean distance.

Euclidean distance (d)

$$= \sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2 + (x_3 - y_3)^2 + \ldots\ldots\ldots + (x_{128} - y_{128})^2}$$

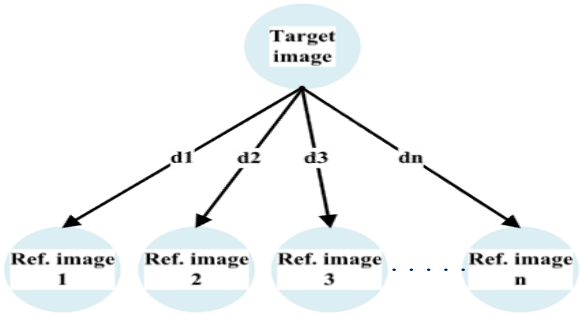

Figure 3.9 Between the Target and Reference Pictures, Determines their Euclidean Distance

K. Flowchart for Evaluating an Unmasked Chart

A sample of three thousand similar face pairings and another example of three thousand face combinations that differ from one another were selected from the LFW dataset, as seen in Figure 3.10. A total of 6000 pairs of faces were utilized to assess the model. The "Fixed model," which has undergone training, is employed to compute an embedding of each team. To determine the Euclidean distance, an embedding pair is used. Furthermore, in the case of the identical facial pair, the gap is evaluated against the predetermined threshold value of 0.8. Suppose the hole is found to be lower than the threshold value. In that case, the tally for accurate predictions on the correct side increases by 1The distance that lies among every set of faces has been roughly measured and contrasted against a threshold of 0.8.

If the estimated length equals or exceeds the minimum value, the count of accurate forecasts is incremented by one.

Similarly, calculating the accuracy rate involves dividing the aggregate number of correct predictions by the total count of facial combos, which amounts to six thousand in the present illustration. Using unmasked face photos, we used this method to assess our trained model. Figure 3.10 depicts the entire workflow for evaluating an unmasked face.

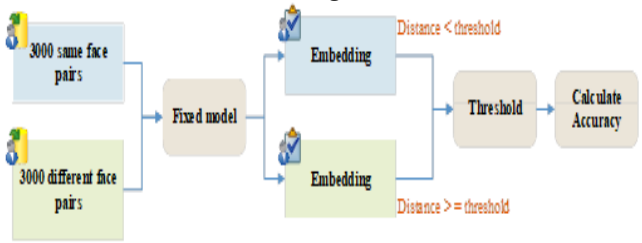

Figure 3.10: Procedure for Unmasking A Face for Evaluation

L. Flowchart for Evaluating Masked Face

Figure 3.11 (b) illustrates how we chose a face database and artificially masked photos (tar_images and ref_images) to test our model. As indicated in Figure 3.4, we used a computer vision technique to produce artificially veiled facesThe dataset utilized in this study was constructed by employing the Microsoft facial database, which encompasses a combination of artificially hidden faces and real-world facial databases. Table 4.2 provides additional information about the evaluation data. The "Fixed model" we have employed can embed tar_images along with ref_images. For the project to proceed, it is necessary to obtain the embeddings of the target along with reference pictures, denoted as tar_embeddings and ref_embeddings, respectively. Based on the data presented in Figure 3.11 (a), the Euclidean distances are computed for every tar_embedding about every ref_embedding. The minimum length is calculated from a set of spaces. If this minimum distance is below the threshold figure of 0.8 and the names of the target and reference faces match, the counter's value for precise forecasts grows by one. Similarly, the distance Euclid measures is computed between every target and reference picture. The accuracy is determined by dividing the number of correct forecasts by the overall number of predictions made, which in this case is 2,000. We found that our trained model had an accuracy of about 97% for evaluating masked faces. We used masked faces to assess our training model in this manner. Figure 3.11 (b) depicts the entire flowchart of examining the masked Face.

