Date of publication xxxx 00, 0000, date of current version xxxx 00, 0000.

Digital Object Identifier 10.1109/ACCESS.2017.DOI

Towards Accurate and Lightweight Masked Face Recognition: an Experimental Evaluation

YOANNA MARTÍNEZ-DÍAZ[1], HEYDI MÉNDEZ-VÁZQUEZ[1], LUIS S. LUEVANO[2], MIGUEL NICOLÁS-DÍAZ[1], LEONARDO CHANG[2] and MIGUEL GONZALEZ-MENDOZA[2]

[1]Advanced Technologies Application Center (CENATAV), Havana, Cuba (e-mail: ymartinez, hmendez@cenatav.co.cu)

[2]Tecnologico de Monterrey, Campus Estado de México, Estado de México, México (e-mail: luis.s.luevano, lchang, mgonza@tec.mx)

Corresponding author: Leonardo Chang (e-mail: lchang@tec.mx).

ABSTRACT Given the current COVID-19 pandemic, most people wear a mask to effectively prevent the spread of the contagious disease. This sanitary measure has caused a significant drop in the effectiveness of current face recognition methods when handling masked faces on practical applications such as face access control, face attendance, and face authentication-based mobile payment. Under this situation, recent efforts have been focused on boosting the performance of the existing face recognition technology on masked faces. Some solutions trying to tackle this issue fine-tune the existing deep learning face recognition models on synthetic masked images, while others use the periocular region as a naive manner to eliminate the adverse effect of COVID-19 masks. Although the accuracy of masked face recognition remains an important issue, in the last few years, the development of efficient and lightweight face recognition methods has received an increased attention in the research community. In this paper, we study the effectiveness of three state-of-the-art lightweight face recognition models for addressing accurate and efficient masked face recognition, considering both fine-tuning on masked faces and periocular images. For the experimental evaluation, we create both real and simulated masked face databases as well as periocular datasets. Extensive experiments are conducted to determine the most effective solution and state further steps for the research community. The obtained results disclose that fine-tuning exiting state-of-the-art face models on masked images achieves better performance than using periocular-based models. Besides, we evaluate and analyze the effectiveness of the trained masked-based models on well-established unmasked benchmarks for face recognition and asses the efficiency of the used lightweight architectures in comparison with state-of-the-art face models.

INDEX TERMS COVID-19 pandemic, lightweight deep models, masked face recognition, periocular recognition

I. INTRODUCTION

The present situation of the COVID-19 pandemic has changed the world in all dimensions. The trend of wearing face masks for all people in public places have imposed new challenges for the research community. Many applications based on face recognition techniques, such as face access control, face attendance, and face authentication based mobile payment, have nearly failed to effectively recognize the masked faces. At the moment, removing masks for passing authentication systems is not recommended since this can increase considerably the transmission of the COVID-19 virus. Furthermore, because the virus can be spread through contact, systems based on passwords or fingerprints are less safer than face recognition solutions which do not need to touch any device. Therefore, masked face recognition has become a crucial computer vision task to help the global society reduces virus infection.

Current advanced face recognition methods are based on deep learning models [10], [21], which have been able to achieve impressive performance on public benchmarks. However, deep face recognition performs poorly under new challenging conditions, such as the occlusion caused by masked faces. Face occlusion has been widely addressed by the research community in the scope of face recognition solutions [43]. Most existing works consider general occlusions that commonly appear in unconstrained capture conditions, such as sunglasses, scarves, or other random objects like books and cups [34]. The performance of these methods tends to degrade by a large margin in front of specific objects like the COVID-19 masks, that occlude a large part of the face, including important facial regions such as the mouth and the nose [29].

This is the reason why recognizing masked faces is currently an active research topic [7], [18], [37]. In the last year, recent works [4], [8], [29] have evaluated the effect of wearing a mask on automatic face recognition systems based on state-of-the-art deep Convolution Neural Networks (CNN). However, these studies focus mainly on the performance of common deep face recognition models with high computational cost. Moreover, since deep learning-based approaches depend on massive training data, databases with real face masks have been collected and tools for generating synthetic masked images have been developed [2]. Nevertheless, most of these collected datasets are not publicly available and the models are trained and evaluated under different experimental settings, which can be difficult to understand their behavior.

The use of some of the ocular traits that have been proposed for human recognition, can be regarded as a naive manner to eliminate the adverse effect of the mask. Unlike other ocular traits such as iris, retinal and conjunctival vasculature, the acquisition of the periocular biometrics does not require high user cooperation and close capture distance [31], which is in particular useful for COVID-19 pandemic. Recent methods for periocular recognition based on deep learning models have shown promising results even for in-the-wild images [13], [16], [36]. Nevertheless, since periocular biometrics encloses only the immediate vicinity of eyes, i.e., a sub-region of a face, it captures relatively less information compared with that of the face.

On the other hand, although the accuracy of face recognition systems is very important, in the last few years, the development of efficient and lightweight face recognition methods has received an increased attention in the literature. This interest has been motivated by the demand for the deployment of face recognition models in the embedded domains and other use-cases constrained by low computational power devices and high throughput requirements. Recently, the effectiveness of several lightweight face architectures was demonstrated for different face recognition scenarios, reaching high levels of accuracy and compactness with a very low computational cost [24].

In this paper, we aim at studying the effectiveness of three state-of-the-art lightweight face recognition models to enable the future development of solutions addressing accurate and efficient masked face recognition. We investigate two different approaches to enhance the performance of these models in front of masked faces. The first approach consists of including masked facial images in the learning process of the lightweight models. Due to the lack of publicly available masked face databases, we create face datasets with synthetic masks and propose a masked face dataset that simulates a realistically variant collaborative face scenario. The created real masked database is the first version of an on-going data collection process, that will be available upon request for future research and comparisons. In the experimental evaluation, we cover the matching of masked faces as well as faces with and without masks. Considering the periocular biometrics as a naive manner to address the adverse effect of the mask, the second approach applies the lightweight networks in the context of periocular recognition. In order to analyze the feasibility of using periocular information and compare its performance against the obtained by the models trained with masked faces, we create periocular datasets from the same datasets used in the masked face recognition scenario. In addition, we investigate the effect of employing the lightweight models trained with masked images on well-established benchmarks of unmasked images. Aiming at evaluating the deployment capacity of the used lightweight face networks, we assess their computational requirements and compare them with state-of-the-art models.

