Vol.67: e24230917, 2024
https://doi.org/10.1590/1678-4324-2024230917
ISSN 1678-4324 Online Edition

BABT
Brazilian Archives of Biology and Technology

Article - Engineering, Technology and Techniques

KInsight: a Robust Framework for Masked Face Recognition

Shivani Sharma[1]
https://orcid.org/0000-0001-6652-2651

Rashmi Chaudhry[2]
https://orcid.org/0000-0002-6842-9917

Dhruv Tewari[2]
https://orcid.org/0000-0001-9700-8401

Saurabh Soreng[2]
https://orcid.org/0000-0003-2555-8943

Sachin Kumar[3*]
https://orcid.org/0000-0003-3949-0302

[1]Thapar Institute of Engineering and Technology, Department of computer Science and Engineering, Patiala, India, [2]Netaji Subhash Technical University, Department of Computer science and Engineering, New Delhi, India, [3]South Ural State University, Big Data and Machine Learning Lab, Chelyabinsk, Russia.

Editor-in-Chief: Alexandre Rasi Aoki
Associate Editor: Fabio Alessandro Guerra

Received: 01-Sep-2023; Accepted: 22-Nov-2023

*Correspondence: sachinagnihotri16@gmail.com; Tel.: +79512471669 (S.K.).

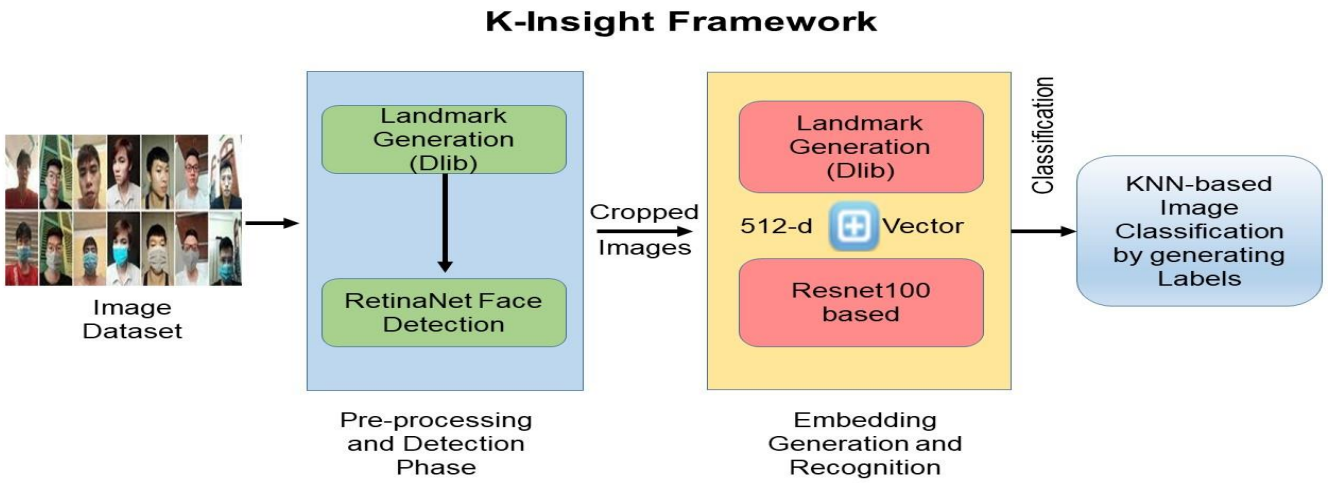
HIGHLIGHTS

- An integrated scheme to improvise masked face detection & recognition is proposed
- The scheme Integrate Dlib & Retina face to detect oversized & low-resolution images
- Results shows an accuracy of 98.5 % over benchmark dataset.

Abstract: The pandemic and other environmental conditions have increased the use of masks as a precautionary measure. It is an effective way to protect ourselves from viruses/pollution/ and other health-affecting environmental factors. However, for smart devices such as smartphones with face locks, attendance systems, and smart surveillance cameras with enabled face recognition, these masks raised another challenge of masked face recognition. Masked face recognition is an increased subset challenge of the standard face recognition problem as they lack facial features. The masked images are occluded, making the structure and the facial features of the non-occluded region of importance. This paper presents a novel two-fold approach KInsight (K Nearest Neighbor-based Insight Face algorithm) for masked face detection using antelopev2, which uses a RetinaFace detection algorithm and a ResNet100 Convolutional Neural Network for face detection and embedding generation. Further, we propose to use a KNN classifier for masked face recognition. Several experiments have compared the proposed scheme's performance with important research contributions. Experimental results show that the scheme significantly outperforms several benchmark approaches with an accuracy of around 98.5%.

Keywords: Face Recognition; Deep Neural Network; Face detection.

GRAPHICAL ABSTRACT



INTRODUCTION

The recent Pandemic, environmental conditions, and other related factors had a huge impact on the lives of people worldwide. Many people lost their lives in the pandemic, and some are struggling with unhealthy environmental conditions, etc. [1]. People are afraid to go to places with wide public gatherings. Many parts of the world are still struggling with some other viruses and variants causing fear in people while going out. Health organizations suggest, and even people themselves are willing to follow guidelines such as social distancing, vaccination, contactless public settings, and wearing masks, etc. for their own safety. Now we are getting back to our lives and with certain cautions, we are going to public places with gatherings like offices, educational institutes, airports, etc. like before. Smart devices supporting contactless operations are more in use by the authorities, People usually prefer to wear masks while accessing these public places. This helps in protecting and preventing them from their very own reasons but raises several challenges for biometric authentication methods with enabled face recognition. Face recognition-based systems are crucial to support contact-less operations and their accuracy is of top-most importance. However, wearing a mask hinders this operation as a large number of facial features get occluded by the mask. This decreases the performance of these systems therefore, it is important to investigate the methods which can help in improving the masked face recognition system's performance. Masked face recognition is a sub-technique of occluded face recognition methods with prior knowledge about the targeted face's occluded area [2]. The domain of occluded face recognition captivated the attention of several researchers in the last few years. Face recognition algorithms learn specific features from an individual's face such as nose, mouth, cheeks, chin eyes, eyebrows, and forehead for recognition purposes. However, with a mask, most of these features get obstructed and only the eyes, eyebrows, and forehead are visible for feature extraction and learning. Therefore, we need to improvise our system by effectively focusing on these unobstructed regions of a masked face and learning important features for recognition with accuracy.

In this paper, we divided the process into two different phases i.e. masked face detection and masked face recognition. The initial detection phase is improvised in order to detect faces in low-resolution and high-occlusion scenarios too. Further, the issue of oversized images is also handled during the face detection phase which was causing zero embedding generation before. Further, masked face recognition, on the other hand, aims to recognize a face with a mask based on the eyes and the forehead regions using an ensemble deep learning approach. The performance of the proposed scheme has been evaluated in terms of accuracy, precision, recall, and F1-Score over several benchmark datasets and with different existing benchmark facial recognition schemes. With a set of experimental results obtained the proposed scheme gains an accuracy of around 98% which is significantly better than another state-of-the-art.

The paper is organized as follows. Section 2 represents a detailed review of existing methods for face recognition. Section 3 presents a brief discussion regarding work, and dataset, and a detailed

discussion of the proposed pre-processing technique with respect to one benchmark dataset. Further, a detailed proposed methodology is presented. Experimental setups and results are discussed in section 4. Finally, the work is concluded in Section 5.

