

Facial recognition for partially occluded faces

Omer Abdulhaleem Naser¹, Sharifah Mumtazah Syed Ahmad¹, Khairulmizam Samsudin¹,
Marsyita Hanafi¹, Siti Mariam Binti Shafie¹, Nor Zarina Zamri²

¹Department of Computer and Communication System Engineering, Faculty of Engineering, University Putra Malaysia (UPM),
Serdang, Malaysia

²Cybersecurity Malaysia, Cyberjaya, Malaysia

Article Info

Article history:

Received Nov 29, 2022

Revised Feb 2, 2023

Accepted Feb 4, 2023

Keywords:

Database

FaceNet

Facial recognition

Landmark

Multi-task cascaded

convolutional neural networks

Occluded faces

Support vector classification

ABSTRACT

Facial recognition is a highly developed method of determining a person's identity just by looking at an image of their face, and it has been used in a wide range of contexts. However, facial recognition models of previous researchers typically have trouble identifying faces behind masks, glasses, or other obstructions. Therefore, this paper aims to efficiently recognise faces obscured with masks and glasses. This research therefore proposes a method to solve the issue of partially obscured faces in facial recognition. The collected datasets for this study include CelebA, MFR2, WiderFace, LFW, and MegaFace Challenge datasets; all of these contain photos of occluded faces. This paper analyses masked facial images using multi-task cascaded convolutional neural networks (MTCNN). FaceNet adds more embeddings and verifications to face recognition. Support vector classification (SVC) labels the datasets to produce a reliable prediction probability. This study achieved around 99.50% accuracy for the training set and 95% for the testing set. This model recognizes partially obscured digital camera faces using the same datasets. We compare our results to comparable dataset studies to show how our method is more effective and accurate.

This is an open access article under the [CC BY-SA](https://creativecommons.org/licenses/by-sa/4.0/) license.



Corresponding Author:

Omer Abdulhaleem Naser

Department of Computer and Communication Systems Engineering, Faculty of Engineering

Universiti Putra Malaysia (UPM)

43300 Serdang, Selangor, Malaysia

Email: omar.abdulhalem592@gmail.com

1. INTRODUCTION

Facial recognition (FR) is the technique of identifying a person based on trained photographs of that individual. FR technology can recognize both objects and faces. FR is a sort of biometrics that is very intelligent, rapid, and does not require fingerprints or complex passwords [1]. However, FR requires the extraction of facial features as well as a huge number of datasets that require substantial storage space and computational time. These are the most significant requirements that may negatively influence the facial recognition process [2]. Moreover, many recent researchers working on facial recognition systems have encountered the following obstacles: illumination, face posture, and partial occlusion. Unfortunately, these challenges have led to a very low level of FR accuracy [2], [3]. In fact, facial occlusion is recognized as one of the most challenging problems addressed by previous researchers. Moreover, as the COVID-19 outbreak spreads, an increasing number of people are wearing masks to prevent infection, making automatic FR difficult [4]. Therefore, this work proposes and applies simple and effective strategies to increase the recognition accuracy of occluded faces. Among these methods is the application of a multi-task cascaded convolutional neural network (MTCNN), which can detect facial positions in photos (face landmarks). MTCNN also contributes to the refining of discovered false negative landmarks [5]. FaceNet is an additional application of machine learning that functions as a verification tool by turning each processed image into 128-dimensional features and then placing these 128-dimensional features in Euclidean

space [6]. To classify data into meaningful categories, we can use a linear support vector classifier (SVC) that fits the given dataset and returns a "best fit" hyperplane. Once we have the hyperplane, we can use it as input into a classifier to determine the "predicted" class [7].

This paper provides an overview of recent facial recognition approaches that are currently available in section 2 [8]-[17]. In addition, an overview of popular facial recognition techniques and their general classifications are provided. Then these methods are compared based on their respective techniques, databases, and recognition rate. The remaining sections are organized as follows. Section 3 provides an overview of the suggested model technical technique and database collecting. The experiments and results are discussed in section 4. Section 5 covers the evaluation of results. Section 6 presents the conclusion and suggested future directions required for an effective face recognition system.

2. COMPREHENSIVE THEORETICAL BASIS

Various researchers from around the world have conducted extensive research on facial recognition for occluded faces over the last few decades. They have used various techniques to produce an acceptable level of system accuracy that can be used and applied in many different applications, including health, security, and educational sectors. In Ouannes *et al.* [8], the authors proposed a combination of Laplacian pyramid blending and CycleGANs methods to achieve accurate facial recognition for partially occluded faces. Moreover, Ouannes *et al.* [8] used two feature extraction techniques, and they are learned features and hand-crafted features, experimented on two types of datasets: EKFD and IST-EURECOM LFFD. The EKFD dataset achieved an accuracy of 94%, while the IST-EURECOM LFFD achieved an accuracy of 72%.