The main contributions of this work are summarized as follows:

- The collection of a real masked face dataset including verification and identification protocols for unconstrained masked face recognition, that will be available upon request to encourage and support future solutions for this problem.
- An extensive evaluation of the performance of lightweight deep models from two different approaches including (real and simulated) masked faces images and periocular images. This covered extended experiments with three state-of-the-art lightweight face architectures, which evidence that the masked face models should be used as the primary solution.
- Performance assessment of face models when matching masked vs. masked face images and masked vs. unmasked face images.
- Comparison of the proposed lightweight masked and periocular models with several state-of-the-art deeper models for the problem of wearing a mask on face recognition.
- A study of the effect of using lightweight masked face models on unmasked face recognition benchmarks with different covariates such as pose, age, and large-scale (unmasked vs. unmasked face images).
- An analysis and comparison of deployment capacity of the proposed lightweight solutions with several face recognition models by taking into account the storage space (model size), the compactness (number of parameters) and the Floating Point Operations Per Second (FLOPs).

The remainder of this paper is organized as follows. In Section 2, we provide an overview of related work covering both masked face recognition and periocular face recognition. In Section 3, the datasets employed for the study are described. The selected lightweight face models and the implementation details are presented in Section 4. The extensive experimental evaluation is provided in Section 5 and finally, conclusions are given in Section 6.

II. RELATED WORK

In this section, we review the face recognition methods that have been proposed to address the effect of wearing a mask in the era of the COVID-19 pandemic. We summarize existing solutions on two different approaches containing the methods that directly leading with the mask and those which use periocular region as biometric trait for recognition.

A. MASKED FACE RECOGNITION

Wearing a mask can be characterized as a kind of facial occlusion. Although there are a large number of approaches developed for face recognition in the presence of occlusions [43], most of them consider the occlusion of small regions of the face due to sunglasses, mustache, bangs or hats [34]. However, the face masks that are used to prevent COVID-19 disease, occlude around 70% of the face area [29], including the mouth, chin, and nose. Thus, specific studies and methods have been arisen during the last year for masked face recognition.

Due to lack of training and testing datasets with face images wearing masks, the first works related to masked face recognition during the COVID-19 pandemic are based on the creation of databases with real or simulated face masks. Wang et al. [38] were the first in proposing real and simulated masked face datasets, including a large Real-world Masked Face Recognition Dataset (RMFRD) and Simulated Masked Face Recognition Dataset (SMFRD). Although the authors claim to enhance the recognition accuracy from 50% to 95%, this dataset is a little hard to be used since it has not clearly defined an evaluation protocol. Moreover, details of their method and baseline are not clearly specified. In [2], a methodology and an open-source masking tool are presented, in order to effectively augment existing face datasets to train masked face recognition algorithms. Also, the authors create a real-world masked face database (MFR2) for testing that contains masked face images from celebrities and politicians. As result, they report a considerable increase in the accuracy of the existing FaceNet model for both masked and unmasked faces, being able to extend out to real life masked faces.

On the other hand, some works have focused on enhancing the recognition performance of masked faces. In [28] a Support Vector Machine classifier is trained using the feature vector embeddings provided by FaceNet model on a collected small database for this purpose. The authors claim 99% of accuracy but the database and evaluation protocols used, are not provided and detailed. An alternative solution based on recovering unmasked faces for feature extraction was proposed in [19]. For this, a de-occlusion distillation framework is introduced, where first, appearance information is recovered by using a generative inpainting network and then, rich structural knowledge is transferred from a high-performance pretrained general recognizer in a teacher - student model. The method presented in [37], is based on the state-of-the-art ArcFace work [10] to extract deep features from the detected and normalized face images, which are then combined with LBP features extracted from eyes and eyebrows. Also, ArcFace network is used in [26], with several modifications for the backbone and the loss function. The network, based on ResNet-50, is modified to output the probability that a face is wearing a mask. In addition, the ArcFace loss is combined with a mask-usage classification loss to train mask robust facial feature embedding.

In the last year, different challenges and studies have been conducted in order to benchmark the performance of masked face recognition methods. The behavior of three face recognition algorithms in the presence of masked face probes is evaluated in [7]. The authors collect their own database for the evaluation and show how two of the top-performing face recognition deep models (ArcFace and SphereFace) and a COTS algorithm (from Neurotechnology) are affected in the presence of masks. The National Institute of Standards and Technology (NIST) reports the performance of a large number of face recognition algorithms on faces occluded by face masks being run under the Ongoing Face Recognition Vendor Test (FRVT) [29]. This study evidence that error rates for unmasked versus masked faces, have been decreasing across algorithms development after the pandemic. However, some of the algorithms that are quite competitive with unmasked faces still fail to authenticate between 10% to 40% of masked images. Although, the results evidence that a number of developers have adapted their algorithms to support masked face recognition, particular design details of the tested algorithms are not provided. Moreover, the dataset used in the NIST study is not publicly available and it contains synthetic masked images from controlled scenarios, thus real masked images in unconstrained scenarios were not considered.

Recently, the IJCB Masked Face Recognition Competition 2021 (IJCB-MFR-2021) [4] evaluates the solutions submitted by 10 teams. The database used in the competition represents a collaborative face verification scenario, but only 47 subjects wearing real face masks were considered. Moreover, most of the submitted solutions, especially the top-performing ones, are based on heavier ResNet architectures with a compactness between 23 and 108 millions of parameters. Another evaluation was conducted in Face Biometrics under COVID Workshop and Masked Face Recognition Challenge in ICCV 2021 [8]. A large number of recent solutions were evaluated on a dataset containing 6,964 masked facial images and 13,928 non-masked facial images. In this case, ResNet is also used as baseline model and the details of the submitted solutions are not provided. Unfortunately, the test data will not be released to the public but we are able to submit our proposal in order to be evaluated in this large dataset. In general, although some restrictions are imposed to the solutions submitted to these competitions, lightweight architectures are not considered.

B. PERIOCULAR FACE RECOGNITION

Under the situation caused by the COVID-19 pandemic, periocular recognition has reached direct relevance. Periocular region refers to the facial area in the immediate vicinity of the eyes [16]. Although there are no specific guidelines for the size and bounds of the periocular region, some studies suggest that considering the eyelids, eyelashes, eyebrow, tear duct, eye shape, and the surrounding skin can result in higher recognition rates [30].