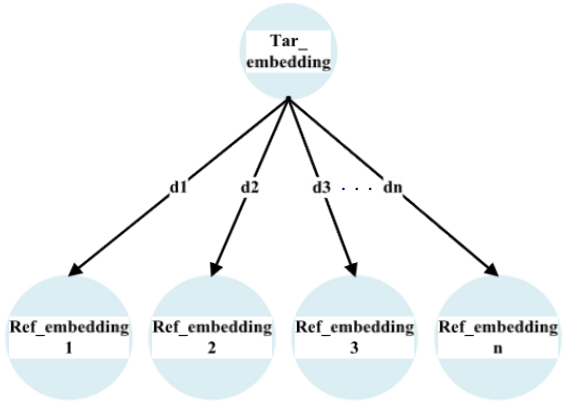

Figure 3.11 (A) Distance Among Single Tar_Embedding and Several Ref_Embeddings Can Be Calculated

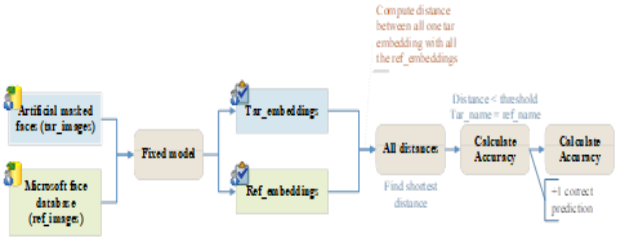

Figure 3.11(b). Flowchart for Evaluating an Unmasked Face

M. Using the Same Dataset, A Comparison of Accuracy with the Facenet Model

Each of the trained models, our fixed model and a pre-trained model from FaceNet [21], were input with the same masked faces. According to research, whereas the FaceNet model performed poorly for masked face photos, our model performed well. The comparisons of the three models' accuracy are presented in Table 3.3. FaceNet is a well-known model with 99% accuracy for standard face recognition; unfortunately, this trained model performed poorly for masked face identification. The oriented architecture, dataset, and testing information set used by both the model we showed, and the FaceNet model are shown in Table 3.3. However, our model achieved MFR precision, which was approximately 97%. Based on the evidence above, it has been inferred that our model showed outstanding results in recognizing masked faces.

Table 3.3 Using our Trained Model, Real-Time Identification of Faces While Wearing Masks

Model Title	Architecture	Training Dataset	Accuracy (Threshold -0.8)	Testing Collection of Data (dataset)
20180408-102900 [15]	Inception ResNet V1	CASIA	45.11 Percent	Artificial masked face
20180402-114759 [15]	Inception ResNet V1	VGGFace2	60.49 Percent	Artificial masked face
Fixed model	Inception ResNet V1	CASIA	96.9 Percent	Artificial masked face

According to the findings presented in Figure 3.8, our system undergoes loading the complete face data of photographs before calculating the embeddings for every different face image. The fixed algorithm is the trained model utilized for MFR. We conducted facial alignment on the facial database to eliminate extra components from the facial photographs. The system is prepped to begin the facial matching procedure after completing the input picture analysis. This research used a facial database of Microsoft image files not used during our model's training phase. The experiment employed a face database, and our photo was included. The Microsoft Face database contains 85744 photos in total.

However, we only used 2000 of them for testing in this project. All the people depicted in those photos were real.Second, to achieve better results, we employed high-resolution input photos. The input picture is obtained via a live video feed captured by a laptop's camera. The SSD (Single Shot Detector) model recognized each input frame's face and mask. Subsequently, the area around the face is extracted from the picture frame and resized to conform to the [112, 112, 3] style. The Fixed approach, which is a variant of the one we use that has undergone training, is further employed to ascertain the embedded elements from an input picture. Embeddings represent the face characteristic in a size 128-dimensional numeric format. And the Euclidean distance is determined using these embeddings.The process uses the face's embedded geometry and stored embedded geometric data from several face databases to figure out the distances using Euclidean geometry. Figure 3.11 (b) shows that the calculation exhibits a one-to-many relationship. Facial recognition works much better when the embedding is already loaded into the algorithm. This is because the input picture does not have to calculate Euclidean distance, which speeds up the processing. Moreover, if two photographs depict an identical person, the measured distance between them should ideally approximate zero. The minimum length is identified from a set of measurements, and if this minimum distance is less than the specified threshold value of 0.8, it is considered the solution. Hence, the threshold result is compared to the minimum Euclidean distance before determining the facial recognition outcome. If the gap between the input picture frame and the threshold value is smaller than the specified threshold, the person's name will appear on the shelf, indicating a successful face match. Otherwise, an unknown text will be written if the gap exceeds the limit.