Zang and Junyong [9] proposed that occluded faces can be accurately recognized by using principal component analysis (PCA) to extract the features of histograms of oriented gradients (HOGs) and local binary patterns (LBPs) and generating a HOG-LBP joint feature. Therefore, this kind of deployment was made to enhance the dictionary of facial occlusion. This method has achieved an accuracy of no more than 88.5%. Xu *et al.* [10] proposed yet another facial occlusion approach, which essentially employs a fusion method comprised of connected-granule labelling features and reinforced centrosymmetric local binary patterns. This method used two types of datasets (UHDB-3 and CelebA) for the purpose of integrating manifold features. Eventually, this approach has reached an accuracy of 95%, which is able to learn facial templates that can help recognise partial facial occlusion. Thus, a local identity-related region was retrieved using an attention method. The authors then reorganized local features into a single template using global representation. A pairwise differential siamese network [11] can also distinguish obstructed faces by comparing the similarity of facial blocks and false feature pieces. The dataset used in this approach is MegaFace Challenge, and the maximum accuracy achieved is approximately 74.40%.

As a result, [12] proposed another technique for developing facial recognition for faces that are mostly obscured by masks, sunglasses, and scarves by deploying various models. FaceNet, MaskNet, CosFace, and ArcFace are examples of these models. The public WiderFace dataset was used to train these models. These models were tested and evaluated using the receiver operating characteristic curve (ROC). ResNet50 is the refined model that is used as a CNN base-line model for these algorithms, and it is responsible for refining the extracted facial features. As a result, the accuracy gained from this methodological deployment is about 78.25%. On the other hand, GANs with partial occlusion-aware stages have been proposed [13]. Then it would be used as an additional input by the latter GANs in order to synthesize the ultimate de-occluded face.

Zhao *et al.* [14] created a new LSTM-autoencoder model from two LSTM components to restore partially obscured faces. The first component sought to encode faces resistant to occlusion, while the second sought to remove recurrent occlusion from faces. Furthermore, Xu *et al.* [15] developed deep GANs in order to eliminate partial face-occlusion while iteratively filling in the missing parts. Although de-occlusion techniques produced better results than occluded-face recognition, there was still room for improvement. In fact, lower recognition rates were caused by only identifying non-occluded parts of faces rather than the entire face. Furthermore, occlusion's location and size affected the recognition performance of the remaining non-occluded part. Li *et al.* [16] proposed that identity-preserving generative adversarial networks could be used to inpaint and recognize occluded faces at the same time. They used an inpainting network, a global-local discriminative network, a parsing network, and an identity network among other networks. The identity diffusion between the real face and its inpainted corresponding was effectively eliminated as a result of the proposed incorporated networks. The model achieved accuracy of 85.50% on the LFW dataset.

As a result, all of the methods proposed in these related works effectively achieved a high accuracy level by overcoming the challenge of faces with varying illumination, pose, and colors. They were also somewhat effective in approaching an acceptable level of accuracy in order to overcome the challenge of occluded facial recognition, but there is still room for improvement in terms of techniques and accuracy rates. Therefore, MTCNN technique is applied to overcome the limitations of techniques and come up with higher accuracy rates. Table 1 summarizes related research work in occluded facerecognition.

Table 1. Summary table of the related facial recognition approaches

Related work	Techniques	Datasets	Recognition rate %
Ouannes <i>et al.</i> [8]	SURF + k-NN & KD-Tree	IST-EURECOM LFFD LFW	72%
Zhang and Junyong [9]	HOG-LPB	ORL/CelebA	88.5%
Xu <i>et al.</i> [10]	OREO	UHDB-31 CelebA	95%
Song <i>et al.</i> [11]	Pairwise differential Siamese network	MegaFace Challenge/LFW	74.40%
Huang <i>et al.</i> [12]	FaceNet, MaskNet, CosFace, and ArcFace	WiderFace/CelebA	78.25%
Dong <i>et al.</i> [13]	OA-GAN	CelebA	–
Zhao <i>et al.</i> [14]	RLA	LFW	–
Xu <i>et al.</i> [15]	DCGANs	AR	–
Li <i>et al.</i> [16]	IP-GANs	LFW	85.50%
Ferdinand <i>et al.</i> [17]	MTCNN/FaceNet/SVM	CelebA/NUAA	79%

3. METHOD

The research framework consists of three main techniques which are: MTNCC performs face detection landmarks, FaceNet performs facial feature extraction and SVC performs facial recognition to find the best-matched identities. FaceNet is a deep neural network trained to infer a mapping from face images to a compact Euclidean distance space to measure the similarity of the face. Following the generation of this space, conventional methods can be readily applied to FaceNet embeddings as feature vectors for applications including face recognition, verification, and clustering [18]. Finally, after training the datasets in the system, SVC will find the optimum hyperplane to fit it. A classifier can then take the hyperplane as input to produce a "predicted" class [7]. This is the proposed model's flowchart, as shown in Figure 1.