Early periocular recognition approaches used monocular information, separating the left region from the right region and performing the matching individually. Park et al. [30] were the first to study the feasibility of using the periocular region as a biometric trait and evaluate its performance using different matchers based on global and local handcrafted feature extractors. The authors also examine the effectiveness of the periocular region for non-ideal scenarios and suggest including eyebrows and using neutral facial expression, as well as combining the results of matching the left and the right sides of the periocular images for more accurate recognition. In [3], different methods using the left, the right and both eyes were evaluated for recognition. It was shown that in all cases an improvement between 3 and 5 percent was achieved when using both eyes instead of only one. Thus, most of the state-of-the art methods use the bi-ocular information, some of them analyzing left and right eyes separately and then combining the results [36], while others use both eyes within a single image [14].

With the emergence of deep learning approach, the focus of the researchers has been moved to learn robust representations by deep Convolutional Neural Networks (CNNs) for periocular recognition, achieving visible improvement in the performance of periocular biometric systems [17], [44], [45]. The semantics-assisted convolutional neural networks (SCNN) [44] was one of the first proposals that use deep learning-based representation for periocular images. By incorporating explicit semantic information (gender and side), it shows to offer better discriminating power with the usage of a relatively smaller number of training samples. In [45] the authors apply existing pre-trained architectures, proposed to classify generic objects, to the task of periocular recognition. The results obtained show that these networks are able to outperform reference periocular features. Similarly, seven different off-the-shelf deep learning based CNN using transfer learning approach were implemented in [17] to analyse the utility of periocular region in non-ideal scenarios. A new method for masked face recognition was proposed in [20] by integrating a cropping-based approach with the Convolutional Block Attention Module (CBAM) to focus on the regions around eyes.

On the other hand, some works propose feature fusion approach which combines handcrafted features (e.g. LBP and HOG) with features extracted using pretrained CNN models [18], [36]. Another hybrid model is introduced in [1] for ocular smartphone authentication (Selfie Biometrics). The proposal is a fusion of a stacked unsupervised convolution-based model with a stacked supervised convolution-based model, which is combined with Root SIFT. A recent selfie periocular verification method is presented in [35], which consists of a two-stage approach based on a CNN with pixel-shuffle, and a new loss function based on a sharpness metric, aiming at enhancing the periocular images with a super-resolution approach.

Although the significant and encouraging research progress gained by the aforementioned works to address the problem of recognizing faces wearing masks, the study of lightweight deep networks for this problem deserves further attention. Moreover, there is a lack in the evaluation and comparison of existing methods for the two different approaches reviewed in this section, under the same scenarios and conditions, which could be very helpful for establishing the most suitable way to deal with the problem of masked face recognition.

III. DATASETS

In this section, we present the datasets used for studying the masked face recognition problem. Due to the lack of publicly available large-scale datasets for training and testing, we generate face images with simulated masks from existing unmasked face databases. We test the trained models in some state-of-the-art masked datasets and we also collect a real masked dataset from subjects of our laboratory. To analyze the feasibility of periocular information, we construct periocular images from the same datasets used for the masked face recognition scenario.

A. SIMULATED MASKED FACE DATASETS

In order to create simulated masked face datasets, we use the open-source tool MaskTheFace [2] to convert existing face datasets into masked face datasets. It uses the face landmarks detector provided by Dlib library [15] to identify the face tilt and six key features of the face necessary for applying a mask. Based on the face tilt, the corresponding mask template selected from the library of masks, is then transformed based on the six key features to fit on the face. MaskTheFace provides five different mask types including cloth, surgical, N95, KN95 and gas, and supporting 24 existing patterns that can be applied to mask types above to create more variations.

For the purpose of training, we select CASIA-WebFace [42] which is a face dataset that contains 494,414 images of 10,575 identities, with an average of 15 images per identity varying in pose, age, ethnicity and illumination. From this dataset, we create a Simulated Masked (SM) CASIA-WebFace by augmenting it using the described MaskTheFace tool. Specifically, for each unmasked face image from every subject, we generate four masked images using cloth, gas, KN95 and surgical-green synthetic mask types. Thus, both the original unmasked image and the created synthetic masked images, compose the SM CASIA-WebFace dataset in order to make sure that the trained networks perform well on both the masked and unmasked images. Figure 1 shows some examples of the synthetic masked face images obtained

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2021.3135255, IEEE Access

for different unmasked subjects from the CASIA-WebFace dataset.

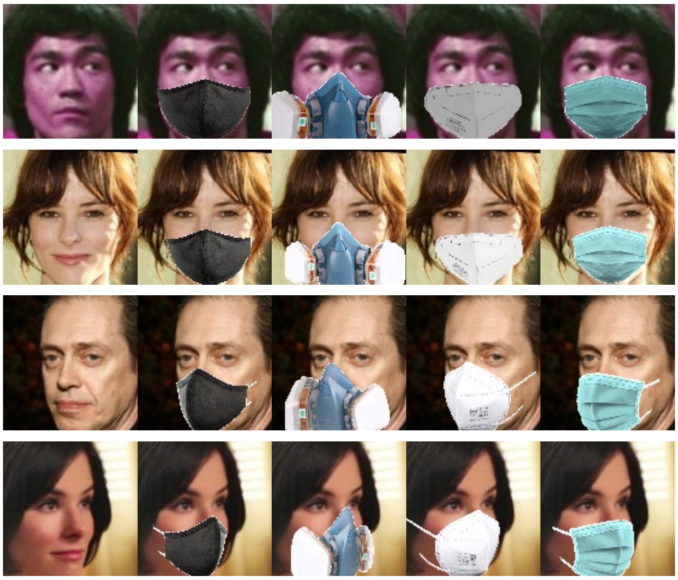

FIGURE 1: Examples of training face images from Simulated Masked CASIA-WebFace dataset.

In the case of testing, we select different face benchmarks including LFW [12], AgeDB-30 [27] and CALFW [46] to generate simulated masked datasets. For this, we use the MaskTheFace tool with one randomly selected mask applied to each image.

Labeled Faces in the Wild (LFW) [12] is a standard face recognition benchmark that contains 13,233 web-collected images from 5,749 different identities, with large variations in pose, expression and illuminations. The AgeDB [27] is an in-the-wild dataset with large variations in pose, expression, illuminations, and age. It contains 16,488 images of 568 distinct subjects. The average age range for each subject is 50.3 years. There are four groups of test data with different year gaps (5, 10, 20 and 30 years, respectively) for age-invariant face verification. In this paper, we only use the most challenging subset, AgeDB-30, to report the performance. Cross-Age LFW (CALFW) [46] is a recently introduced dataset that shows higher age variations, with the same identities from LFW database. These three databases define an evaluation protocol based on 6,000 face pairs matching, which are divided into ten subsets, each having 300 positive pairs and 300 negative pairs. To analyze the performance of trained networks, we compute the verification accuracy (Acc) and the Equal Error Rate (EER) metrics on the established 6,000 face pairs of each database.