In the same way, this method displays the textual representation of "Mask" or "No Mask" depending on whether the user is wearing a face mask or not. The methodology employed for current time Mask Facial Recognition (MFR) can be seen in Figure 3.12. As a result, an integrated system was effectively constructed, demonstrating a notable level of precision in accurately identifying both masked and uncovered faces. The efficacy and feasibility of the suggested approach have been established. The results of current time-veiled facial recognition are depicted in Figure 3.12.

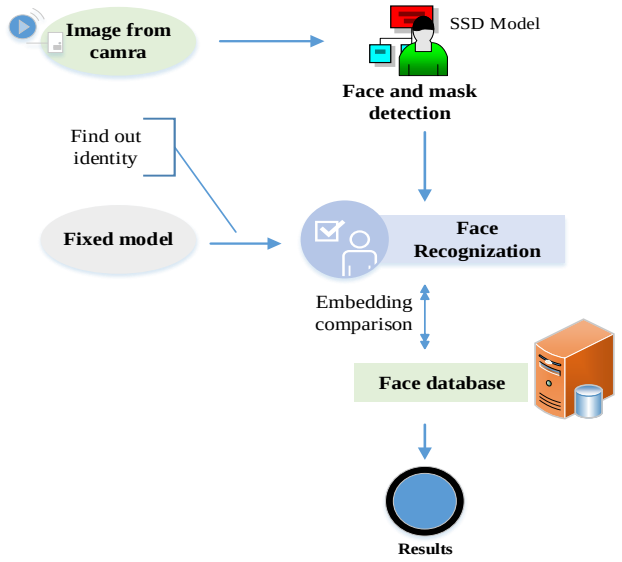

Figure 3.12 Face Recognition in Real Time Using A Mask

N. Datasets and Evaluation Performance

To train our model, it was necessary to include photographs of faces with masks and those without masks. The CASIA datasets are utilized for unmasked photos following image preprocessing, creating false data. The pictures from the CASIA dataset were subjected to masking techniques to conceal some portions of the visual content. The augmentation method employed in this study is based on the Dlib library. There are two distinct categories of prepared facial expressions. The graphical representations are depicted in Figure 8. Furthermore, the present study aims to investigate an image processing technique for generating identical facial images employing various visual aesthetics such as random cropping and random noise. The methods employed in this study include arbitrary angle, random flip, and random brightness techniques. In addition, the model was trained using both imbalanced and balanced datasets. It has been observed that the model's accuracy significantly improves when prepared with the balanced dataset. Consequently, we conducted training on three distinct models, utilizing batches of five, ten, and fifteen training photos per class, respectively.

Figure 3.13: Masked and Unmasked Training Images

Figure 3.14 illustrates the comparison between our research and many alternative methodologies. Based on the obtained data, our study focused on examining the substantial impact of the MFR (Masked Facial Recognition) method. The model demonstrates superior performance compared to the other five models. Furthermore, we have accomplished. The accuracy of the top-performing model is 1.9 percent better than the second-best model (Attention-based), and approximately 49 percent higher. Compared to the least performing model (ResNet-50) in the context of masked and

The topic of discussion pertains to the concept of unmasked facial recognition. The paradigm proposed in this study, namely Inception, ResNet V1 is a composite model that incorporates the Inception architecture [22]. The residual network has been shown to offer improved recognition capabilities. The utilization of residual connections in the training process of Inception networks has been observed to enhance performance and expedite the training procedure.

Retrieval Number: 100.1/ijrte.E79700112524
DOI: 10.35940/ijrte.E7970.12050124
Journal Website: www.ijrte.org

Published By:
Blue Eyes Intelligence Engineering and Sciences Publication (BEIESP)

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%

Figure 3.14: Comparison of Current Approach Against Other Methods

IV. RESULT AND DISCUSSION

We have developed a system that can distinguish between photos of masked and unmasked faces. The trained model outputs embeddings in a particular initial image, subsequently utilized for face comparison. The embedding process represents facial characteristics, wherein a 128-dimensional array of numerical values is used. We employed Dlib, OpenCV, and the SSD model to develop an innovative approach for generating a dataset of masked faces. Subsequently, we performed image alignment and data cleansing procedures. In the same way, we used photographs of faces that were evenly spread out from each category. The model that was trained on the unevenly spread images did much worse than the model that was introduced on the pictures that were evenly spread out. The amount of masked and unmasked facial images in balanced photos is equal. The study of LFW and face mask datasets revealed that they outperformed all current models [8, 17, 22, 23, 24]. The LFW dataset and artificially masked photos that weren't utilized to train the evaluation model were included.We created images that had been artificially obscured from people's photos in the Microsoft Face database. There are more than 8500 different faces. The data sets used for training and testing are in Tables 4.1 and 4.2, respectively.

