6th International Conference on Advanced Technologies
for Signal and Image Processing - ATSIP'2022
May 24-27, 2022, Canada-Tunisia



2022 6th International Conference on Advanced Technologies for Signal and Image Processing (ATSIP) | 978-1-6654-5116-1/22/$31.00 ©2022 IEEE | DOI: 10.1109/ATSIP55956.2022.9805960

# Transfer Learning for Robust Masked Face Recognition

Omar Adel Muhi
ATISP research unit
École Nationale d'Électronique
et de Télécommunications de Sfax (ENET'Com)
omar.muhi@yahoo.com

Mariem Farhat
ATISP research unit
Higher Institute of Computing
and Multimedia of Gabes
mariem.farhat@isimg.tn

Mondher Frikha
ATISP research unit
École Nationale d'Électronique
et de Télécommunications de Sfax (ENET'Com)
mondher.frikha@enetcom.usf.tn

***Abstract*— Over the past three years, it has seen many outbreaks of different coronavirus diseases around the world. One of the ways of transmission of COVID 19 is air transport. This transmission occurs when humans breathe in droplets released by an infected person by breathing, speaking, coughing or sneezing. The World Health Organization (WHO) has given orders to wear a face mask in the public places.**

**Despite the impressive results achieved by deep learning methods in face recognition, the performance of these methods deteriorates when wearing a mask. The issue of masked face recognition is attracting more attention.**

**In this work, a simple and effective method is proposed to deal with the recognition of people who wear a mask. We divide the problem into three phases: feature detection, feature extraction and recognition. The first stage of our model with MediaPipe Face Mesh automatically produces a segmentation of the masked area as feature detection and several points for cropping the area of interest. Then, the second stage extracts the features gained using Resnet50. The selected network, based on ResNet-50, is modified so that the third stage performs the classification process with Soft Max as an redaction and activation function.**

**To train our system, we used the Labeled Faces in the Wild (LFW). From the original facial recognition data set, a masked copy is generated using data augmentation, and the two data sets are combined during the training process. Our model achieved a test accuracy of 98%.**

***Keywords—masked face recognition, LFW, ResNet-50, SoftMax***

## I. Introduction

In April 2020 there is a research has been published by CSIS it discussed the matter of facial recognition system that has absolute precision in the ideal condition, reached to 99.97% of recognition accuracy level [1]

Moreover, while the face recognition has reached an unprecedented level in perfect conditions, this does not scale to daily and real-world conditions. In fact, Ageing, Low resolution, lighting, facial covering, are the factors which subject under real world condition, they affect by accuracy of facial recognition technology [2].

Go back to our real life and for the last 3 years, people have been wearing a mask to prevent covid-19 disease. The World Health Organization has proven that when we wear face masks, it prevents the transmission of the Corona virus. This face mask covers Large part from face including the nose and mouth. This is a special case of occlusion that makes the task of recognition more difficult.

Thus, masked face recognition is considered as a crucial, urgent and immediate challenge to defeat with new face recognition methods and systems. NIST [2] reviewed the performance of FR algorithms before and after the COVID-19 pandemic. They evaluated existing (pre-pandemic) algorithms after they had been modified to deal with maskers. The main conclusion from this study is that the accuracy of the recognition system drops from 100% to 96% when a person wears a face mask.

Researchers have noted many difficult to deal with this issue. Firstly, Most of the current advanced face recognition methods are based on deep learning, which mainly based on a large number of training samples. However, there is currently a shortage of large, publicly available facial data set with masks [3]. Second, because the mask covers a large part of the face where there are abundant features such as the mouth and nose, the traditional facial recognition algorithm may not work effectively. Finally, it is difficult to determine which face is wearing the mask.

To meet the challenge of wearing masks, It is necessary to improve existing facial recognition algorithms.

Hence, we state a masked face recognition model that depends on deep transfer learning using ResNet50.