- MTCNN is used in the pre-processing stage to perform facial detection. Additionally, the loaded images will be resized, and the colors of the images will be converted to RGB colors in a clearer manner. Furthermore, the image pixels will be normalized to values ranging from 1 to 0.
 - Feature extraction stage, in which FaceNet is used to extract additional facial features and then normalize those extracted features.
 - Classification stage: This is where the model is defined and training takes place by using SVC to perform model classification, which can aid in improving probability prediction.
 - Prediction stage, in which the trained model identifies faces from the testing set or from a webcam.
- This flowchart, shown in Figure 1, clearly shows how the overall model's mechanism works in four stages.

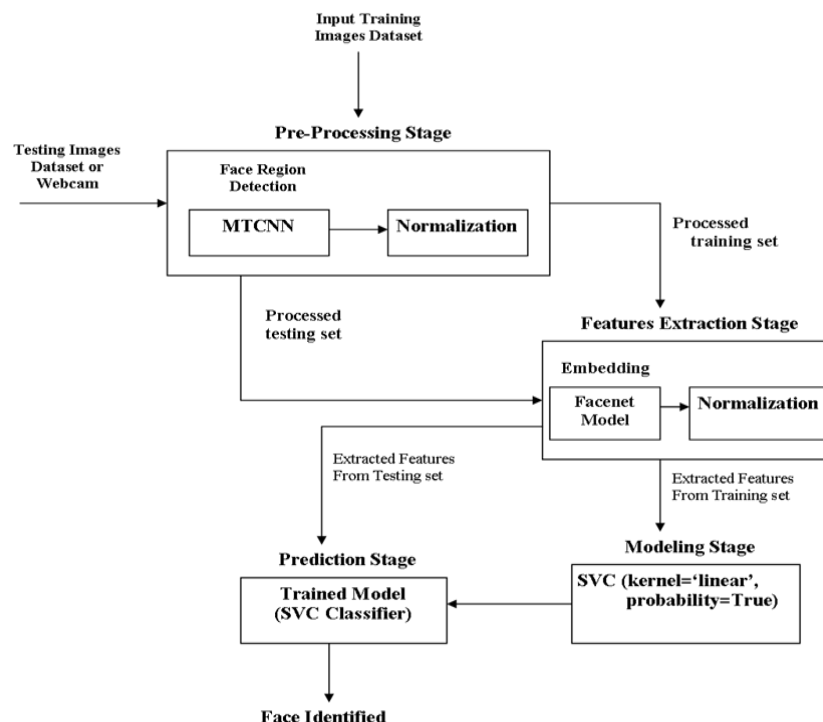


Figure 1. Research framework

3.1. MTCNN

Operates in three steps. The first step is to detect face landmarks, which essentially plays an efficient role in processing stored images and predicting face position. The second and third steps of the MTCNN process involve further refining the predicted images to eliminate false predictions and adding more landmarks to the images to improve the accuracy of the resulting images [19].

As a result, Figure 2 shows that a CNN is used to process the images as they are scaled down in the proposal network (P-Net) stage of the process. Image features are extracted in the second and third stages, referred to as the refine network (R-Net) and the output network (O-Net), respectively, to build bounding boxes. In addition, for each bounding box, O-Net computes the five-face landmark areas [20]. Furthermore, MTCNN's theme framework is similar to that of cascaded CNNs, but it deals with both face area detection and facial feature detection simultaneously. MTCNN, like many other CNN models aimed at addressing image issues, employs image pyramids, bounding box regression, non-maximum suppression (NMS), and a variety of CNN technology as clearly illustrated in Figure 3 [21].

The MTCNN library was used to perform three tasks in the preprocessing phase: face detection, face scaling, and face cropping. After using detection to pinpoint the face's location in a picture, the resulting bounding box was rendered. The cropping was then implemented after the bounding box was calculated. After receiving a facial scan, the image would be resized to match the dimensions specified by the model, which is 160×160 Pixels [22].

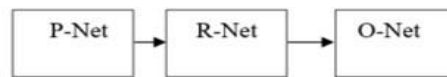


Figure 2. Block diagram of MTCNN [20]