B. REAL MASKED FACE DATASETS

Aiming at assessing the performance of the trained models on real masked face images, we create a small face database from persons of our laboratory wearing real masks. This database (RMFR-CEN) is an initial version and further data collection efforts are ongoing. The data tries to simulate a collaborative, yet varying, scenario where the mask, illumination, pose and background can change on each of the participants. In total, we collected 395 images from 100 identities. Each identity has on average of 4 images with both masked and unmasked faces. The dataset is processed in terms of face alignment and image dimensions. As result, each image has a dimension of (112 × 112 × 3). Figure 2 shows some examples of the images collected from our laboratory.

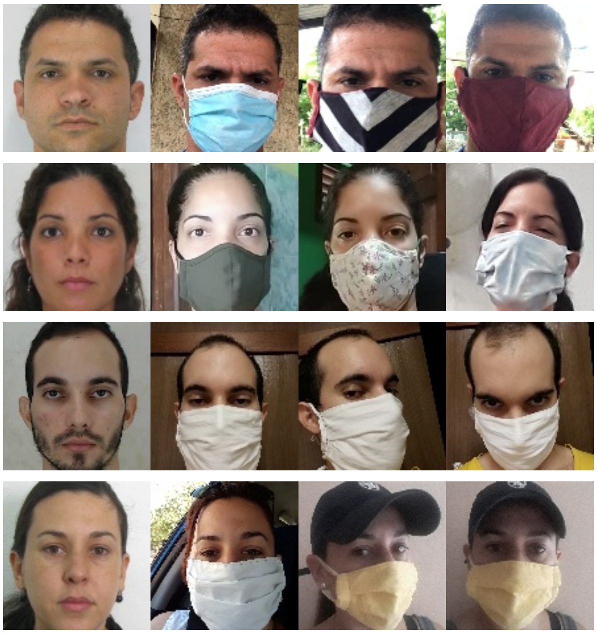

FIGURE 2: Examples of collected face images from RMFR-CEN dataset.

For performance evaluation, we design both face verification and identification protocols. The verification protocol specifies 469 positive pairs and 10,000 negative pairs composed of one masked face and one unmasked face. For performance measurement, each pair is evaluated by computing a matching similarity score, and the paired True Acceptance Rate (TAR) at different False Acceptance Rates (FAR), the Equal Error Rate (EER) and the Area Under ROC (AUC) are used as evaluation metrics. In the case of face identification, we construct the evaluation setup for the closed-set scenario, where for each subject, we use the most frontal unmasked image as the gallery, while the masked ones are used as probes. To report the identification performance, we select the Cumulative Matching Characteristic (CMC) [32] and the mean Average Precision (mAP) measures.

In order to enlarge the evaluation and compare the trained lightweight masked models with some existing masked face recognition solutions, we use the Masked faces in real-world for face recognition (MFR2) [2], the test set of the InsightFace-Track in Masked Face Recognition Challenge of ICCV 2021 [8] and AR Face [23] datasets.

MFR2 is a small dataset with 53 identities of celebrities and politicians with a total of 269 images collected from the internet. Each identity has on average 5 images, including both masked and unmasked faces. In Figure 3, we show some sample images from the MFR2 dataset. For the network performance evaluation, a total of 848 image pairs (424 positive pairs, and 424 negative pairs) are defined and Max Accuracy and TPR@FAR=0.2% metrics are used to report the verification performance.

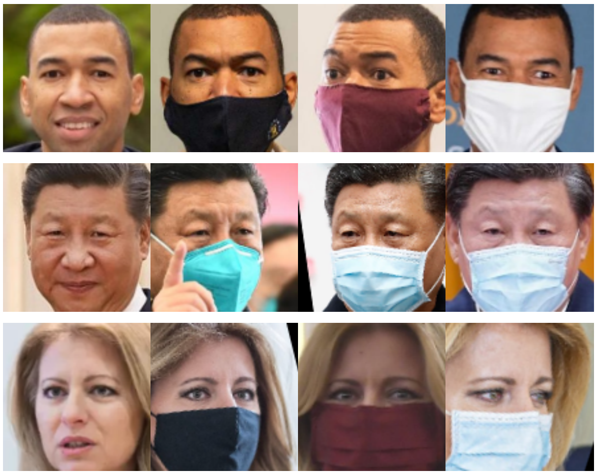

FIGURE 3: Examples of face images from MFR2 dataset.

The Masked Test Set of the InsightFace Track of ICCV 2021 [8] is a private dataset which contains 6,964 masked facial images and 13,928 non-masked facial images of 6,964 identities. In total, there are 13,928 positive pairs and 96,983,824 negative pairs for the verification evaluation. Unlike existing face recognition test sets, this dataset is not collected from celebrities, thus the identity-overlapping problem is naturally avoided. As evaluation metric 1:1 face verification is used and the results are reported in terms of True Positive Rate (TPR) @ False Positive Rate (FPR) = 1e-4. In Figure 4 we show some examples images from this dataset.

The AR Face Database [23] contains around 4,000 images from 126 subjects captured on two different sessions. Each person has up to 13 images per session with different expressions, illuminations and occlusions. The occlusions included in the dataset are the presence of sunglasses and scarves. Although this dataset is not a masked face dataset, it has been used for evaluating masked face solutions [19], [34] since the scarf occlusions cover more or less the same region than a face mask (See Figure 5). Thus, we have decided to use the Scarf subset of this database in the evaluation in order to be able to compare with state-of-the-art methods. Following previous protocols, we randomly select 100 subjects (50 males and 50 females) and conduct identification experiments by using one neutral image per subject (the first image in the first session) to conform the gallery.

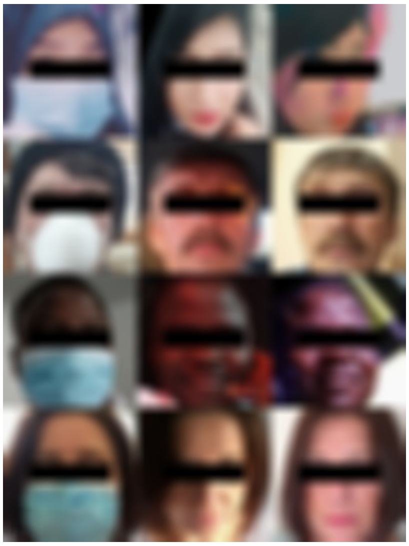

FIGURE 4: Examples images from the Masked Test Set of the InsightFace Track of ICCV 2021.