978-1-6654-5116-1/22/$31.00 ©2022 IEEE

The organization for the rest of the paper is as follows. Section 2 reviews relevant previous work. Section 3 explains the proposed model in detail. Section 4 describes the characteristics of the dataset. Section 5 includes there reporters and analysis of results and Section 6 includes the possibilities and conclusions of the next work.

## II. RELATED WORKS

Face recognition (FR) has a long standing research topic that related to computer vision. With deep learning techniques, the majority of FR systems have been transformed to implement deep learning models and accuracy has been greatly enhanced to over 99.80% in just a few years approaching human performance in an unconstrained state [4]. Hence, the face recognition tends to make far from satisfactorily to encounter challenges like large pose variation, low resolution facial expressions that are difference and conclusion.

Occluded face recognition (OFR) It considers one of the difficult problems because the researcher can't reach to the knowledge of the conclude part, that could be in everywhere and any form in the face image. While, masked face recognition (MFR) is a specific case of the OFR where the occluded part of the face is known. Thus, the MFR task can be considered easier to solve and over the past three years there has been a rapid growth in the amount of research work in the field of MFR.

Exhaustive surveys on FR [4]–[6][8] [9], OFR [2], [10][11] and MFR [12] have been published.

Although it is an important part of recognition systems, the problem of occluded facial images, including masks, has not been fully addressed.

Since the emergence of COVID-19, many works have been submitted to solve the task of recognizing masked faces by using different methods that can be categorized into three groups.

The first group is based on matching based methods [11], [16] Which attempts to compare the similarity between images using the matching process.

The face image is sampled at a number of points of the same size. The feature extraction is then applied to each patch. Finally, the matching process is applied between the interface of the probe and the gallery. These methods treat MFR as a normal face recognition problem and do not take into account damage to the masked area and lack of texture features in the area around the mouth and nose. Moreover, the performance of these is very sensitive to different sampling strategies.

The second group is based on discarding methods [14] [15] [16] which rejects the occluded parts completely to avoid inefficient features, so they use only features extracted from the visible parts for recognition. This strategy become more interesting when dealing with masks because we already know the position of the occluded parts on the face.

For instance, Hariri [14] proposed a masked-aware face feature embedding By cropping the eyes and forehead, and then using a quantization-based clustering method on a pre-trained VGG-16 model to extract deep features from unmasked areas.

This approach makes the MFR applicable in real-time applications because it handles only the most interesting part of the image.

Masker recovery which aims to recover the covered parts of the images based on the images in the training in the third group [17]. According to this strategy, it decreases the difficulties of recognition largely.

For example, System et al. [17] The second stage removes the mask and aggregates the affected region in fine detail while retaining the global coherence of the face structure using a GAN-based network after producing a binary segmentation of the mask region.

In [18], the authors suggested drawing the closed segment using GAN integration with the previously improved CNN recognizer. With the goal, a set of identity-centric features are applied to the recognizer as a moderation to enable colored faces to converge towards the centers of their own identity. In this way, their approach can take advantage of GAN for reconstruction and CNN for representation, while simultaneously addressing two difficult tasks, namely, face painting and face recognition.

The reconstructed faces are synthetic and their reliability depends on the quality of the data, network and training process in addition, this strategy is not good and is not preferred for real-world applications because it is time-consuming. In our work, we set two challenges: (1) a robust system that can deal with both masked and non-masked faces that it could achieve a good performance in both evaluation sets, (2) a solution for the lack of public masked dataset.

In order to reach these goals, we propose a system based on occlusion discard strategy using a transfer learning approach.

On the one hand, The most advanced face recognition based on deep learning [19] [20] introduce several CNN architecture that already achieved a very good performance in the FR task. On the other hand, these architectures don't require a huge number of images to fine tuning them compared with a solution from scratch [21]. Thus, we propose a system based on ResNet-50 [22] to extract deep features from the non-occluded part of the masked faces. Nevertheless, we propose a data augmentation through the generation of masked images using MediaPipe face Mesh [23] applied on the LFW dataset [24].