Motivation and Contribution of Paper

Face recognition is a cutting-edge technique in the deep learning and computer vision domain. However, mask-wearing can reduce the overall accuracy of the existing recognition system as most of the part of the face becomes occluded. Hence, we require a novel approach that can improve the accuracy of masked face recognition by utilizing and enhancing the face features that are clearly available after wearing a mask. In this paper, a deep ensemble method has been proposed for masked face recognition using deep-learning models with high accuracy which is best suited to face recognition-enabled services. From the literature studied it can be stated that the proposed scheme is a novel two-fold ensemble scheme for improved facial detection and recognition from images. The step-by-step novel contribution of the approach is as follows:

- The very first novelty exists in the proposed pre-processing of the dataset in order to make it more suitable for better detection of facial images, especially in occluded and oversized scenarios which were one of the major challenges in the detection phase as well as most important supporting step for the entire process.
- The scheme is novel as it handles hazel, low resolution, and oversized images for better detection and avoids null embedding generation.
- Secondly, for generating rich embeddings of masked faces, it is proposed to use the combination of both Dlib and RetinaNet. This helps in generating the embedding of images that were nil when detected by any of the existing models individually.
- Further, embedding is proposed to be extracted from a non-occluded image of the subject + non-occluded part of the subject's image when wearing a mask.
- Transfer learning using KNN model uses deep features extracted from non-masked images to recognize masked faces.
- Finally, the obtained embeddings are used for training the simple K-NN Classifier model for the purpose of making the whole process lightweight, considering distance metrics such as cosine and Euclidean.

RELATED WORK

Face recognition is one of the critical as well as captivating problems for researchers in the computer vision domain especially when we are using biometric authentication systems with enabled face recognition in our daily lives. Masks have a huge impact on the efficiency of the Face recognition system and the accuracy of such systems is of prime importance. Existing face recognition techniques acquire an accuracy of around 99.3% however, this took a steep downfall when dealing with masked face recognition. Existing research work for face recognition can be broadly categorized into three directions [3] as Occlusion robust feature Extraction [4]-[14], Occlusion-aware face recognition [15]-[21] and occlusion recovery-based face recognition.

In this paper we are exploring the latest category of methods and review some important literature in context to the same. The set of methods falls in the occlusion recovery-based face recognition category. These methods try to recover occluded areas and then use traditional face recognition models to perform verification. Hence, causes dependency on the accuracy of recovered areas. These approaches have been implemented using deep learning, linear reconstruction, and sparse classifier representation. PCA reconstruction has been used in [22] to handle eye area occlusion caused due to glass wearing. [23] PCA variants have been used to reconstruct the occluded areas. Authors in a[24], a linear combination of samples to represent occluded areas. Sparse representation has been improved using historical knowledge of pixel distribution [25]. The benefits of robust sparse representation have been exploited for error images using adjective block occlusion taking tailored scores [26]. Deep learning is another captivating solution for solving the occlusion face recognition challenges. In [27] authorized LSTM and autoencoder are used collaboratively for modeling spatial and temporal characteristics of occluded faces. GAN is the powerful tool for blind reconstruction [28],[29].

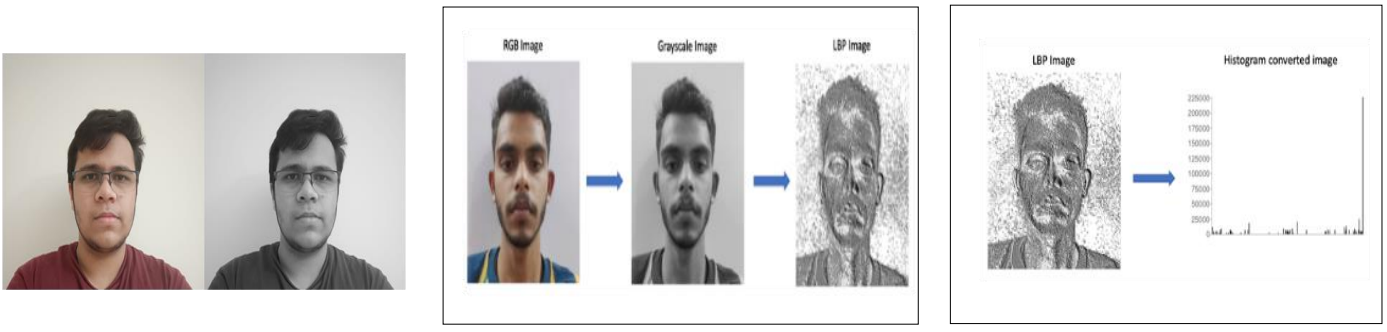
The author in [30] used GAN to reconstruct the corrupted facial areas. In [3] pre-trained GAN model is trained over non-occluded images and then is used to recover the facial regions that are occluded. Further, an analytical study of masked face recognition has been presented in [31], representing the efficiency

of several existing models for face recognition over masked face recognition and showed that the maximum of their performance falls down when working with a masked faceas compared to non-mask images. This provides analytical proof of the improvisation needed in terms of masked face recognition. Some very recent work showed other dimensions in the domain such as [32] Efficient and Robust Training of Face Recognition CNNs by Partial FC i.e. by sparsely updating variant of the FC layer, named Partial FC (PFC). This helps in reducing the memory and computing costs required. Further, [33] resolves the issue of trainingthrough the datasets which are limited to a small set of available mages with ground-truth labels. They explored webly supervised learning to learn from the large scale of web images and corresponding tags without any manual annotations along with limited fully annotated datasets. In particular, inspired by the recent success of webly supervised learning in deep neural networks andIn [34] blending techniques i.e. a mask-to-face image blending approach basedon UV texture mapping is introduced, A self-learning-based cleaning pipeline for processing noisy training datasets is been introduced while considering the impacts of the long-tail distribution and hard faces samples, a loss function named Balanced Curricular Loss is also introduced. The scheme achieved an accuracy of 84.528%. Masked Face Recognition Using MobileNet V2 with Transfer Learning is presented in [48] where detection and recognition are improved. The proposed work uses deep models with feature extraction for resolving masked face recognition. Initially the approach detects masks and further employ VGG16, VGG 19, ResNet50, and ResNet 101 for recognition and claims to achieve the accuracy within the range of 78 to 99 %. The basic contribution claimed in the work is to monitor and decrease the pace of coronavirus and to detect persons wearing face masks. Another scheme called the joint Holistic and Masked Face Recognition scheme is presented in [49] using CNN and plain Vision Transformers (ViTs) and uses the proxy task of patch reconstruction. This helps in improvising the detection and in turn helps in improving the classification process. The parameters of the model are initialized using proxy tasks which in turn exhibit improvised the training stability and performance of face recognition. Furthermore, two different methods for integrating holistic withmasked face recognition in one framework, namely FaceT are given. FaceT is claimed to be performing better than CNNs on both holistic and masked face recognition benchmarks.

From the literature, it can be stated that a number of contributions have been made using different methods to improve mask face recognition considering different challenges. However, still, the accuracy lies between the ranges of 85 to 90 % for most recent works which is not significant when we deal with real-time authentication systems where accuracy must be as close to 100%.