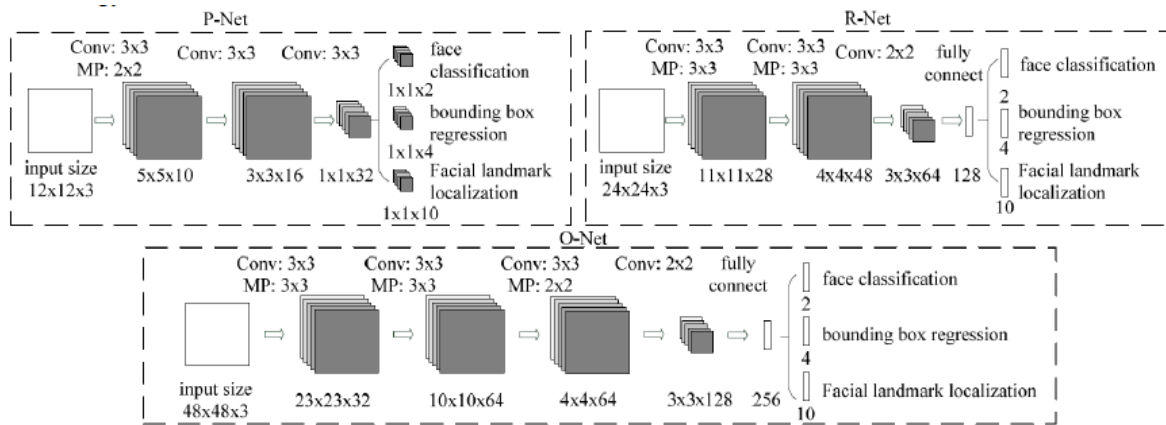


Figure 3. The three networks' architectures, where "MP" stands for maximum pooling and "Conv" stands for convolution. Convolution and pooling have step sizes of 1 and 2, respectively [21]

3.2. MTCNN loss function

The loss function in MTCNN is composed of three subnets. Face recognition makes direct use of the cross-entropy cost function, while box regression and key point positioning make use of the L2 loss. The overall loss is calculated by adding the three components together and then using their relative importance weights. Training P-Net and R-Net focuses primarily on target frame accuracy and less on critical point judgement loss. As a result, the impairment of judgement at crucial positions is less significant. For O-Net, the loss of the critical points is catastrophic because their position is more crucial [23].

a) FR loss function: When solving the FR issue, the cross-entropy cost function is applied to the (xi) input sample.

$$Lidet = - (yidet \log(pi) + (1 - yidet) (1 - \log(pi))) \tag{1}$$

the genuine sample label is denoted by *yidet*, while the likelihood that the network's output is a face is shown as *pi*.

b) Box Regression: We use Euclidean distance for regression of the target box as in (2):

$$L_{box} = \| the\ y^{boxi} - the\ y_{boxi} \| \quad (2)$$

where the y_{boxi} is the object's ground-truth bounding box, while the y^{boxi} denotes its new coordinates as determined by the network's output.

c) Key point loss function: In this function, we also use Euclidean distance as shown in (3):

$$L_{landmark} = \| the\ y_i^{landmark} - the\ y_{landmark} \| \quad (3)$$

where the $y_i^{landmark}$ represents the key points' coordinates received after the network output, while the $y_{landmark}$ indicates the actual key points' coordinates.

3.3. FaceNet model

FaceNet model is also used because it is effective at capturing and extracting facial features by combing through all of the images in the dataset. FaceNet is a deep convolutional neural network (CNN) component that was radically developed in 2015 by a group of international researchers because FaceNet can amazingly contribute to improving false detection and recognition by capturing all the similar and different facial features for the purpose of yielding a more precise output [19]. Figure 4 depicts a detailed block diagram of the FaceNet architecture.



Figure 4. Block diagram of FaceNet architecture [19]

FaceNet is made up of batch layer, deep architecture (CNN), L2 normalization, and triplet loss, as shown in Figure 4. FaceNet consist of eleven layers, which are known as convolutional layers. These eleven layers contain approximately 140 million parameters that are responsible for learning how to map from the dataset's stored images. However, triplet loss is trained and applied for the purposes of this paper. In triplet loss, there is a concept known as "anchor," which is responsible for identifying people's identities based on the images in the dataset. The identity is the same if the distance between the anchor and the positive sample is minimized. Otherwise, the maximized distance indicates that the identity is different. Figure 5 depicts the triplet loss training process [19].

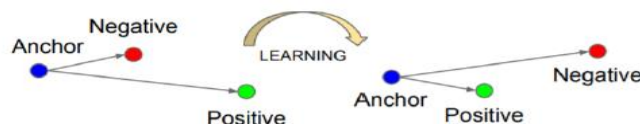


Figure 5. The training process of triplet loss [19]

FaceNet differs from other methods primarily in that it does not rely on a bottleneck layer for recognition or verification jobs but rather learns the mapping from the photos and builds embeddings. After the embeddings are constructed, they can be used as the feature vector in subsequent tasks such as verification, and recognition, using the established methods from that area. In the case of face recognition, for instance, K-NN can be used with embeddings as the feature vector, and any clustering algorithm can be used to group faces into groups; all that's needed for verification is a threshold value to be set [24].

3.4. Model classification

SVC is a learning algorithm that is thought to be very effective in terms of processing datasets and classifying them in a useful manner. Furthermore, SVC is used to categorise the features. The family of related supervised learning techniques known as support vector machines (SVMs) is widely employed for these purposes. SVM is another name for a classification and prediction tool that makes use of machine learning theory to improve upon the accuracy of its predictions. Fuzzy clustering-based support vector machines are systems that use the hypothesis space of linear functions in a high-dimensional feature space and are taught

using an optimization-theory-inspired learning method. The goal of cluster analysis is to identify hidden patterns in a batch of data that is otherwise completely unlabeled. SVM that takes pixel maps as input is more accurate than even the most advanced neural networks [25].