Tables 4.1 Training Datasets

CASIA Collection of data	No. of Class	Augmented by	No. of Unmasked Pictures per class	No. of Masked Pictures per class	No. of Pictures every class	Total Pictures
Dataset 1	10585	4	10	10	5	211700
Dataset 2	10585	4	20	20	10	423400
Dataset 3	10585	4	30	30	15	635100

Tables 4.2 Testing Datasets

Collection of data (Dataset)	Category	Testing Pair	No. of Class	No. of Pictures	No. of Pictures per class
Microsoft face database	Artificial masked images	2,000	85,744	85,744	1
LWF	Real-people unmasked images	Same face and different face pair	5,749	64,973	11.3

We did tests for added testing to see how well our developed model worked with facial recognition tasks that included masked and revealed faces. These experiments encompassed many factors, such as sex, complexion, age, and various masks. Consequently, we successfully trained three separate models, which yielded a Masked Facial Recognition (MFR) precision of roughly 97%. The facial images utilized for testing and training purposes are standardized to a size of [112, 112, 3]. The evaluation of the model we trained encompasses all masked and uncovered faces. For individuals wearing masks, the region surrounding the eyes, eyebrows, and forehead is employed for taking out. Facial features. Conversely, for individuals not wearing masks, the entirety of the human face serves this purpose. Our model exhibited exceptional performance for faces that masked compared to the FaceNet model. Furthermore, we have developed a unified system capable of accurately identifying each masked and unmasked face and multiple faces concurrently.

Additionally, to enhance recognition accuracy, we conducted training on a small-scale model of around 96 MB, developing three distinct trained models. The dimension of our training system was minimized through several techniques, including selecting a smaller training picture, applying face symmetry, and utilizing reduced filter size. To modify the channel counts while training the model and eventually lessen the model measurements, 1*1 convolution were employed. The smaller model's reduced computational requirements facilitate a more expeditious inference process. In the same way, we enhanced the efficiency of our training model. We partitioned the training pictures into multiple batch sizes and then leveraged GPU acceleration to expedite the process. Three distinct models were trained, yielding an average training duration of 30 hours. In Figure 4.1, we show our real-time veiled face recognition results.

Figure 4.1 Displays the Outcomes of Our Real-Time Masked Face Recognition System

V. CONCLUSION

This research work has provided a method for correctly distinguishing between faces that are masked and those that aren't. Around 97% of the time, the MFR (Masked Facial Recognition) system offered was accurate. The construction of the veiled face datasets also involved utilizing computer vision technology. The model was trained using ASIA databases for image preparation. Additionally, its effectiveness was estimated using the LFW database. (Labelled Faces in the Wild) and artificially hidden faces.

Furthermore, an extensive study has been conducted on the performance of three separate models in the context of MFR.

In addition, our study looked at how well our developed model could recognize facial features when people were wearing or not wearing masks, and it did this for a range of factors, including gender, age, skin tone, and different types of covers. Since veiled and exposed facial recognition and detection technologies aim to keep people safe and secure, the suggested strategy can be easily and quickly added to these technologies.

FUTURE RECOMMENDATIONS

This means addressing the issue of improperly fitting rotated faces to ensure our prepared veiled photos. It is imperative to augment the quantity of well-balanced photographs for every level to enhance the standard and variety of the method. In our experiment, we limited the utilization of 60 facial images in each class. In addition, we will create a moderately complex automatic facial recognition framework that could make the whole system better at recognizing faces in general. Furthermore, our proposed research method would be beneficial for unlocking devices like laptops and phones. For security check gates, we would further improve the accuracy of tilt and flip image recognition in the future.

ACKNOWLEDGMENT

I am thankful to my supervisor, GAOMING YANG, for selecting a research topic for me and helping me greatly during my research. I want to thank my colleagues for helping me in the writing process of this paper.

DECLARATION STATEMENT

Funding	No, I did not receive.
Conflicts of Interest	No conflicts of interest to the best of our knowledge.
Ethical Approval and Consent to Participate	No, the article does not require ethical approval and consent to participate with evidence.
Availability of Data and Material	Not relevant.
Authors Contributions	All authors have equal participation in this article.