CHALLENGES AND PROPOSED METHODOLOGY

Most of the recent research contributions focused on extracting the features from the images using techniques like Linear Binary Pattern, Dlib, Insight face, faceNet, etc. following the pipeline as shown in Figure 1, which takes grayscale images as input and produces a pattern of the image presented in Figure 2 in order to produce the histogram in Figure 3. These histograms generated for different images then can be compared for correct classification. LBP uses Euclidean distance between the extracted histograms as a measure for comparison. Dlib also follows the pipeline by taking grayscale images as input, which gets converted into 128-dimensional embedded feature vectors. It has been observed that the pipeline works fine with completely detected faces however when wearing a mask most of the feature-rich areas like the mouth, nose, etc. become occluded hence degrading the performance [31].



a) b) c)

Figure 1: (a) Conversion of Raw Image into Grayscale. **(b)** Image Conversion from RGB to Grayscale which in turn Changes to LBP image. **(c)** LBP image converted to histogram equivalent

Other approaches such as Deep-learning models such as In-sight face, ArcFace, etc. provide the huge capability of learning and extracting features even when we have less scope for doing so. However, these models are data intensive. Hence, for resolving the challenge, we can use image augmentation by ensuring the alignment, and use of non-occluded feature regions to enhance the singularity weight of the region while using the images. Therefore, the overall challenge is to come up with all-round improvements in each sub-step of masked face recognition i.e. data preparation, lightweight model selection, efficient face detection, important feature extraction, rich embedding generation and low data-intensive image classification. In this paper, we proposed an all-around improvement in the whole process i.e. from the data preprocessing step to classification results.

In this paper, we used multiple datasets for experimental purposes such as the CoMask20 dataset [35], RMFRD [36] SMFRD [37], MDMFR [38], and Facemask [39] [40] as shown in Table 1. However, to make understanding easier for the reader and remove the redundancy, we explained the proposed model and its working with respect to one standard dataset i.e. CoMask20 dataset, and discussed the experimental results with respect to all other benchmark datasets.

CoMask20 dataset was generated by Hoai Nam Vu et.al, at their institution [2] called CoMask20. They captured the video of 5 to 10 seconds of each subject and then separated each frame by 0.5 seconds. Each subject's frame is stored in separate folders. To improve the quality of the data set all obscured or occluded images have been eliminated. Folders carry an unequal number of images as well as have different qualities such as light intensity, different angles, head scales, backgrounds, etc. The split frames also carry different sizes due to both vertical and horizontal capture of videos. The dataset contains 2754 total facial images with 300 different identities. A view of the dataset is presented in Figure 2.

Table 1. Datasets Considered

Dataset	Number of Masked Images	Number of Un-Masked Images
CoMask20 dataset	2754	2754
RMFRD dataset	5000	90000
SMFRD dataset	500000	500000
MDMFR dataset	3174	2832
FaceMask dataset	690	686
FaceMask Detection Dataset	10000	10000

Figure 2. CoMask20 Dataset [35]

Phase-I: Data Pre-Processing

The data collected has to be augmented and pre-processed to be invariant and accepted, to be provided as input to various different models based on the back-end architecture used in those models. As discussed in the previous section, we have used the CoMask20 dataset for testing the performance of the benchmark and proposed schemes. We propose to prioritize the non-occluded face regions such that the model extracts the maximum of the facial features from them. The original dataset contains 312 folders without any segregation of the train data and the test data. All the images of the individuals were placed in sub-folders for each distinct person.

In this form of the dataset, the images were not correctly placed into train and test folders leading to images with the wrong label 'id' being used for the wrong application. To resolve this issue, a new augmented version of the CoMask20 dataset was created in which the images of respective subjects were separated into Masked and Unmasked folders which were classified as Test and Train folders respectively. The original dataset also had misaligned images of individuals which were used for training thus reducing the recognition ability of the model. Upon segmentation, all images were aligned in the new dataset, and the first 5 images were used in training. The reason for choosing 5 images was an experimental analysis that was conducted, which provided an understanding that a maximum of the first 5 images contained a meaningful representation of an individual's data points that led to fruitful training results.

Finally, the processed dataset contains 312 folders that have masked as well as unmasked images of 312 different subjects. Image clusters of each subject are processed thereby segregating the unmasked and masked images and storing them in different folders while retaining their class label or subject name. This segregation leads to the creation of the train and test sets respectively. The model is trained on the unmasked images, and it is tested on the masked images. To make the dataset more diverse and robust 20 images of new subjects have been added in both trains as well as test sets. At least, 5 images per class label have been used to train the model.

Phase-II: Proposed Methodology

This paper presents a novel approach to recognizing masked faces using the InsightFace Python Library. The processes of masked face recognition can be divided into three phases i.e. Face Detection, Feature Extraction, and Masked Face Recognition. The overall view of the proposed methodology is presented in Figure 3 and explained in further section.