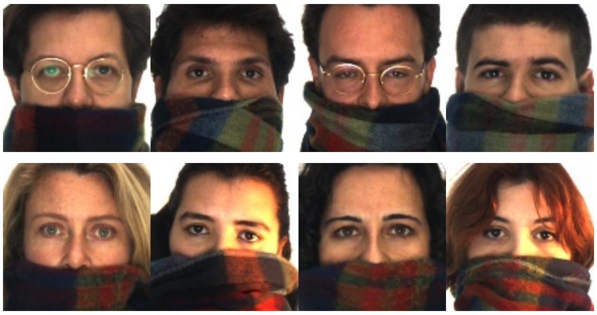

FIGURE 5: Examples of scarf face images from AR dataset.

C. PERIOCULAR DATASETS

In the case of periocular face datasets, we use the bi-ocular information including, in a single image, both eyes and considering the eyelids, eyelashes, eyebrow, tear duct, eye shape and the surrounding skin. For obtaining this periocular region, we crop face images based on the algorithm used in [18] for extracting the region of interest. This algorithm considers the canthus points as reference points, which were detected automatically through Dlib landmarks detector. Finally, the obtained periocular regions are geometrically normalized.

For a fair comparison between periocular-based lightweight models against mask-based lightweight models, we employed the same datasets and evaluation metrics. Thus, we use Periocular CASIA-WebFace for training, while Periocular LFW, AgeDB-30, CALFW y RMFR-CEN are used for testing. Figure 6 shows some examples of the training images from Periocular CASIA-WebFace. In addition, to enlarge our study, we compare the obtained periocular lightweight models with the state-of-the-art methods reported on the periocular images from the AR Face database.

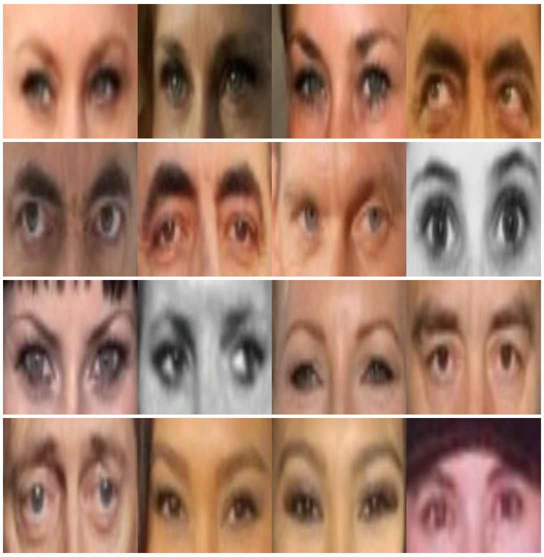

FIGURE 6: Examples of training face images from Periocular CASIA-WebFace dataset.

IV. LIGHTWEIGHT FACE RECOGNITION MODELS

In the last years, developing very efficient and lightweight face recognition networks has become an active research topic in order to make deep CNNs feasible on real-time applications or resource-limited devices. Existing lightweight models have shown to be able to perform very similar to larger and heavier deep models in different face recognition scenarios. In this study, we select three state-of-the-art lightweight CNN face models that were the top-performing in [24]: VarGFaceNet [41], MobileFaceNet [5] and ShuffleFaceNet [22]

VarGFaceNet [41] consists of an efficient variable group convolutional network based on VarGNet for lightweight face recognition. Different from the blocks in VarGNet, it adds squeeze and excitation (SE) block and employes PReLU activation function instead of ReLU to increase the discriminative ability of their blocks. Moreover, VarGFaceNet removes the downsample process at the start of network to preserve more information and applies variable group convolution after last convolution to shrink the feature tensor to $1\times1\times512$ before FC layer. Moreover, 3×3 Convolution with stride 1 is used at the start of network instead of 3×3 Convolution with stride 2 as in VarGNet, which reserves the discriminative ability in lightweight networks.

MobileFaceNet [5] and ShuffleFaceNet [22] have shown competitive performance with respect to high-accurate very deep face models on several benchmarks for unconstrained face recognition. The major contribution of these networks lie in the use of a Global Depth-wise Convolution (GDC) layer instead of a Global Average Pooling (GAP) layer in order to obtain a more discriminative face representation; and Parametric Rectified Linear Unit (PReLU) as non-linear activation function due to its accuracy improvement over the Rectified Linear Unit (ReLU) function.

Specifically, MobileFaceNet [5] uses the residual bottlenecks proposed in MobileNetV2 as their main building blocks, while ShuffleFaceNet [22] is based on the extremely efficient network ShuffeNetV2, where the building blocks in stages 2-4 consist of DenseNet blocks and the number of channels in each block is scaled to generate four networks of different complexities, denoted as 0.5×, 1×, 1.5× and 2×. Taking into account the results obtained in [22] where ShuffeFaceNet 1.5× presented the best trade-off between speed and accuracy, we will use this model and we will refer to it as ShuffeFaceNet in the remaining of our work. In addition, both lightweight networks adopt a fast downsampling strategy at the beginning of the networks, an early dimension-reduction strategy at the last several convolutional layers, and a linear 1×1 convolution layer following a linear global depthwise convolution layer as the feature output layer.

A. IMPLEMENTATION DETAILS

The lightweight face CNN networks, pretrained on the cleaned MS1M dataset [11], are independently fine-tuned on masked and periocular images created from the CASIA-WebFace dataset. For all the models, random horizontal flip is used as augmentation strategy. We adopt Stochastic Gradient Descent (SGD) optimizer with the batch size of 128/256/512 due to limited GPU memory, and the models fine-tuning is carried out on two Nvidia GeForce GTX 1080Ti (11GB) GPUs. The learning rate is initialized to 0.1 and decreased by a factor of 10 periodically at 100K, 140K, 160K iterations. The total iteration step is set as 200K. The momentum parameter is set to 0.9 and weight decay at 5e-4. The parameter initialization for convolution is Xavier with random sampling from a Gaussian normal distribution. For VarGFaceNet, MobileFaceNet and ShuffeFaceNet, we use ArcFace loss function with an angular margin $m = 0.5$, that turned out to be the best as it was specified in [24]. For all face models, we directly take the embedding feature after the last convolutional layer as face representation, and use the cosine similarity to obtain the matching scores.