REFERENCES

1. Soyata, Tolga, et al. "Cloud-vision: Real-time face recognition using a mobile-cloudletcloud acceleration architecture." 2012 IEEE symposium on computers and communications (ISCC). IEEE, 2012. https://doi.org/10.1109/ISCC.2012.6249269
2. T. Schenkel, O. Ringhage, N. Branding," A COMPARATIVE STUDY OF FACIAL RECOGNITION TECHNIQUES With a focus on low computational power", 2019.
3. X. Liu, S. Zhang, COVID-19: Face masks and human-to-human transmission, Influenza Other Respirat. Viruses, vol. n/a, no. n/a, doi: 10.1111/irv.12740 https://doi.org/10.1111/irv.12740
4. L. Wang, A. A. Siddique," Facial recognition system using LBPH face recognizer for anti-theft and surveillance application based on drone technology", Measurement and Control (2020), Vol. 53(7-8), pp. 1070-1077, doi: 10.1177/0020294020932344 https://doi.org/10.1177/0020294020932344
5. "Paris Tests Face-Mask Recognition Software on Metro Riders," Bloomberg.com, May 07, 2020
6. 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). https://doi.org/10.1016/j.jksuci.2021.12.017
7. Das, M. Wasif Ansari, R. Basak, Covid-19 Face Mask Detection Using TensorFlow, Keras and OpenCV, 2020 IEEE 17th India Counc. Int. Conf. INDICON 2020. (2020). https://doi.org/10.1109/INDICON49873.2020.9342585
8. Joshi AS, Joshi SS, Kanahasabai G, Kapil R, Gupta S. Deep learning framework to detect
9. face masks from video footage. In2020 12th International Conference on computational intelligence and Communication Networks (CICN) 2020 Sep 25 (pp. 435-440). IEEE.
10. A. Cabani, K. Hammoudi, H. Benhabiles, M. Melkemi, MaskedFaceNet – A dataset of correctly/incorrectly masked face images in the context of COVID-19, Smart Heal. 19 (2021) 100144. https://doi.org/10.1016/j.smhl.2020.100144.
11. D. Meena and R. Sharan, "An approach to face detection and recognition," 2016 International Conference on Recent Advances and Innovations in Engineering (ICRAIE), Jaipur, 2016, pp. 1-6, doi: 10.1109/ICRAIE.2016.7939462. https://doi.org/10.1109/ICRAIE.2016.7939462
12. S. Ge, J. Li, Q. Ye, and Z. Luo, "Detecting Masked Faces in the Wild with LLE-CNNs," 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, 2017, pp. 426-434, doi: 10.1109/CVPR.2017.53. https://doi.org/10.1109/CVPR.2017.53
13. S. Ghosh, N. Das, and M. Nasipuri, "Reshaping inputs for a convolutional neural network: Some common and uncommon methods", Pattern Recognition, vol. 93, pp. 79-94, 2019. Available: 10.1016/j.patcog.2019.04.009. https://doi.org/10.1016/j.patcog.2019.04.009
14. CASIA dataset: https://github.com/SamYuen101234/Masked_Face_Recognition
15. Liu, Wei, et al. "Ssd: Single shot multi-box detector." European conference on computer vision. Springer, Cham, 2016. https://doi.org/10.1007/978-3-319-46448-0_2
16. FaceNet Pretrained Model: https://github.com/davidsandberg/facenet
17. Mandal, Bishwas, Adaeze Okeukwu, and Yihong Theis. "Masked face recognition using resnet-50." arXiv preprint arXiv:2104.08997 (2021).
18. Dlib: https://github.com/davisking/dlib
19. OpenCV: https://github.com/opencv/opencv
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. https://doi.org/10.1609/aaai.v31i1.11231
21. 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. https://doi.org/10.1109/CVPR.2015.7298682
22. 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. https://doi.org/10.1109/ICASERT.2019.8934543
23. Wang, Zhongyuan, et al. "Masked face recognition dataset and application." arXiv preprint arXiv:2003.09093 (2020).
24. P, B., & Geetha D, D. M. (2019). Design and Experimentation of Face liveliness Detection using Temperature Gradient and Image Quality Assessment. In International Journal of Recent Technology and Engineering (IJRTE) (Vol. 8, Issue 3, pp. 4360–4362). Blue Eyes Intelligence Engineering and Sciences Engineering and Sciences Publication - BEIESP. https://doi.org/10.35940/ijrte.c5516.098319
25. Begum, N., & Mustafa, A. S. (2020). CNN BLSTM Joint Technique on Dynamic Shape and Appearance of FACS. In International Journal of Engineering and Advanced Technology (Vol. 9, Issue 4, pp. 1754–1757). Blue Eyes Intelligence Engineering and Sciences Engineering and Sciences Publication - BEIESP. https://doi.org/10.35940/ijeat.d7308.049420
26. A., O., & O, B. (2020). An Iris Recognition and Detection System Implementation. In International Journal of Inventive Engineering and Sciences (Vol. 5, Issue 8, pp. 8–10). Blue Eyes Intelligence Engineering and Sciences Engineering and Sciences Publication - BEIESP. https://doi.org/10.35940/ijies.h0958.025820
27. P A, J., & N, A. (2022). Faceium–Face Tracking. In Indian Journal of Data Communication and Networking (Vol. 2, Issue 5, pp. 1–4). Lattice Science Publication (LSP). https://doi.org/10.54105/ijdcn.b3923.082522