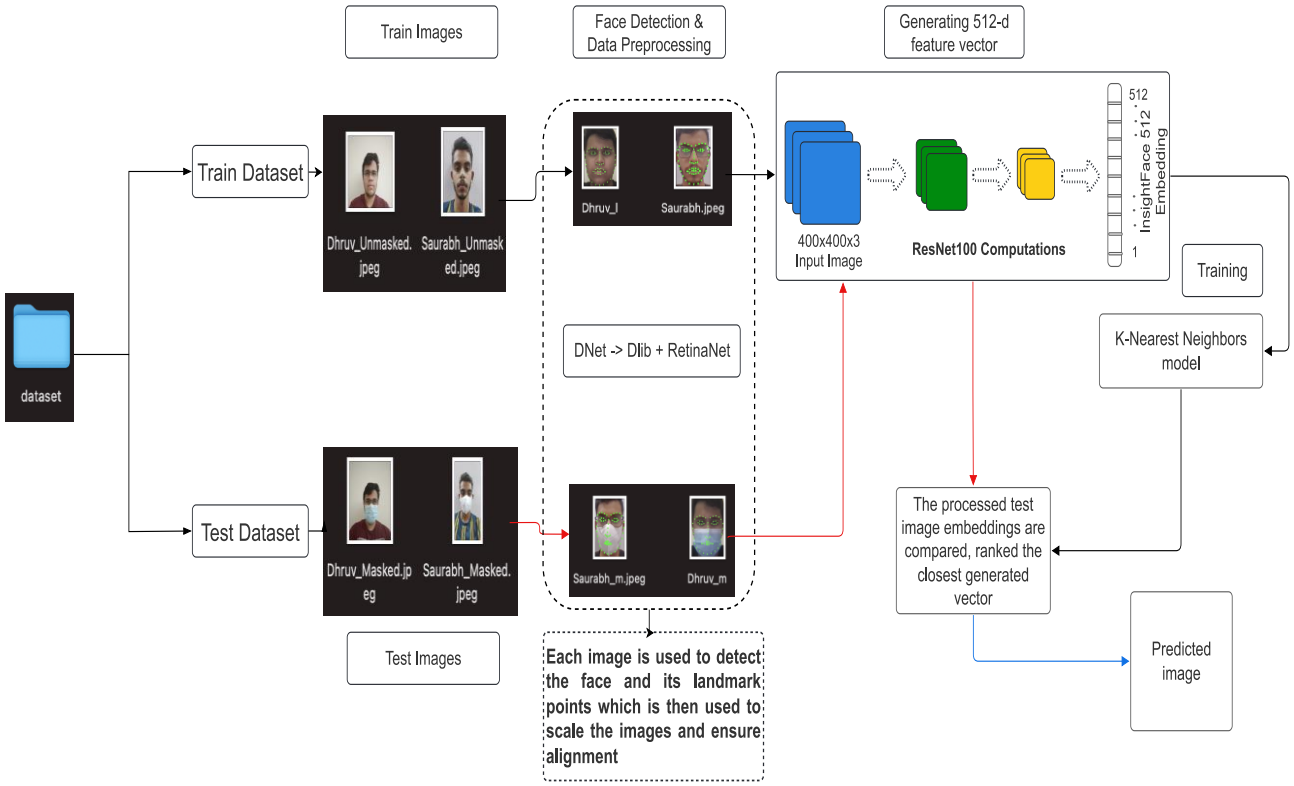


Figure 3: Proposed K-Insight framework

Face Detection

The initial challenge is to detect the individual's face in an image. Various available methods such as dlib [41], MobileNet [42] and RetinaNet [43], etc. produce good results for face detection. RetinaNet is one of the best one-stage object detection models that has proven to work well with dense and small-scale objects. It utilizes the Feature Pyramidal Networks (FPN)[44] for generating rich semantic features out of an image without any loss in terms of power, speed, or memory. RetinaNet provides promising face detection with various scales. RetinaNet clearly focuses on detection, alignment, pixel-level parsing, and 3D regression [2]. However, certain images, like over-scaled face images, and unclear or dense face images, could lead to nil embedding generation and ultimately reduce the data quality. For Example, in some cases, the image is a fully scaled image of a face as shown in Figure 4. Both images are of the same individual that contains the entire face covered in the image. If we pass the image on the right that is the entire face and not just the face-bounded region RetinaNet cannot rightly predict the face and thus there will be no face detection. Therefore, the method should be able to detect the faces and must be able to generate rich embeddings for the next step such that it becomes suitable for developing deep learning models. To resolve this challenge, we propose to collaboratively use dlib+RetinaNet. For the images that are not solely detected using RetinaNet due to the entire image acting as a face as shown in Figure 4, Dlib helps in doing initial face detection and generate 64 landmark vector called Face Set which is then passed to RetinaNet for further processing. When we explicitly use the left image using only the bounded region with the landmark points entitled in bounding annotations in the proposed KInsight algorithm, it is capable of generating the embeddings. This helps in generating a richer data set and allows for the detection of those face images too which may get missed by RetinaNet alone. So, in this paper, we exploited the advantages of both these algorithms for face detection.

Face embedding

We need to extract facial features in order to discriminate between faces. The feature vector extracted must be rich enough to differentiate faces not only when the full face is visible but also in scenarios where the subject is wearing a mask, i.e. with occluded faces as well. Mask-wearing reduces the features available for the system to differentiate as only a non-occluded part of the face is visible. Vector of size 128-d or 256-d generated by models such as Dlib, FaceNet, etc. is not enough for building a robust system. We need a rich feature vector that can learn dense features out of a face such that even with an occluded face, it has enough information to differentiate between while maintaining real processing time. In this work, we explored Arcface which employs ResNet100 Convolutional Neural Network to generate 512-dimensional feature vectors to fulfill the requirement to identify large-scale identities. ArcFace makes predictions based on the features and weight angle. The class weight in ArcFace is not only the nearest weight to embedding features i.e. θ but also the nearest weight when the angle is added. This causes better performance as the conditions become stricter and boundaries become hard between neighbors as shown in Figure 5.