## III. THE PROPOSED MODEL

In this section, we will present the detailed structure for our proposed model based on MediaPipe Landmark activation, Resnet-50 and SoftMax.

Our model is divided into three components Fig. 1: first, the face detection using Media Pipe face Mesh which estimate 468 3D face landmarks on devices in real-time. Second, the segmentation component which extract the region of interest that contains most important features in the non-occluded part of the face. Finally, the ResNet50 component which performs the feature extraction and the classification process using the SoftMax activation function.

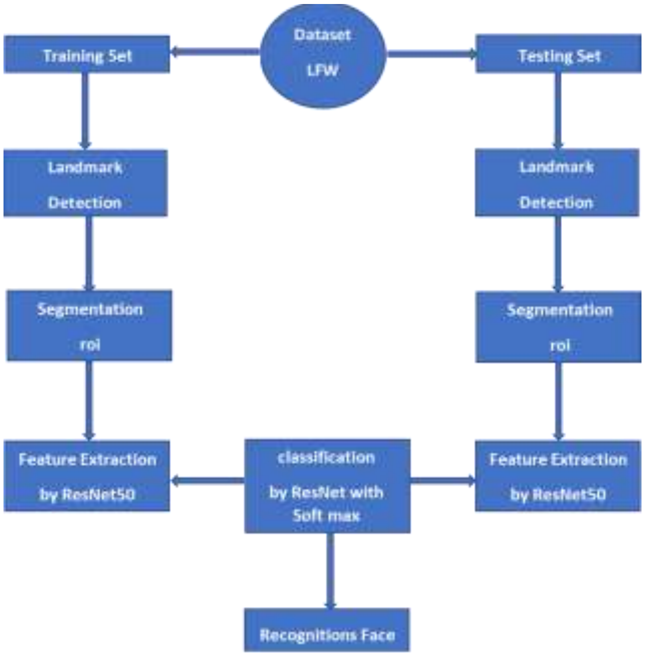


Fig. 1: The Proposed System

### A. Face Detection

First, the masked face is detected using a landmark detector.

There are many landmark face detection, like Dlib (68, 85) [25] and the Media Pipe Face Mesh [23] Fig.2. the Media pipe landmark have two advantages; first it is more suitable for a real time application, second it provides a huge number of points around the face. These landmarks are used for face correction before feature extraction and for data augmentation in an offline process to generate new masked images.

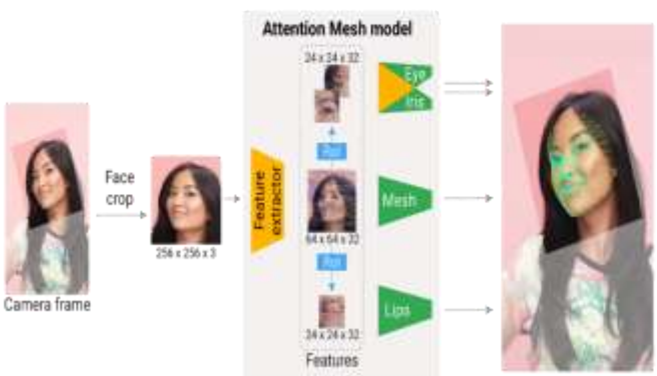


Fig. 2: Face landmarks: The black number dots represent the 468 landmarks in 3D [28].

### B. Segmentation

In order to crop the region of interest and maintain only the non-occluded region, we set four static points of this area using extracted vertex of the detected mesh in the last component.

To get the correct image dimensions, we multiply the point by the length and width. Using Region of interest (ROI) function we cover the area with a rectangle. We tried to extract this region using 10, 8, 6 and 4 points. After several attempts, we retain only 4 points in order reduce the computational time. We make a full mask for each covering image and make it black, except for the selected area, which will be in white. Finally, we apply an AND operation to this image with the original morphological image and we got a picture of the selected part sized 100*100 Fig. 3.