28. Kumari, J., Patidar, K., Saxena, Mr. G., & Kushwaha, Mr. R. (2021). A Hybrid Enhanced Real-Time Face Recognition Model using Machine Learning Method with Dimension Reduction. In Indian Journal of Artificial Intelligence and Neural Networking (Vol. 1, Issue 3, pp. 12–16). Lattice Science Publication (LSP). https://doi.org/10.54105/ijainn.b1027.061321

AUTHORS PROFILE

Yamin iqra, is a master student in the School of computer science and Engineering, at Anhui University of Science and Technology. Her research direction is face detection based on machine learning (artificial intelligence). She is mainly interested in artificial intelligence that revolve around computer vision, machine learning, and artificial intelligence, with a particular emphasis on face detection and recognition. She is passionate about developing algorithms that can accurately detect and recognize faces in various conditions, including low-quality images, occlusions, and variations in pose and lighting. Her work often involves exploring techniques such as deep learning, convolutional neural networks, and transfer learning to improve the accuracy and robustness of face detection systems.

Yang Gaoming received the master's degree in computer application from Guizhou University, in 2003, and the Ph.D. degree in computer application technology from Harbin Engineering University, in 2012. He is currently a Professor and a Doc. Supervisor with the Anhui University of Science and Technology. He used to be a Visiting Scholar at the University of Arkansas for one year. His research interests include machine learning and privacy preserving. His research interests contribute to the advancement of privacy-preserving machine learning by presenting his research at conferences, workshops, and seminars. He serves as a reviewer for several prestigious journals and conferences, providing valuable feedback to fellow researchers. Additionally, he actively engages with the privacy community by participating in standardization efforts and sharing his expertise through tutorials and open-source contributions.

Marcel Merimée, BAKALA MBOUNGOU, master student, majored in the school of computer science and engineering, Anhui University of Science and Technology. His research direction is sentiment analysis based on e-commerce platform. I'm mainly interested in artificial intelligence. His research interests lie in the areas of sentiment analysis, natural language processing, and machine learning. He is particularly interested in developing algorithms that can accurately capture the nuances of customer sentiments expressed on e-commerce platforms, such as product reviews and social media comments. His work often involves applying machine learning techniques to analyze large volumes of unstructured data and extract meaningful insights.

Muhammad Asad Yamin studies the University of Rostock in Germany to pursue a doctorate. The focus of his research is computational thermodynamics. He has worked in the fields of thermodynamics and computational fluid dynamics (CFD) for five years. Heat transfer and flow analysis are his areas of interest. He currently works in the field of machine learning (ML), and in the future, he aims to apply these ML approaches to the fields of thermodynamics and CFD.

Masood Usama is a master student in the School of Mechanical Engineering, at Anhui University of Science and Technology, Huainan, China. His research direction interests revolve around AI-Powered Adaptive Rehabilitation Exoskeletons, with a particular emphasis on personalized and assistive technologies. He is passionate about designing intelligent systems that can autonomously adapt to individual users, optimize rehabilitation protocols, and improve functional outcomes. His work also involves studying human-robot interaction and developing intuitive control interfaces for exoskeleton devices, his research area is Robotics including Robot Design and Control, Robot Kinematics and Dynamics, Robot Arms Coordinated Operation. He has specialization in the intersection of robotics, artificial intelligence, and rehabilitation.