3.5. Dataset collection

CelebA, MFR2, WiderFace, LFW, and MegaFace Challenge are the datasets collected for this paper, and the total number of images gathered is approximately 6,075, with approximately 125 identities as illustrated in Table 2.

Table 2. The used databases

Database	No. of subjects	No. of Images per subject	Total images per database	Description
CelebA	111	30	2404	Includes over 200K celebrity images of 10K subjects with 40 attribute annotations per-image [26].
MFR2	14	12	130	Has 53 celebrities and politicians in 269 photographs. Each identity has images of masked and unmasked faces [27]. The size of each image is (160x160x3)
WiderFace	150	13	1950	WIDER FACE leverages public WIDER dataset photographs to create a baseline for face detection. 32k pictures and 394k faces are labelled [28].
LFW	136	15	2040	13,233 LFW face images. 1,680 of 5,749 identities have multiple photographs. Standard LFW evaluation provides verification accuracies on 6,000 face pairs [29].
MegaFace Challenge	113	12	1356	MegaFace is a massive database of 1 million images of 690 thousand people who were free to pose and adjust their lighting, expressions, and exposure [30].

4. EXPERIMENTS AND RESULTS

In the pre-processing stage, MTCNN is employed to extract faces from the images in the four datasets, which are of varying sizes and colors. The images are first resized and normalized to a consistent size and color format, enabling MTCNN to identify the faces. MTCNN then utilizes a multi-step process to detect and align the faces, while eliminating any non-face regions. The resulting output contains only the extracted faces from the images. This pre-processing step is depicted in Figure 6. Figure 6(a) clearly depicts the pre-processing stage, in which image sizes are resized from (599,316) pixels as illustrated in Figure 6(b) to (160,160) pixels, where only the face is extracted and aligned via MTCNN and then normalized and further feature extraction is done via using FaceNet model.

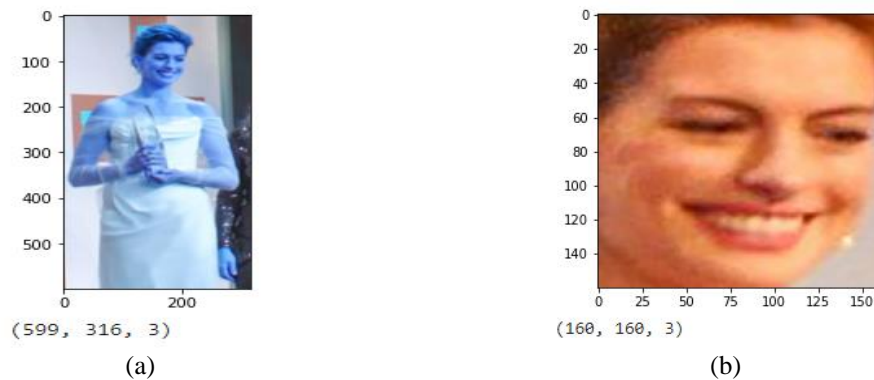


Figure 6. Images prior and after pre-processing using MTCNN (a) prior-processing and (b) post processing using MTCNN

Then, as shown in Table 3, FaceNet plays a role in the feature extraction stage, where additional features are extracted and normalized. The input for this extraction method was 3 channel (RGB) photos, and the output was 128-dimensional vectors [31]. With its 22 layers, FaceNet is able to accurately and efficiently extract features from facial images, and its output/features may be trained into 128-dimensional embeddings [32].

Table 3. Training and testing set images

		No. of extracted features
No. of Training Set	6075	128-dimentional
No. of Testing Set	130	128-dimentional

The results of this stage, which is the modelling stage, where training takes place and SVC plays an effective role in classifying images that can fit the model. SVC creates boundaries between two neighbouring classes. When a class's nearest point is on the Hyperplane, we say the class's SVM is at the margin. Surprisingly, both the testing and training sets were extremely accurate. As shown in Table 4, the training set has an accuracy of 99.5% and the testing set has an accuracy of around 94%.

Table 4. Accuracy of the training and testing sets

Proposed method	MTCNN+ FaceNet+SVC	No. of extracted features	Accuracy%
No. of Training Set	6075	128	99.5%
No. of Testing Set	130	128	95%

The essential components of the proposed model are presented in Table 4, together with the number of training and tested datasets. It also includes the accuracy percentage of the trained and tested datasets. The final phase of this model is the prediction phase, in which a random face is chosen from the test set, and the name of that randomly chosen face is expected and plotted. As a result, the expected name will correctly predict the face and plot it, as shown in Figure 7.