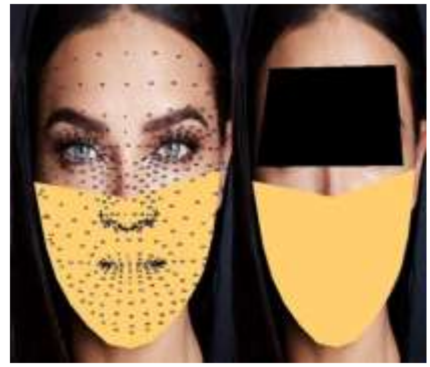

Fig. 3: Landmark detection and a Region of Interest cropped

### C. Feature-Extractor and Classifier

They proven that ResNet50 [26] reached to the best result because it has used as feature extractor. ResNet50 represents deep learning based on residual learning [22]. Also ResNet.50 with about 50 layers, starting involution layer, then ending with connection layers. It contains 16 residual bottleneck block which everyone joins 3 layers.

The SoftMax loss and its variants are widely used as objectives for face recognition [27]. We used the following definition of the SoftMax function:

$$softmax(z_i) = \frac{\exp(z_i)}{\sum_j z_j} \qquad (1)$$

Where $z_i$ represents values from the neurons of the output layer. The exponential works as a non-linear function. Later these values are divided by the sum of the exponential values in order to normalize them and then convert them to probabilities.

SoftMax is used in polynomial logistic regression and is often used as the redaction and activation function of a neural network to normalize the network's output to a probability distribution on expected output classes, based on the loss selection axiom, a logistic generalization that operates on multiple dimensions [29].

## IV. Experimentation Results

### A. Dataset

The first step is to create a training dataset containing pre-processed input images for mask face recognition. We apply a mask on each image of the LFW dataset [24] in offline using a

tool Mask The Face[1]. We generate for each person five masked images. The total number of images without the mask 20 photos * 100 people = 2000 photos.
And we split our dataset in three categories as following:

1- Number of images used in training A (unmasked) = 2000
2- Number of images used in training B (masked and unmasked) = 2500
3- The number of images used in the test (only masked faces) = 500

During the training, people are chosen at random.

### B. Implementation Details

All the experimental trials were conducted on a computer HP sever has equipped by processor (2 GHz), Core (TM) i7. All implementation was done using Python 3.8**.**

The initial learning rate is set to 0.1. We have tried 10, 50, 100 epochs, and stop training at 100 epochs. For loss function pool from 20.7 to 0.28.

### C. Results

We evaluate our approach using the following metrics:

$$Accuracy = \frac{(TP+TN)}{(TP+TN)+(TN+FN)} \quad (2)$$

We have a lot of symbols, one of them is TP for True Positive, also FP is abbreviated to False Positive and FN for False negative sample.
The researchers [28] build their approach based on deep transfer learning in order to enhance image of classification accuracy, whereas the last is not acceptable. We are going to discuss the achieved results for the joint use of ResNet and softmax classifier in this work. We will look also for the results using two methods of face detection Dlib, and Media Pipe Face Mesh as shown in the Table (1).
Fig. 4 illustrates the achieved percentage for the training accuracy evolution. We have achieved an accuracy around **100%** for each scenario.