For data preprocessing, RetinaFace detector [9] is applied to detect all faces and landmark points, which are used to align and crop each face into a template with the size of 112×112, where each pixel (ranged between [0; 255]) in RGB images is then normalized into [-1; 1] by subtracting the mean pixel value, i.e. 127.5, and divided by 128.

V. EXPERIMENTAL EVALUATION

In this section, we assess the effectiveness of trained ShuffleFaceNet, MobileFaceNet and VarGFaceNet on both masked and periocular face recognition datasets and compare them with state-of-the-art solutions. Moreover, we analyze the effect of using the lightweight masked face models for face recognition in unmasked benchmarks. In addition, we analyze the computational efficiency of these lightweight models compared with some state-of-the-art deep face models used for the problem of masked face recognition.

A. MASKED FACE RECOGNITION

In this scenario, all lightweight deep models that were trained on the Simulated Masked CASIA-WebFace dataset are tested on several datasets for both, simulated masked and real face recognition. To baseline the performance, we compare these models with their original version without fine-tuning them on masked face images.

1) Results on simulated masked datasets

In order to asses the performance of recognizing faces with and without the masks on, we first conduct experiments on Simulated Masked datasets: LFW, AgeDB-30 and CALFW. We evaluate two different configurations in order to draw effective matching comparisons: a) we carried out the matching of face pairs with the simulated masked (masked vs. masked) and b) we test pairs composed by one masked face and one unmasked face (masked vs. unmasked). In all cases we evaluate the original models and their fine-tuned versions.

In Table 1 and Table 2, we present the face verification results obtained for the two matching configurations, in terms of Equal Error Rate (EER) and Accuracy (Acc).

TABLE 1: Face verification performance (%) on the Simulated Masked LFW, AgeDB-30 and CALFW datasets by matching masked face pairs.

Method	Masked vs. Masked					
	LFW		AgeDB-30		CALFW	
	Acc	EER	Acc	EER	Acc	EER
VarGFaceNet [41]	96.4	3.8	85.9	13.8	85.9	14.1
ShuffleFaceNet [22]	96.9	3.4	86.3	13.6	85.5	15.1
MobileFaceNet [5]	97.1	3.2	87.8	12.3	86.8	13.8
VarGFaceNet-Mask	97.8	2.4	88.9	11.3	87.0	13.5
ShuffleFaceNet-Mask	98.0	2.1	89.8	10.0	88.7	11.8
MobileFaceNet-Mask	**98.5**	**1.6**	**91.6**	**8.7**	**89.9**	**10.8**

TABLE 2: Face verification performance (%) on the LFW, AgeDB-30 and CALFW datasets by matching masked faces versus unmasked faces.

Method	Masked vs. Unmasked					
	LFW		AgeDB-30		CALFW	
	Acc	EER	Acc	EER	Acc	EER
VarGFaceNet [41]	96.9	3.0	89.2	10.8	88.5	11.6
ShuffleFaceNet [22]	97.2	3.0	88.9	11.0	87.9	12.4
MobileFaceNet [5]	97.4	2.5	90.0	9.9	89.5	11.5
VarGFaceNet-Mask	98.3	1.9	89.4	10.7	87.7	13.2
ShuffleFaceNet-Mask	98.4	1.8	90.9	9.1	89.8	10.7
MobileFaceNet-Mask	**98.8**	**1.3**	**92.3**	**7.7**	**90.3**	**10.0**

It can be observed in the tables that for the three datasets, all the lightweight models fine-tuned with the masked images enhance the verification performance of the models that has not been trained with masked facial images. We can see that the greater improvements are obtained in front masked faces with age variations. Among the models, MobileFaceNet-Mask achieves the best results in the three databases. If we compare Table 1 against Table 2, it can be seen that better results are obtained when at least one of the images is unmasked. This is a desire property for real applications where usually the enrolled images are in normal condition (unmasked).

2) Results on RMFR-CEN dataset

In order to study the effectiveness of face models trained with simulated masks on real-world masked faces, we use the dataset collected in our laboratory, RMFR-CEN dataset. We follow the verification and identification protocols defined for this dataset on Section III-B, that consider comparisons of masked face images against unmasked images. The obtained results are presented in Table 3 and Table 4, respectively.

TABLE 3: Face verification (%) results on RMFR-CEN dataset.

Method	TAR%@FAR			EER	AUC
	30%	10%	1%		
VarGFaceNet [41]	86.7	75.7	49.7	19.2	89.5
ShuffleFaceNet [22]	87.4	74.8	52.7	17.5	89.6
MobileFaceNet [5]	90.0	81.0	63.3	14.7	92.5
VarGFaceNet-Mask	**94.9**	**87.2**	64.4	**11.3**	**94.9**
ShuffleFaceNet-Mask	91.9	82.3	70.6	15.1	93.0
MobileFaceNet-Mask	93.8	85.7	**72.9**	12.8	94.5

TABLE 4: Closed-face identification (%) results on RMFR-CEN dataset.

Method	Rank-1	Rank-10	Rank-20	mAP
ShuffleFaceNet [22]	59.4	82.7	87.6	63.4
VarGFaceNet [41]	60.9	80.7	87.1	64.2
MobileFaceNet [5]	68.8	84.7	88.6	71.7
ShuffleFaceNet-Mask	68.8	86.1	92.1	72.0
VarGFaceNet-Mask	68.8	88.1	**93.1**	72.0
MobileFaceNet-Mask	**77.2**	**91.1**	92.1	**79.6**

It can be seen that also in the real masked images, for both verification and identification protocols, the models fine-tuned with synthetic masks are able to enhance considerably the performance of originals face models. Specifically, all mask-based models are able to increase in at least 9% the TAR@FAR=1% and the Rank-1 results with respect to unmasked models. Among the masked models, VarGFaceNet-Mask obtains better verification results, while MobileFaceNet-Mask achieves the highest identification performance, especially at Rank-1. As we can appreciate, when we test the lightweight masked models in face images wearing real masks, the improvements over the original models are more remarkable. However, the overall performance, is not as good as when we tested these models in synthetic datasets, which still leaves a large margin of improvement.