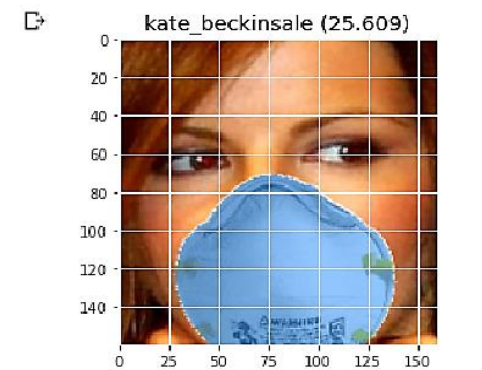


Figure 7. The correctly predicted face

Therefore, Figure 7 clearly demonstrates the correctly predicted face, and how it is correctly plotted. As a result, we can see that faces are all correctly predicted regardless of how they were occluded or covered because they will still be recognized and verified, and here is the predicted face that was captured by the webcam as shown in Figure 8. Figure 8 clearly shows that a computer camera recognized and verified the face even when it was obscured by an eyeglass or a mask. These pictures have also been added to the datasets and trained as well, so it was easy for the model to recognize the identity of this person.

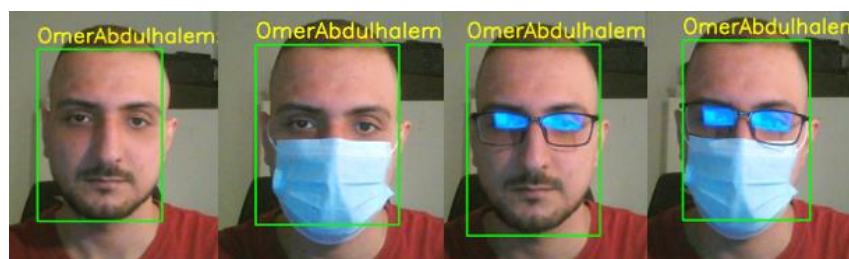


Figure 8. Faces captured via webcam

5. RESULTS EVALUATION

An evaluation of this paper's findings is provided in Table 5, which summarises the evaluation in terms of a comparison between the paper's proposed techniques, recognition rates, and database types with that of previous journal research papers. Table 5 illustrates the previous approaches along with their techniques, datasets, and accuracy percentage. Our proposed model uses different technique than the rest of the previous researchers, but it uses two exact same datasets that most of the mentioned research techniques share in common, such as Celeb A dataset and LFW dataset. Therefore, as you may clearly see that the proposed approach has much higher accuracy percentage.

Table 5. Results comparison with previous techniques

Related work	Techniques	Datasets	Recognition rate %
Ouannes <i>et al.</i> [8]	SURF + k-NN & KD-Tree	IST-EURECOM LFFD	72%
Zhang and Junyong [9]	HOG-LPB	LFW	88.5%
Xu <i>et al.</i> [10]	OREO	ORL/CelebA UHDB-31 CelebA	95%
Song <i>et al.</i> [11]	Pairwise differential Siamese network	MegaFace Challenge/LFW	74.40%
Huang <i>et al.</i> [12]	FaceNet, MaskNet, CosFace, and ArcFace	WiderFace/CelebA	78.25%
Dong <i>et al.</i> [13]	OA-GAN	CelebA	–
Zhao <i>et al.</i> [14]	RLA	LFW	–
Xu <i>et al.</i> [15]	DCGANs	AR	–
Li <i>et al.</i> [16]	IP-GANs	LFW	85.50%
Ferdinand <i>et al.</i> [17]	MTCNN/FaceNet/SVM	CelebA/NUAA	79%
Our Proposed System	MTCNN + FaceNet + SVC	CelebA & MFR2 & WiderFace & LFW & MegaFace Challenge	99.6%

6. CONCLUSION AND FUTURE WORK

The primary goal of this paper is to accurately recognize faces obscured by masks, hats, eyeglasses, or scarves. This goal has been achieved by applying three types of python-based algorithms: MTCNN, FaceNet, and SVC. MTCNN is in charge of detecting only the faces of the images that have been chosen as a dataset. This model has been trained on five types of datasets: CelebA, MFR2, WiderFace, LFW, and MegaFace Challenge. After MTCNN has correctly positioned and detected the face landmarks, the FaceNet deep learning algorithm begins to function as a facial feature extractor. FaceNet also performs additional normalization on the extracted features. SVC eventually begins to process the entire dataset and classifies it in a way that can fit the model and help to increase the probability prediction and accuracy. This proposed method has an accuracy of 99.50% of the occluded faces, making it more effective and accurate than previous proposed methods and techniques.

In the future, the same methodology will be used, but a different classifier will be used. Following that, a rate of accuracy comparison will be made between the two classifiers. As a result, the one with the highest accuracy rate will be chosen and used. In addition, more datasets will be added to include as many samples as possible, making the model more applicable and accurate.