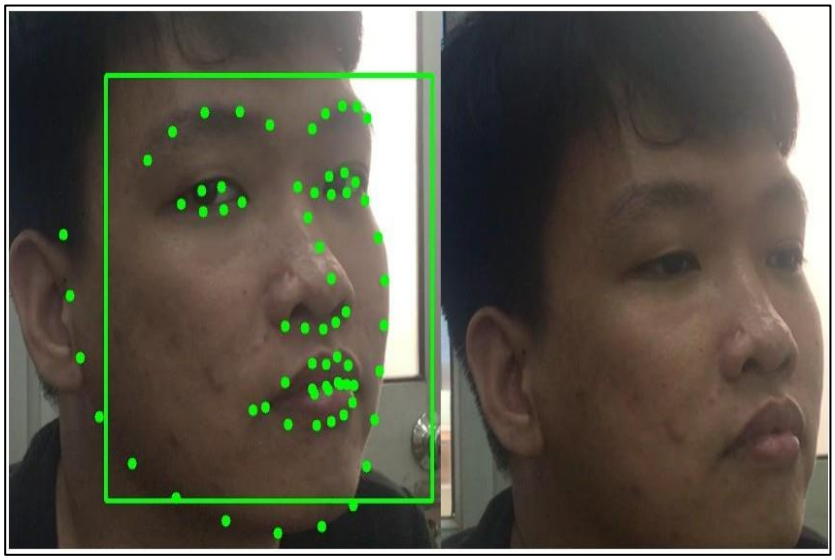


Figure 4. Over-scaled images of the face

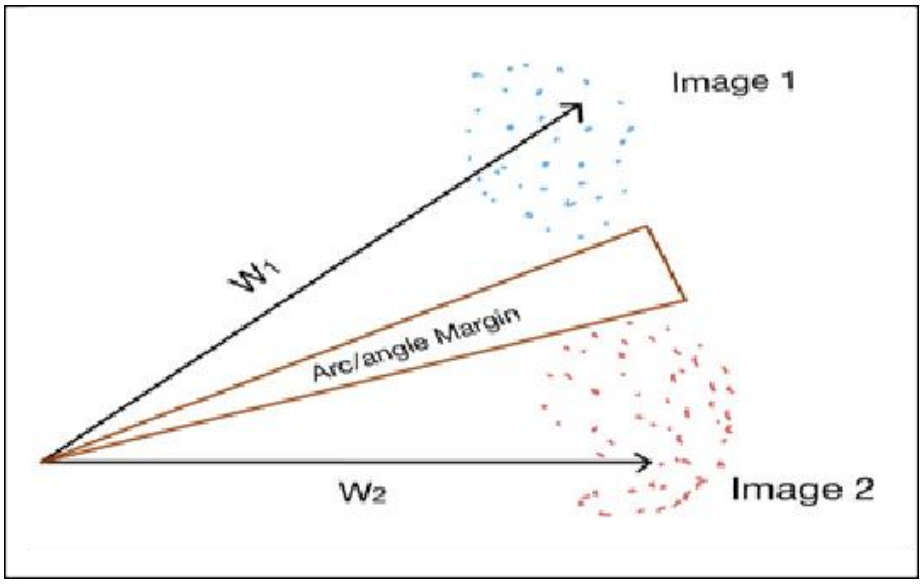


Figure 5. Geometric Interception of ArcFace

Our proposed scheme uses Insight Face whose face detector backend is based on Retina-Net, the detected face is passed to the InsightFace algorithm with an antelope2 backend that generates 512-d facial embedding as shown in Figure 6, which would be used to further train the deep learning model for the further.

```
[[Face(bbox=array([ 745.5539 ,  277.13324, 1165.7688 ,  854.5919 ], dtype=float32), kps=array([[ 931.1964 ,  513.8652
3],
        [1112.2714 ,  497.74374],
        [1089.5594 ,  603.8185 ],
        [ 992.6133 ,  731.87823],
        [1128.8704 ,  715.79425]], dtype=float32), det_score=0.78777796, embedding=array([ 2.16315031e+00, -8.992915
75e-01, -1.10100999e-01,  9.70849395e-02,
        -8.52053344e-01,  3.46632957e+00,  1.64410219e-01, -1.12430429e+00,
         2.60110021e+00, -3.88756096e-01, -6.20923579e-01, -1.14099935e-01,
         1.46500003e+00,  8.22728574e-02, -3.97115082e-01,  9.09278154e-01,
         3.70794646e-02, -5.55469841e-03,  1.67960835e+00, -8.58457685e-01,
        -1.25978160e+00,  5.11125982e-01, -1.26891959e+00,  6.57947958e-02,
         8.74833226e-01,  1.13748908e+00,  2.62492836e-01,  6.70501292e-01,
         1.61935902e+00, -2.09315300e-01, -1.26159620e+00, -4.63950425e-01,
        -7.80007362e-01,  1.74706960e+00, -1.86122552e-01, -2.09789366e-01,
        -2.45171237e+00, -1.76420259e+00, -3.45456362e-01, -1.26739550e+00,
        -1.33950222e+00, -1.60314739e+00,  2.93679059e-01,  8.41005445e-01,
        -9.02849078e-01,  1.46650505e+00,  1.65767014e-01,  2.19304115e-02,
         1.33734846e+00,  1.17782664e+00, -1.40179169e+00,  8.61993432e-02,
         1.93630481e+00,  8.68141770e-01,  9.97751594e-01, -2.46190459e-01,
         3.98745239e-01, -6.42061412e-01, -1.54093647e+00, -2.27590576e-01,
        -8.60453486e-01,  1.41187525e+00, -2.28241116e-01, -1.83509588e-02,
         7.91480422e-01, -6.63641021e-02,  1.65024173e+00, -1.36819351e+00,
         8.13324004e-02,  1.08925596e-01,  1.51091766e+00, -6.84291184e-01,
         8.27549398e-03,  5.98804593e-01,  8.45231175e-01,  6.53182149e-01,
        -1.27480900e+00, -4.66067851e-01,  4.25472409e-01,  8.67347360e-01,
         2.94086123e+00,  2.15562433e-01, -2.81489313e-01,  2.70932150e+00,
        -2.42942691e-01,  1.50820744e+00, -1.15759015e+00, -1.12770236e+00,
         4.75344896e-01, -5.99340200e-01,  1.13649940e+00, -2.52234899e-02,
        -1.08669305e+00,  2.31339073e+00,  3.69666606e-01,  6.81873798e-01,
```

Figure 6. Over-scaled image embeddings generated by Dlib+RetineNet

Masked face recognition

Several classifiers are present such as K-NN, SVM, and CNN architecture for the purpose of classification. We propose to use KNN for classification, as a simple and robust classifier for the masked face recognition system. The reason to choose this over the other classifier is that our task requires the classification to be based on the closest value of a random test data point with a probably trained image thus giving the result of what is the closest image as per the image data point, This would be recognition task possible in the K-NN architecture with the utmost efficiency. KNN uses a 512-d vector for prediction. The simple yet effective KNN ensures no loss of data during the learning procedure. Furthermore, KNN facilitates getting updated with new knowledge without the need to re-training from scratch. Different values of K are being selected for prediction and classification purposes which is not an easy task for other complicated classifiers such as deep neural networks. Easy parameter tuning and simplicity make KNN the first choice in spite of its high computational cost with growing data size.

KNN-based classification

The 512-d embedding created out of facial features can be provided as input to the KNN which in turn generates K's closest neighbors. When a vector is given as input to the classifier, the similarity with the known data is measured one by one in terms of their distance from the queried vector. These measured distance values are sorted in ascending order and then the first K-sorted entries are considered the K-closest vectors as a result. The distance is calculated between the vectors using a function that is pre-defined such as Jaccard, Euclidean, cosine, etc. In our proposed method, we used the cosine similarity measure as it is used to compare two vectors in multi-dimensional space and produce a ranking depending on their similarity measure with the queried vector. Let X and y be two vectors then cosine similarity can be measured as shown in Equation 1.

$$Sim(x,y) = Cos(\theta) = \frac{x.y}{||x||.||y||} \quad (1)$$

Where, ||x|| and $||y||$ is the Euclidean norm of vector x= $\{x_1, x_2, x_3 ... x_p\}$ and y. It helps to evaluate the angle between the queried and the data vector. If the value of the angle inclined towards '0' represents lower similarity means they are orthogonal to each other whereas, if tending towards '1', shows a high similarity between vectors. Hence, we need to do the classification for a maximum value of cos (θ) means the minimum value of θ as they are inversely proportional to each other.

The training of the model has been done on non-occluded, augmented faces which provides us with the maximum information with the help of the facial embeddings. Further, we tested our model over test data with varying values of K *i.e.* K= {1, 3, 5, 7}, and achieved the best results for k=1. A detailed discussion of the result and proposed model performance is presented in the following section.

EXPERIMENTS AND DISCUSSION

In this section, the experiments performed and a detailed discussion on the results will be provided as follows.

Experimental settings and performance metrics

The experiments were built and run on a 2.3 GHz Quad-Core Intel Core i5 processor with 8GB of RAM on a Unix-based macOS system. The entire codebase was made in python. The models were trained on an online Kaggle notebook which provided us with a GPU of 12 GB. Performance metrics provide a way to evaluate the reliability and accuracy of a prediction model. The common metrics Accuracy, Precision, Recall and F-measure are used in the study for performance evaluation.

Prediction Accuracy

It measures the correctness i.e. total number of images that are correctly classified and recognized by the classifier and is presented in Equation 2.

$$Accuracy = \frac{TP + TN}{Total\ Number\ of\ Transactions} \quad (2)$$

Where, TP and TN present the sample images which have been classified as positive and negative correctly respectively.

Precision

This metric measures the correctness i.e. ratio of the total number of images that are correctly predicted and all classified images, given in Equation 3.

$$P = \frac{TP}{TP + FP} \quad (3)$$

Where, FP denotes the images that have been classified as positive but are actually negative.

Recall

It measures the completeness i.e. total number of correctly classified images by the classifier with respect to the total positive samples and is presented in Equation 4.

$$R = \frac{TP}{TP + FN} \quad (4)$$

F-measure

It is the combination of precision and recall stated above and is the harmonic means of both as shown in Equation 5.

$$F - Measure = 2 * \frac{P * R}{P + R} \quad (5)$$

After pre-processing the data, the augmented RGB images of size 400x400x3 were given as input to KInsight (proposed model). It then generated 512-d feature vectors that acted as the facial embedding on each image in the training set. These embedding were trained on a K-Nearest Neighbor classifier by varying values of K and the obtained results of the performance metric have been studied

Performance evaluation on face detection

Initially, we compare the first phase of the approach i.e. face detection with state of art, considering different benchmark datasets. Table 2 represents, a comparative analysis of the proposed scheme over a variety of datasets. The first column of the table represents datasets. Further, columns 2 and 3 provide details of masked and unmasked images, the dataset consists of. Further, columns provide a detailed view of performance in terms of accuracy, precision, recall, and F1-Score. It can be clearly observed from the table that in all cases the proposed scheme approaches 100% detection accuracy. This is because our approach successfully resolves the challenge of nil embedding generation.