3) Comparison with state-of-the-art

In order to compare the lightweight masked face models with existing solutions for the masked face recognition problem, we assess their performance on the MFR2, the InsightFace-Track in Masked Face Recognition Challenge of ICCV 2021 and the AR Face datasets.

Table 5 presents comparative results achieved on the MFR2 dataset by the lightweight face models and FaceNet model with and without fine-tuning. The results are reported based on the evaluation criteria Max Accuracy and TPR@FAR=0.2%, described in [2]. As can be seen, all the lightweight-masked face models achieved higher verification performance than FaceNet-FT, being the MobileFaceNet-Mask the best one in terms of TPR@FAR=0.2%. In addition, we can observe that fine-tuning face recognition models with masked face images (real or simulated) improves the verification performance.

TABLE 5: Face verification (%) results on MFR2 dataset.

Method	TPR@FAR=0.2%	Max Accuracy
FaceNet [2]	48.9	90.3
ShuffleFaceNet [22]	75.7	95.0
VarGFaceNet [41]	88.2	95.5
MobileFaceNet [5]	90.3	94.3
FaceNet-FT [2]	82.8	96.0
ShuffleFaceNet-Mask	86.3	96.1
VarGFaceNet-Mask	90.3	96.1
MobileFaceNet-Mask	**91.3**	**96.1**

In Table 6, we present the verification performance on the InsightFace-Track in Masked Face Recognition Challenge of ICCV 2021, where TPR is measured on mask-to-non-mask 1:1 protocol, with FAR less than 0.01%(1e-4). Also, further details are presented such as the training dataset, as well as the size and the inference time of the models. In all cases ArcFace loss function was used. We compare our lightweight masked face models with the provided baseline solutions based on the ResNet architecture. It is important to note that, we do not participate in the competition, we only asses the performance of our models on the test set. Thus, the comparison with the reported baseline results is not fair enough since most of them employ different and bigger training sets such as Glint360K and MS1MV3, which contributes to the differences in the performance. It can be seen that, all ResNet baseline models significantly increase their verification accuracy by using Glint360K dataset. For example, the R100 backbone trained on the Glint360K dataset, outperforms in more than 40% the results obtained by using the CASIA dataset, which is the one we used. Although our lightweight masked face models were not trained with the datasets provided by the competition (MS1M, Glint360K), they considerably improve the verification performance of the R100 trained on the CASIA. In particular, the MobileFaceNet-Mask, the best performing one, surpass the R100 trained on CASIA in more than 28%. Moreover, we can observe that MobileFaceNet-Mask is capable to obtain better results than both versions of R18 trained with more powerful training sets. In the future, we plan to employ some of these datasets for fine-tuning our masked face models in order to increase their accuracy. On the other hand, we can appreciate that all lightweight models present the smallest model sizes with very low inference times.

In Table 7, recognition accuracy at Rank-1 is reported on the Scarf subset of the AR Face database. The performance of lightweight face models fine-tuned with mask images is compared with those reported by state-of-the-art methods proposed for face recognition under occlusions. As we can see, the three masked face models are able to achieve perfect recognition rates under this kind of occlusion, which is somehow similar to the one caused by the presence of masks. These results outperform specific methods devoted to handling occlusions such as ArcFace-FT [34] and PDifferentialSiamese [34] that have been evaluated on this database.

TABLE 6: Verification performance (%) on the InsightFace-Track in Masked Face Recognition Challenge of ICCV 2021. TPR denotes the TPR@FAR=1e-4 measured on mask-to-non-mask 1:1 protocol. Inference time is evaluated on Tesla V100 GPU using onnxruntime-gpu==1.6.

Backbone	Dataset	TPR	Size(MB)	Time(ms)
R18	MS1MV3	47.85	91.66	**1.86**
R18	Glint360K	53.32	91.66	2.01
R34	MS1MV3	58.72	130.25	3.05
R34	Glint360K	65.11	130.25	3.04
R50	MS1MV3	63.85	166.31	4.26
R50	Glint360K	70.23	166.31	4.34
R100	CASIA	26.62	248.90	7.07
R100	MS1MV2	65.77	248.90	7.03
R100	MS1MV3	69.09	248.59	7.03
R100	Glint360K	**75.57**	248.59	7.04
VarGFaceNet-Mask	CASIA	39.24	19.18	2.52
ShuffleFaceNet-Mask	CASIA	38.15	9.31	2.27
MobileFaceNet-Mask	CASIA	55.07	**7.87**	2.69

TABLE 7: Rank-1 face identification accuracy (%) on AR Face dataset with natural scarf occlusions.

Method	Rank-1
RPSM [39]	90.2
Stringface [6]	92.0
LMA [25]	93.7
DeOccDistillation [19]	94.1
ArcFace-FT [34]	96.4
PDifferentialSiamese [34]	98.3
ShuffleFaceNet-Mask	**100**
MobileFaceNet-Mask	**100**
VarGFaceNet-Mask	**100**

B. PERIOCULAR FACE RECOGNITION

In this section, we test the lightweight face models that were trained with periocular CASIA-WebFace dataset on the periocular datasets obtained by cropping the images of the LFW, AgeDB-30, CALFW and RMFR-CEN databases. To analyze and evaluate the effectiveness of using periocular region, we compare their performance w.r.t. the results obtained by finetuning with masked face images. Moreover, we compare the periocular models with some state-of-the-art periocular algorithms.

1) Results on periocular datasets

Table 8 shows the verification performance obtained by the lightweight face models trained with periocular region in the Periocular LFW, AgeDB-30 and CALFW datasets. As we can observe, among the tested lightweight periocular-based

TABLE 8: Face verification (%) results on Periocular LFW, AgeDB-30 and CALFW datasets.

Method	LFW		AgeDB-30		CALFW	
	Acc	EER	Acc	EER	Acc	EER
VarGFaceNet-Periocular	97.7	2.3	87.4	12.4	86.6	14.1
ShuffleFaceNet-Periocular	97.3	2.9	87.9	11.9	88.7	11.7
MobileFaceNet-Periocular	**98.1**	**2.1**	**90.2**	**9.9**	**89.6**	**11.0**

models, the MobileFaceNet-Periocular achieves the highest verification results for the three benchmarks. However, if we compare the performance of these models with the results obtained by the lightweight masked models on Tables 1 and 2, it can be appreciated that by using the periocular information we are not able to improve the results achieved by the models fine-tuned on masked images, especially in front of age variations.