REFERENCES

- [1] S. Bahroun, R. Abed, and E. Zagrouba, "KS-FQA: Keyframe selection based on face quality assessment for efficient face recognition in video," *IET Image Processing*, 2021, doi: 10.1049/ijpr.12008.
- [2] M. T. Ghazal and K. Abdullah, "Face recognition based on curvelets, invariant moments features and SVM," *TELKOMNIKA (Telecommunication Computing Electronics and Control)*, vol. 18, no. 2, pp. 733-739, 2020, doi: 10.12928/telkomnika.v18i2.14106.
- [3] A. S. Sanchez-Moreno *et al.*, "Efficient face recognition system for operating in unconstrained environments," *Journal of Imaging*, vol. 7, no. 9, p. 161, 2021, doi: 10.3390/jimaging7090161.
- [4] D. A. Maharani, C. Machbub, P. H. Rusmin, and L. Yulianti, "Improving the capability of real-time face masked recognition using cosine distance," in *2020 6th International Conference on Interactive Digital Media (ICIDM)*, 2020, pp. 1-6, doi: 10.1109/ICIDM51048.2020.9339677.
- [5] M. Aria, V. Agnihotri, A. Rohra, and R. Sekhar, "Secure online payment with facial recognition using MTCNN," *International Journal of Applied Engineering Research*, vol. 15, no. 3, pp. 249-252, 2020, doi: 10.47577/ijaer.v15i3.2121.
- [6] S. Sreesuba, G. Anitha, A. Irumporai, S. Usha, and P. S. Devi, "Facial recognition-based attendance marking system," *Annals of the Romanian Society for Cell Biology*, vol. 25, no. 6, pp. 6452-6459, 2021, doi: 10.2478/arscb-2021-0103.
- [7] T. Liu, L. Jin, C. Zhong, and F. Xue, "Study of thermal sensation prediction model based on SVC algorithm with data preprocessing," *Journal of Building Engineering*, vol. 48, p. 103919, 2022, doi: 10.1016/j.job.2021.103919.





- [8] L. Ouannes, A. B. Khalifa, and N. E. B. Amara, "Comparative study based on de-occlusion and reconstruction of face images in degraded conditions," *Traitement du Signal*, vol. 38, no. 3, pp. 251-262, 2021, doi: 10.18280/ts.380305.
- [9] J. Zhang and W. Junyong, "Local occluded face recognition based on HOG-LBP and sparse representation," in *2020 IEEE International Conference on Artificial Intelligence and Computer Applications (ICAICA)*, 2020, pp. 85-90.
- [10] X. Xu, N. Sarafianos and I. A. Kakadiaris, "On improving the generalization of face recognition in the presence of occlusions," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, 2020, pp. 3996-4005, doi: 10.1109/CVPRW50498.2020.00407.
- [11] L. Song, D. Gong, Z. Li, C. Liu, and W. Liu, "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, pp. 4100-4109, doi: 10.1109/ICCV.2019.00086.
- [12] B. Huang *et al.*, "When face recognition meets occlusion: a new benchmark," in *ICASSP 2021-2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2021, pp. 1482-1486, doi: 10.1109/ICASSP39728.2021.9413893.
- [13] J. Dong, L. Zhang, H. Zhang, and W. Liu, "Occlusion-aware gan for face de-occlusion in the wild," in *2020 IEEE International Conference on Multimedia and Expo (ICME)*, 2020, pp. 1-6, doi: 10.1109/ICME46284.2020.9102788.
- [14] F. Zhao, J. Feng, J. Zhao, W. Yang, and S. Yan, "Robust LSTM-autoencoders for face de-occlusion in the wild," *IEEE Transactions on Image Processing*, vol. 27, no. 2, pp. 778-790, 2018, doi: 10.1109/TIP.2017.2771408.
- [15] L. Xu, H. Zhang, J. Raitoharju, and M. Gabbouj, "Unsupervised facial image de-occlusion with optimized deep generative models," in *2018 Eighth International Conference on Image Processing Theory, Tools and Applications (IPTA)*, 2018, pp. 1-6, doi: 10.1109/IPTA.2018.8608127.
- [16] C. Li, S. Ge, Y. Hua, H. Liu, and X. Jin, "Occluded face recognition by identity-preserving inpainting," in *International Symposium on Artificial Intelligence and Robotics*, 2020, pp. 291-302.
- [17] J. Ferdinand, C. Wijaya, A. N. Ronal, I. S. Edbert, and D. Suhartono, "ATM security system modeling using face recognition with FaceNet and haar cascade," in *2022 6th International Conference on Informatics and Computational Sciences (ICICoS)*, 2022, pp. 111-116, doi: 10.1109/ICICoS56336.2022.9930563.
- [18] F. Schroff, D. Kalenichenko, and J. Philbin, "Facenet: A unified embedding for face recognition and clustering," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2015, pp. 815-823, doi: 10.1109/CVPR.2015.7298682.
- [19] E. Jose, M. Greeshma, M. T. P. Haridas, and M. H. Supriya, "Face recognition-based surveillance system using facenet and mtcnn on jetson tx2," in *2019 5th International Conference on Advanced Computing and Communication Systems (ICACCS)*, 2019, pp. 608-613, doi: 10.1109/ICACCS.2019.8728466.
- [20] G. C. Suguna, H. S. Kavitha, and S. Shirahatti, "Face recognition system for real time applications using SVM combined with facenet and MTCNN," *International Journal of Electrical Engineering and Technology (IJEET)*, vol. 12, pp. 328-335, 2021.
- [21] J. Du, "High-precision portrait classification based on MTCNN and its application on similarity judgement," *Journal of Physics: Conference Series*, vol. 1518, no. 1, IOP Publishing, 2020, doi: 10.1088/1742-6596/1518/1/012066.
- [22] F. Cahyono, W. Wirawan, and R. F. Rachmadi, "Face recognition system using facenet algorithm for employee presence," in *2020 4th International Conference on Vocational Education and Training (ICOVET)*, 2020, pp. 63-67, doi: 10.1109/ICOVET50258.2020.9229888.
- [23] V. D. Rangari, "Dominant facial emotion recognition using mtcnn in Python," *International Research Journal of Modernization in Engineering Technology and Science*, vol. 04, no. 10, 2022, doi: 10.56726/IRJMETS30585.
- [24] D. Kumar, "Introduction to facenet: a unified embedding for face recognition and clustering," *Medium, Analytics Vidhya*, 2020.
- [25] A. Joseph and P. Ramamoorthy, "Facial expression recognition using SVC classification and INGI method," *International Journal of Science and Research (IJSR)*, vol. 2, no. 4, pp. 296-300, 2013.
- [26] "Large-Scale CelebFaces Attributes (CelebA) Dataset," CelebA Dataset, <http://mmlab.ie.cuhk.edu.hk/projects/CelebA.html>, (accessed on: 28 November 2022)
- [27] A. Aqeel, "Aqeelanwar/mask the face: convert face dataset to masked dataset," GitHub, <https://github.com/aeqelanwar/MaskTheFace>, (accessed on: 28 November 2022)
- [28] "Wider_face: tensorflow datasets," TensorFlow, https://www.tensorflow.org/datasets/catalog/wider_face, (accessed on: 28 November 2022)
- [29] "Papers with code - LFW dataset," Papers with Code, <https://paperswithcode.com/dataset/lfw>, (accessed on: 28 November 2022)
- [30] "Papers with code - megaface dataset," Papers with Code, <https://paperswithcode.com/dataset/megaface>, (accessed on 26 November 2022)
- [31] J. A. C. Moreano and N. B. L. S. Palomino, "Global facial recognition using gabor wavelet, support vector machines and 3D face models," *Journal of Advances in Information Technology*, vol. 11, no. 3, pp. 169-176, 2020, doi: 10.12720/jait.11.3.143-148.
- [32] N. Darapaneni *et al.*, "Automatic face detection and recognition for attendance maintenance," in *2020 IEEE 15th International Conference on Industrial and Information Systems (ICIIS)*, 2020, pp. 1-6, doi: 10.1109/ICIIS51140.2020.9342670.