Further, we compared the performance of our detection scheme with existing benchmark schemes as shown in Figure 7. It can be clearly observed that the proposed detection approach performs significantly better. This is due to the ability of the proposed method to detect oversized and low-resolution schemes, which resolves the challenge of nil embedding generation and produces rich embedding for the next layer of the model.

Table 2. Performance of proposed scheme over benchmark datasets

Dataset	Images with Mask	Images without Mask	Accuracy	Precision	Recall	F1-Score
RMFRD [36]	5000	90000	99.64	99.34	99.2	99.63
SMFRD [37]	500000	500000	97.6	98	98	99.1
MFMRD [38]	3174	2832	100	100	100	100
Facemask [39]	690	686	100	100	100	100
Facemask Detection [40]	10000	10000	100	100	100	100
CoMask20 [35]	2754	2754	100	100	100	100

Performance evaluation on face recognition

Initial, experiments were conducted over the CoMask20 dataset and then expanded to other datasets. The following section presents a detailed discussion with respect to the CoMask20 dataset and then we provide a cumulative result over several benchmark datasets.

We compared the proposed scheme with benchmark models such as dlib- face, LBPH, DeepFace, etc., as shown in Table 3 over the CoMask20 dataset. It can be observed, that the best accuracy of 98.54% has been achieved on the proposed scheme. The proposed architecture is compared with other existing models such as the images were trained on the dlib face recognition module and LBPH algorithm which had provided accuracy of 13.2% and 24.63% respectively. Further, the image dataset was also trained on the ArcFace model of InsightFace using the deep face library which compared the similarity of 2 images based on their cosine distance. After training the images on the deep face model and testing the model, an accuracy of 66.66% was achieved.

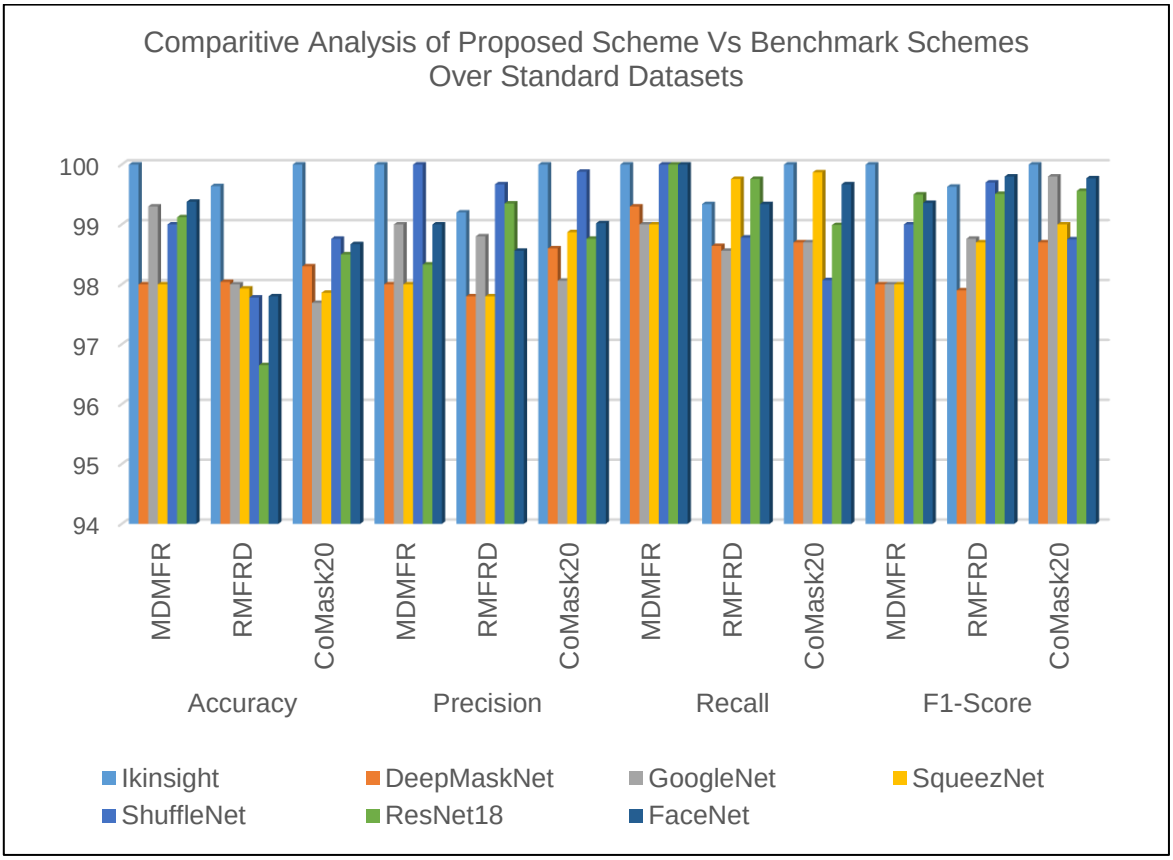


Figure 7. Performance of Proposed and other scheme in terms of Face Detection

Table 3. Comparative Performance of Existing Models vs. Proposed Scheme Over CoMask20 dataset

Performance metric	Dlibface recognition, 128-d	LBPH Algorithm	Deepface (ArcFace)	InsightFace (antelopev2) 512-d (Proposed Method)
Accuracy	0.132	0.2643	0.666	0.985
Precision	0.121	0.2332	0.616	0.982
Recall	0.132	0.264	0.664	0.985
F1-Score	0.126	0.247	0.639	0.982

Table 4. Similarity Measure Comparison for K=1

Similarity Measure	Accuracy	Precision	Recall	F1-Score
Euclidean	0.985	0.981	0.985	0.981
Cosine	0.9752	0.9855	0.9714	0.9784

A major reason behind the proposed KInsight architecture yielding the best results is the pre-processing of data and the rich embedding produced due to the combination of both Dlib and RetinaNet during the detection phase. Further, ResNet outperforms all the other Convolution Neural Networks in image processing as it is successful in vanishing as well as exploding gradient problems.

Furthermore, we have experimented with two similarity measures i.e. cosine and Euclidean, and compared the performance in terms of both as shown in Table 4. It can be observed that for both of these similarity measures the performance of the proposed scheme is good. However, the Cosine measure is a little more significant than the Euclidean distance. Therefore, we chose to move with cosine similarity measure in further experiments.

Another set of experiments was conducted to check the performance of the proposed approach with varying values of K i.e. k= 1, 3, 5, 7. It can be observed from Table 5 that the performance of the scheme is best at k=1 and degrades as the value increases. With k=1, the accuracy of 98.5%, precision of 98.1%, recall of 98.5%, and an F1-Score of 0.981 is achieved. This is because k = 1 finds the first closest image that

corresponds to its feature vector thus giving the desired result. The value of k increases the number of images compared to give the final prediction, which causes disparity from the initial closest image that should resemble a face recognition task.