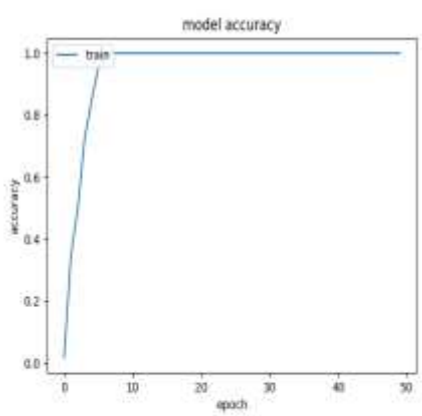


Fig. 4: Accuracy function on the training set

TABLE I. RESULTS FOR DIFFERENT SCENARIOS

| Scenario | Epoch | Detection model | Train Image | Test Image | Train Acc (%) | Test Acc (%) |
|---|---|---|---|---|---|---|
| Scenario 1 | 100 | MediaPipe | faces without mask | Masked Faces | 100 | 97.34 |
| Scenario 2 | 50 | MediaPipe | faces without mask | Masked Faces | 100 | 97 |
| Scenario 3 | 100 | MediaPipe | faces without mask | Masked Faces | 100 | **98.3** |
| Scenario 4 | 50 | MediaPipe | faces without mask | Masked Faces | 100 | 98.1 |
| Scenario 5 | 100 | Dlib | faces without mask | Masked Faces | 98 | 96 |
| Scenario 6 | 50 | Dlib | faces without mask | Masked Faces | 95 | 95.2 |
| Scenario 7 | 100 | Dlib | faces without mask + Masked Faces | Masked Faces | 100 | 97.1 |
| Scenario 8 | 50 | Dlib | faces without mask + Masked Faces | Masked Faces | 100 | 97 |

Table I. presents a comparison between different scenarios where both masked and unmasked faces are used independently and jointly. Using unmasked images and masked images for the training of the considered face recognition model, MediaPipe, ResNet-50 and SoftMax, achieve a very high verification performance **98.3%** accuracy. The verification performances of the considered models are substantially degraded when only unmasked face images are considered.
We make also the observation that the Media pipe Landmark return better recognition accuracy than Dlib in all scenarios.

## V. COMPARISON WITH RELATED WORKS

The work presented in [3] used the same data sets, the LFW pseudo-disguised data set. The authors [3] achieved test accuracy ranging from 50% to 95%. In the presented work, the accuracy of the test was tested with a ResNet-50 classifier. In this work, we report an accuracy of between 97% and 98%.

## VI. CONCLUSION AND FUTURE WORKS

In this work, we seek to investigate the performance of face recognition models on masked faces. The proposed model suggests that feature extraction has used ResNet50. It represents models in transfer learning. The Res Net with Soft max classifier in LFW database achieved 98.3% testing accuracy. The result after making the comparison has been applied with same works. Also the proposed model has been achieved the successful in the testing accuracy. Show our model

[1] https://github.com/aqeelanwar/MaskTheFace

tested on these datasets shows better recognition rates. So ResNet-50 gives better accuracy for simple recognition of masked faces. In subsequent work, it is of paramount importance to enhance and extend our work to address the case of various extreme masks for face recognition.

## VII. References


[1] P. Grother, M. Ngan, and K. Hanaoka, "Ongoing Face Recognition Vendor Test (FRVT) Part 1: Verification," 2020.

[2] D. Zeng, R. Veldhuis, and L. Spreeuwers, "A survey of face recognition techniques under occlusion," IET Biometrics, vol. 10, no. 6, pp. 581–606, 2021.

[3] Z. Wang et al., "Masked face recognition dataset and application," arXiv, pp. 1–3, 2020.

[4] M. Wang and W. Deng, "Deep face recognition: A survey," Neurocomputing, pp. 1–31, 2020.

[5] U. Jayaraman, P. Gupta, S. Gupta, G. Arora, and K. Tiwari, "Recent development in face recognition," Neurocomputing, vol. 408, pp. 231–245, 2020.

[6] G. Guo and N. Zhang, "A survey on deep learning based face recognition," Comput. Vis. Image Underst., vol. 189, no. July, p. 102805, 2019.

[7] S. Madhavan and N. Kumar, Incremental methods in face recognition: a survey, vol. 54, no. 1. Springer Netherlands, 2021.

[8] M. O. Oloyede, G. P. Hancke, and H. C. Myburgh, "A review on face recognition systems: recent approaches and challenges," Multimed. Tools Appl., vol. 79, no. 37–38, pp. 27891–27922, 2020.