BIOGRAPHIES OF AUTHORS







Omer Abdulhaleem Naseer     is a third-year Ph.D. student at University Putra Malaysia, Faculty of Engineering. Majored in Computational Methods in Engineering. The specified research field is bio-metrics, particularly, facial recognition for occluded faces. He can be contacted via this email: omar.abdulhaleem592@gmail.com.







Sharifah Mumtazah Syed Ahmad (Assoc. Prof. Dr.)     An associate professor who currently works as a senior lecturer and supervisor at University Putra Malaysia, faculty of Engineering. Dr. Sharifah is specialized in many different areas, including Biometrics, Security, Image and Signal Processing, and Machine Learning. She has earned her Ph.D. in Electronics in 2004 from the University of Kent, United Kingdom. Currently she has over 50 cited publications with scopus h index of 12. She can be contacted via this email: s_mumtazah@eng.upm.edu.my.







Khairulmizam Samsudin (Dr.)     A Dr. who currently works as a senior lecturer and supervisor at University Putra Malaysia, Faculty of Engineering. Dr. Khairulmizam's areas of specialties include Embedded Systems, Robotics, and Intelligent Systems. He can be contacted via this email: khairulmizam@upm.edu.my.







Dr. Marsyita Hanafi     A Dr. who currently works as a senior lecturer and supervisor at University Putra Malaysia, faculty of Engineering. Dr. Marsyita is specialized in many different areas, including precision agriculture, smart monitoring systems, biometrics, security, image processing, and machine learning. She can be contacted via this email: marsyita@upm.edu.my.



Siti Mariam Binti Shafie     received her M.Sc. and B.Eng. degrees in Faculty of Engineering from Universiti Putra Malaysia in 2003 and 1999, respectively. She is currently working at the same university in the Department of Computer and Communication Systems Engineering. Her research interests include embedded system, IoT, and machine learning. She can be contacted at email: mariam@upm.edu.my.



Ts. Nor Zarina Zamri     was appointed as Research Lead for Malaysian Security Evaluation Facility Department under CyberSecurity Malaysia (CSM MySEF). Background in Bachelor (BCs Hons) Computer Science (Software Engineering) and Master of Science (MSc) in Information Technology. She specializes in smart card, biometric and hardware security and leads teams in cyber security research exploration, explores new technologies, and produces in-house test tools, test methods, guidelines, and training. She can be contacted via this email: norzarina.zamri@cybersecurity.my.