Table 5. Comparison of Proposed KInsight Model on different values of K

Value of K	Accuracy	Precision	Recall	F1-Score
1	0.985	0.981	0.985	0.981
3	0.970	0.965	0.970	0.963
5	0.656	0.645	0.656	0.614
7	0.465	0.414	0.465	0.402

Table 6. Comparison of the proposed model with recent approaches

Method	Technique	Dataset	No. of Images	Accuracy
K-Insight	InsightFace(antelopV2), 512-D (Proposed Method	CoMask20 RMFRD MDMFR	2754 5000 3174	**98.5** **96.45** **94.3**
[2]	InsightFace(ArcFace)+LBP Based Voting	CoMask20	2754	87
[47]	FaceMaskNet	User Data RMFRD	2000 5000	88.92 88.82
[48]	DeepMaskNet	MDMFR	4006	93.33
[49]	CNN+BoF	RFMRD SMFRD	5000 500000	91.3 88.9

Another set of experiments has been conducted to compare the performance of the proposed model with some of the latest work with similar goal of masked face recognition. A comparative analysis of recent research work has been presented in Table 6. The table's first column represents the reference to research work, the second column presents the technique used in the corresponding work. Column 3 and 4 shows the dataset considered for evaluating the performance of the proposed scheme and the number of images in the respective dataset. The last column shows the accuracy achieved for masked face detection during the testing phase. It can be clearly seen that with the CoMask20 dataset, which has around 2754 images, the proposed scheme works significantly well.

However, the same dataset has been used by another work [2], which achieved an accuracy of 87 %. Further, the proposed method also shows promising results over other benchmark datasets i.e. with RMFRD [36], and MFMRD [38] with an accuracy of 96.45% and 94.3% respectively which is a significant improvement compared to other existing as well.

Additionally, the performance is measured with respect to different face recognition models available over three significant datasets named MDMFR, RMFRD, and CoMask20 as presented in Table [7] in terms of accuracy. It can be clearly observed that the proposed KinsightNet provides the highest accuracy with respect to all three datasets i.e. of 94.3%, 96.45%, and 98.5%.

A summary of the proposed work is as follows: Our proposed method is divided into the following stages

- Image augmentation by ensuring the alignment, and use of non-occluded feature regions to enhance the singularity weight of the entire image for the entire dataset.
- Face detection using the Retina Net model.
- Generation of 512-d face vector embedding using ResNet100 architecture.
- Used the 512-d feature vectors and trained a K-NN classifier for various values of K to judge the best possible result.
- K value is taken as 1 which bore the best result.
- The proposed method yielded a result that provided around 98% accuracy.

Table 7. Comparison of the accuracy of proposed model with recent approaches with respect to benchmark datasets

Method	MDMFR (%)	MDMFR (%)	MDMFR (%)
DarkNet53	89	87	91.9
ResNet18	84.2	83	87
VGG19	90.91	87	9.01
ShuffleNet	88.57	86	90
AlexNet	88.79	86	90
DenseNet	88.15	87	90.46
DeepMaakNet	93.33	90	95.86
CNN+LBP	84.67	86	87
Kinsight	**94.3**	**96.45**	**98.5**

CONCLUSION AND FUTURE SCOPE

In this paper, a two-fold methodology is proposed which detects the face and recognizes the facial features of an individual. In the first step, augmentation has been done to make the proposed method more robust to have the dataset more organized and the computations faster while maintaining image integrity. Once the data is augmented, the need to detect the face from an image to use it as the input stream is produced. This is achieved using the dlib library and RetinaNet convolution neural network. The second step of the method- ology involves the generation of 512-d feature vectors by passing the input stream to the InsightFace architecture with the backend of ResNet100 and the AntelopeV2 model. The generated vector is then used for the facial recognition of an individual. Experiments were conducted to verify the proposed method, and performance has been compared to other existing benchmark methodologies. Evaluation metrics illustrate that the proposed technique is an integral improvement over other cutting-edge face recognition techniques. Reconstruction techniques like GAN can be explored to improve the recognition and classification of masked facial images. Further, other cutting-edge techniques for improving resolution which in turn may help in improving embedding generation would also help in generating better recognition results.

Funding: This research received no external funding.
Conflicts of Interest: Authors declare no conflict of interests

REFERENCES

1. Chen S, Liu Y, Gao X, Han Z. Mobilefacenets: efficient cnns for accurate real-time face verification on mobile devices. In: Chinese Conference on Biometric Recognition; 2018 Aug 11-12; Urumchi, China: Springer; c2018. p. 428-38.
2. Vu HN, Nguyen MH, Pham C. Masked face recognition with convolutional neural networks and local binary patterns. Appl Intell (Dordr). 2022 Aug; 52(5):5497-512.
3. Chen Y-A, Chen W-C, Wei C-P, Wang Y-CF. Occlusion-aware face in painting via generative adversarial networks. In: IEEE International Conference on Image Processing (ICIP). 2017 Sep 17-20; Beijing, China: IEEE xplore; c2017. p. 1202-6.
4. Liao S, Zhu X, Lei Z, Zhang L, Li SZ. Learning multi-scale block local binary patterns for face recognition. In: International Conference on Biometrics. 2007 August 27-29; Seoul, Korea: Springer; c2007. p. 828-37.
5. Zhao G, Pietikäinen M. Dynamic texture recognition using local binary patterns with an application to facial expressions. IEEE Trans Pattern Anal Mach Intell. 2007 Apr; 29(6):915-28.
6. Dalal N, Triggs B. Histograms of oriented gradients for human detection. In: IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05); 2005 Jun 20; San Diego, California: IEEE xplore; c2005; p. 886-93.
7. Zhang B, Shan S, Chen X, Gao W. Histogram of Gabor phase patterns (hgpp): A novel object representation approach for face recognition. IEEE Trans Image Process. 2006 Dec; 16(1):57-68.
8. Zou J, Ji Q, Nagy G. A comparative study of local matching approach for face recognition. IEEE Trans Image Process. 2007 Sep; 16(10):2617-28.
9. Wiskott Wiskott L, Fellous J-M, Kuiger N, von der Malsburg C. Face recognition by elastic bunch graph matching. IEEE Trans Pattern Anal Mach Intell. 1997 Jul; 19(7):775-9.
10. Heheb I, Al-Maadeed N, Al-Madeed S, Bouridane A, Jiang R. Random sampling for patch-based face recognition. In: 5th International Workshop on Biometrics and Forensics (IWBF); 2017 Apr 4-5; Coventry, UK: IEEE xplore; c2017. p. 1-5.