[9] B. Lahasan, S. L. Lutfi, and R. San-Segundo, "A survey on techniques to handle face recognition challenges: occlusion, single sample per subject and expression," Artif. Intell. Rev., vol. 52, no. 2, pp. 949–979, 2019.

[10] L. Ouannes, A. Ben Khalifa, and N. E. Ben Amara, "Deep Learning vs Hand-Crafted Features for Face Recognition under Uncontrolled Conditions," 2019 Int. Conf. Signal, Control Commun. SCC 2019, pp. 185–190, 2019.

[11] D. Zeng, R. Veldhuis, and L. Spreeuwers, "A survey of face recognition techniques under occlusion," arXiv, pp. 1–23, 2020.

[12] A. Alzu, F. Albalas, T. Al-hadhrami, L. B. Younis, and A. Bashayreh, "Masked Face Recognition Using Deep Learning : A Review," 2021.

[13] V. Aswal, O. Tupe, S. Shaikh, and N. N. Charniya, "Single Camera Masked Face Identification," pp. 57–60, 2021.

[14] W. Hariri, "Efficient Masked Face Recognition Method during the COVID-19 Pandemic," 2020.

[15] Q. Wang, T. Wu, H. Zheng, and G. Guo, "Hierarchical Pyramid Diverse Attention Networks for Face Recognition," pp. 8326–8335.

[16] Y. Li, K. Guo, Y. Lu, and L. Liu, "Cropping and attention based approach for masked face recognition," Appl. Intell., 2021.

[17] N. Ud Din, K. Javed, S. Bae, and J. Yi, "A Novel GAN-Based Network for Unmasking of Masked Face," IEEE Access, vol. 8, pp. 44276–44287, 2020.

[18] S. Ge, C. Li, S. Zhao, and D. Zeng, "Occluded face recognition in the wild by identity-diversity inpainting," IEEE Trans. Circuits Syst. Video Technol., vol. 30, no. 10, pp. 3387–3397, 2020.

[19] Y. Taigman, M. A. Ranzato, T. Aviv, and M. Park, "DeepFace : Closing the Gap to Human-Level Performance in Face Verification."

[20] F. Schroff and J. Philbin, "FaceNet: A Unified Embedding for Face Recognition and Clustering."

[21] A. Brodzicki, M. Piekarski, and D. Kucharski, "Transfer Learning Methods as a New Approach in Computer Vision Tasks with Small Datasets Introduction to Transfer Learning Methods," vol. 45, no. 3, 2020.

[22] K. He and J. Sun, "Deep Residual Learning for Image Recognition," pp. 1–9.

[23] C. Lugaresi et al., "MediaPipe: A Framework for Building Perception Pipelines."

[24] G. B. Huang et al., "Labeled Faces in the Wild : A Database forStudying Face Recognition in Unconstrained Environments To cite this version : HAL Id : inria-00321923 Labeled Faces in the Wild : A Database for Studying Face Recognition in Unconstrained Environments," 2008.

[25] S. Sharma, K. Shanmugasundaram, and S. K. Ramasamy, "FAREC - CNN Based Efficient Face Recognition Technique using Dlib," no. 978, pp. 192–195, 2016.

[26] M. Loey, G. Manogaran, M. Hamed, N. Taha, N. Eldeen, and M. Khalifa, "A hybrid deep transfer learning model with machine learning methods for face mask detection in the era of the COVID-19 pandemic," Measurement, vol. 167, no. July 2020, p. 108288, 2021.

[27] X. An et al., "Partial FC : Training 10 Million Identities on a Single Machine," pp. 1445–1449.

[28] B. Mandal, "Masked Face Recognition using ResNet-50," 2012.

[29] A. GravesIan, G. Yoshua, B. Geoffrey, H. Yann, L. Andrew, N. Demis, H. David, S. Fei-Fei Li. https://en.wikipedia.org/wiki/Softmax_function October 2020 (UTC).