11. Seo J, Park H. A robust face recognition through statistical learning of local features. In: International Conference on Neural Information Processing; 2011 Nov 13-17; Shanghai, China: Springer; c2011. p. 335-341.
12. He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. In: Proceedings of the IEEE conference on computer vision and pattern recognition; 2016 Jun 26- Jul 1; Las Vegas, Nevada: IEEE xplore; c2016. p. 770-8.
13. Krizhevsky A, Sutskever I, Hinton GE. Imagenet classification with deep convolutional neural networks. Advances in neural information processing systems (NIPS); 2012 Dec 3-6; Lake Tahoe, Nevada, USA: ACM; c2012;p. 25-29.
14. Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition; 3rd International Conference on Learning Representations (ICLR 2015); 2015 May 7-9; San Diego, California: Computational and biological learning society; c2015. P. 1409.1556.
15. Min R, Hadid A, Dugelay J-L. Improving the recognition of faces occluded by facial accessories. In: IEEE International Conference on Automatic Face & Gesture Recognition (FG); 2011 Marc 21; Santa Barbara, CA, USA: IEEE xplore; c2011. p. 442-7.
16. Chen Z, Xu T, Han Z. Occluded face recognition based on the improved svm and block weighted lbp. In: International Conference on Image Analysis and Signal Processing; 2011 Oct 21; Wuhan, China: IEEE xplore; c2011. p. 118-22.
17. Morelli Andrés AM, Padovani S, Tepper M, Jacobo-Berlles J. Face recognition on partially occluded images using compressed sensing. Pattern Recog Lett. 2014 Jan; 36(1):235-42.
18. Xie J, Xu L, Chen E. Image denoising and inpainting with deep neural networks. Adv Neural Inf Process Syst.(NIPS); 2012 Dec 3-6; Lake Tahoe, Nevada, USA: ACM; c2012; p. 25-29.
19. Ou W, Luan X, Gou J, Zhou Q, Xiao W, Xiong X, et al. Robust discriminative nonnegative dictionary learning for occluded face recognition. Pattern Recog Lett. 2018 May; 107(1):41-9.
20. Wan W, Chen J. Occlusion robust face recognition based on mask learning. In: IEEE International Conference on Image Processing (ICIP); 2017 Sep 17; Beijing, China: IEEE Publications; c2017. p. 3795-9.
21. Song L, Gong D, Li Z, Liu C, Liu W. Occlusion robust face recognition based on mask learning with pairwise differential Siamese network. In: Proceedings of the IEEE/CVF international conference on computer vision; 2019 27 Oct- 2 Nov; Seoul Korea: IEEE xplore; c2019. p. 773-82.
22. Park JS, Oh YH, Ahn SC, Lee SW. Glasses removal from a facial image using recursive error compensation. IEEE Trans Pattern Anal Mach Intell. 2005 Mar; 27(5):805-11.
23. De La Torre F, Black MJ. A framework for robust subspace learning. Int J Comput Vis. 2003 Aug; 54(1):117-42.
24. Wright J, Yang AY, Ganesh A, Sastry SS, Ma Y. Robust face recognition via sparse representation. IEEE Trans Patt Anal Mach Intelli. 2008 Apr; 31(2):210-27.
25. Zhou Z, Wagner A, Mobahi H, Wright J, Ma Y. Face recognition with contiguous occlusion using Markov random fields. In: 12th International Conference on Computer Vision; 2009 Sep 29; Kyoto, Japan: IEEE xplore; c2009. p. 1050-7.
26. Zhao F, Feng J, Zhao J, Yang W, Yan S. Robust lstm-autoencoders for face de-occlusion in the wild. IEEE Trans Image Process. 2017 Nov; 27(2):778-90.
27. Van Oord A, Kalchbrenner N, Kavukcuoglu K. Pixel recurrent neural networks. In: PMLR International Conference on Machine Learning; 2016 Jun 11; New York City, NY, USA: ACM; c2016. p. 1747-56.
28. Kingma DP, Welling M. Auto-encoding variational bayes. arXiv preprint, 20 Dec 2013. Available from: arXiv:1312.6114.
29. Ulyanov D, Vedaldi A, Victor L. Deep image prior. Int J Comput Vis. 2020 Jul; 128(7):1867-88.
30. Damer N, Grebe JH, Chen C, Boutros F, Kirchbuchner F, Kuijper A. The effect of wearing a mask on face recognition performance: an exploratory study. In 2020 International Conference of the Biometrics Special Interest Group (BIOSIG). 2020 Sep 16; Darmstadt, Germany: IEEE xplore; c2020. p. 1-6.
31. An X, Deng J, Guo J, Feng Z, Zhu X, Yang J, et al. Killing two birds with one stone: efficient and robust training of face recognition cnns by partial fc. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition; 2022 Jun 21-24; New Orleans, Louisiana: IEEE xplore; c2022. p. 4032-41.
32. Zheng W, Yan L, Wang F-Y, Gou C. Learning from the web: Webly supervised meta-learning for masked face recognition; 2021 Oct 18; Virtual: IEEE xplore; c2021. p. 4299-308.
33. Feng T, Xu L, Yuan H, Zhao Y, Tang M, Wang M. Towards mask- robust face recognition. In: Proceedings of the IEEE/CVF international conference on computer vision; 2021 Oct 18; Virtual: IEEE xplore; c2021. p. 1492-1496.
34. CoMask-20 dataset [cited 2022-1-12]. Available from: https://github.com/tuminguyen/COMASK20.
35. RMFRD dataset. Available from: https://www.kaggle.com/datasets/muhammeddalkran/masked- face recognition. [Accessed on 2022-1-12].
36. SMFRD dataset; 2022-01-12. Accessed: 22-01-MDMFR dataset. Available from: https://github.com/X-zhangyang/Real-World-Masked-Face-Dataset. Available at: https://drive.google.com/drive/folders/1YgHEQzb4yL2FigYPfAccessed
37. Facemask-dataset [Internet]. [place unknown: publisher unknown]; [Cited 2022 Jan 12]. Available from: https://www.kaggle.com/sumansid/facemask-dataset.
38. Facemask detection dataset [Internet]. [place unknown: publisher unknown]. [2022 Jan 12]. Available from: https://www.kaggle.com/luka77/facemask-detection-dataset-20000-images.

39. Boyko N, Basystiuk O, Shakhovska N. Performance evaluation and comparison of software for face recognition, based on dlib and OpenCV library; Second International Conference on Data Stream Mining & Processing (DSMP); 2018 Aug 21-25; Lviv, Ukraine: IEEE xplore; c2018. p. 478-82.
40. Zhou Y, Liu Y, Han G, Fu Y. Face recognition based on the improved mobilenet. In 2019 IEEE Symposium Series on Computational Intelligence (SSCI) IEEE Symposium Series on Computational Intelligence (SSCI). 2019 May 19-23; San Francisco, CA, USA: IEEE xplore; c2019. p. 2776-81.
41. Sitepu SE, Jati G, Alhamidi MR, Caesarendra W, Jatmiko W. Facenet with retina face to identify masked face. In 2021 6th International Workshop on Big Data and Information Security (IWBIS); 2021; Depok, Indonesia;2021. 86 p.
42. Lin T-Y, Dolĺar P, Girshick R, He K, Hariharan B, Belongie S. Feature pyramid networks for object detection. In Proceedings of the IEEE conference on computer vision and pattern recognition; 2017 July 21-26; Honolulu, Hawaii: IEEE xplore; c2017. p. 936-44.
43. Anwar A, Ray CA. Masked face recognition for secure authentication. 2020 Jun 13-19; Seattle, WA, USA: IEEE xplore; c2020. arXiv preprint Available from: arXiv:2008.11104.
44. Martindez-Diaz Y, Luevano LS, Mendez-Vazquez H, Nicolas-Diaz M, Chang L, Gonzalez-Mendoza M. Shufflefacenet: A lightweight face architecture for efficient and highly-accurate face recognition. In Proceedings of the IEEE/CVF International Conference on Computer Vision Workshops. 2019 27 Oct-2 Nov; Seoul, Korea: IEEE xplore; c2019. p. 1-8.
45. Golwalkar R, Mehendale N. Masked-face recognition using deep metric learning and facemask-net-21. Appl Intell; 2023 Aug; 52(1):13268–79.
46. Ullah N, Javed A, Ali Ghazanfar M, Alsufyani A, Bourouis S. A novel deep masknet model for face mask detection and masked facial recognition. J King Saud Univ Comput Inf Sci. 2022 Jan; 34(10):9905-14.
47. Hariri W. Efficient masked face recognition method during the COVID-19 pandemic. Signal Image Video Process. 2022 Apr; 16(3):605-12.