In Table 9 and Table 10, we present the verification and identification performance obtained in Periocular RMFR-CEN dataset, respectively. Specifically, if we compare the performance of the periocular models w.r.t. the masked models (in Tables 3 and 4, respectively) we can observe that in the case of verification, for high FAR values (e.g. 30%), the degradation on the performance is bigger, while in the case of identification experiments, for a lower FAR=1% the results are more closer. Also in this case, MobileFaceNet achieves the best results and exhibits the greater differences between masked and periocular versions.

TABLE 9: Face verification (%) results on Periocular RMFR-CEN dataset.

Method	TAR@FAR			EER	AUC
	30%	10%	1%		
VarGFaceNet-Periocular	84.9	74.6	64.0	19.6	88.7
ShuffleFaceNet-Periocular	83.8	77.8	66.5	18.8	89.5
MobileFaceNet-Periocular	**88.7**	**82.1**	**72.5**	**15.4**	**92.2**

TABLE 10: Closed-face identification (%) on Periocular RMFR-CEN.

Method	Rank-1	Rank-10	Rank-20	mAP
ShuffleFaceNet-Periocular	68.3	88.1	94.1	71.3
VarGFaceNet-Periocular	68.1	84.2	88.6	71.4
MobileFaceNet-Periocular	**73.8**	**88.6**	**91.6**	**76.1**

2) Comparison with state-of-the-art

In order to compare the periocular lightweight models with state-of-the-art periocular algorithms, we follow the protocol used in [36] for the AR Face database, where a large number of periocular methods have been evaluated. The obtained identification accuracy for Rank-1 and Rank-5 are listed in Table 11. As we can observe, similar to masked-based model, the three periocular lightweight models reach 100% of identification, outperforming all the reported state-of-the-art methods.

TABLE 11: Face identification (%) performance and comparison with state-of-the-art methods on Periocular AR dataset using protocol defined in [36].

Method	Rank-1	Rank-5
AlexNet	93.6	96.8
VGG16	94.2	97.6
FaceNet	94.2	97.8
LCNN29	94.3	97.5
DeepIrisNet-B	94.4	97.2
DeepIrisNet-A	95.2	98.4
Multi-fusion CNN	96.1	98.7
OCLBCP dual-stream CNN	96.3	98.8
ShuffleFaceNet-Periocular	**100**	**100**
MobileFaceNet-Periocular	**100**	**100**
VarGFaceNet-Periocular	**100**	**100**

C. EFFECT OF USING MASKED FACE MODELS ON UNMASKED FACE DATASETS

In order to ask the question if face models trained on masked datasets can be used to recognize normal unmasked faces, we test the MobileFaceNet-Mask, ShuffleFaceNet-Mask and VarGFaceNet-Mask on well-established face recognition benchmarks. We compare them with their unmasked face models, as well as with state-of-the-art methods reported on these benchmarks. Specifically, we selected face databases which cover different covariates such as unconstrained scenario (LFW), age (AgeDB-30, CALFW), pose (CPLFW) and large-scale (IJB-B).

TABLE 12: Verification accuracy (%) and comparison with state-of-the-art methods reported in [24] in original LFW and AgeDB-30 databases.

Method	LFW	AgeDB-30
VGG-Face	98.9	85.1
CenterLoss	99.3	90.7
Marginal Loss	99.5	95.7
Seesaw-shuffleFaceNet	99.6	96.9
VarGFaceNet	99.6	97.1
ShuffleFaceNet	99.7	97.3
MobileFaceNet	99.7	97.6
ResNet100-ArcFace	**99.8**	**98.2**
VarGFaceNet-Mask	98.9	90.9
ShuffleFaceNet-Mask	99.0	92.8
MobileFaceNet-Mask	99.2	93.9

Table 12 presents the verification results on LFW and AgeDB-30 datasets, while Table 13 shows those obtained on CALFW and CPLFW benchmarks. We can see that, for general unconstrained images in the LFW dataset, the lightweight masked models obtain very close results to their unmasked versions and even to the state-of-the art. However, for age and pose variations (Table 13) , the impact on the performance is greater. Among the masked models, the MobileFaceNet-Mask is the less affected by the different covariates.

In order to evaluate the masked models on large-scale datasets, we use the Janus Benchmark-B (IJB-B) dataset [40] which consists of 1,845 subjects with 21,798 still images and 55,026 frames from 7,011 videos. Specifically, we follow

TABLE 13: Face verification accuracy (%) and comparison with state-of-the-art methods reported in [24] in original CALFW and CPLFW databases.

Method	CALFW	CPLFW
Human-Individual	82.3	81.2
Human-Fusion	86.5	85.2
CenterLoss	85.5	77.5
SphereFace	90.3	81.4
VGG-Face2	90.6	84.0
ResNet100-ArcFace	**95.9**	91.2
DDL	-	**93.4**
MobileFaceNet	95.1	87.7
VarGFaceNet	94.8	87.7
ShuffleFaceNet	94.7	86.9
MobileFaceNet-Mask	91.9	83.2
ShuffleFaceNet-Mask	91.3	81.1
VarGFaceNet-Mask	89.0	81.2

TABLE 14: Verification TAR (%) at different FARs and 1:N (mixed media) Identification on the IJB-B database.

Method	TAR@FAR		Identification		
	1e-4	1e-3	IET@ FPIR=0.1%	Rank-1	Rank-5
VGG-Face2	80.0	88.7	83.9	90.1	94.5
RKD	89.6	94.7	87.6	93.4	96.5
SP	89.8	94.9	88.0	93.8	96.6
R50-ArcFace	89.9	94.5	88.2	93.6	96.5
DDL	90.7	95.2	89.5	93.9	**96.6**
MobileFaceNet	**92.8**	**95.6**	**92.2**	**94.0**	96.5
VarGFaceNet	92.9	95.6	92.1	94.0	96.5
ShuffleFaceNet	92.3	95.2	91.4	93.6	96.2
MobileFaceNet-Mask	87.3	92.7	85.4	91.6	95.0
VarGFaceNet-Mask	78.6	88.4	74.7	87.3	92.7
ShuffleFaceNet-Mask	79.4	90.2	73.5	88.3	92.8

the evaluation protocols 1:1 verification and 1:N (mixed media) identification including both closed and open-set protocols. As performance metrics, the paired TAR@FAR are reported for the 1:1 verification protocol, while Cumulative Match Characteristic (CMC) and Identification Error Trade-off (IET) curves are reported for the 1:N closed-set and 1:N open-set identification protocols, respectively.

Table 14 presents the verification and the identification performance of masked face models and state-of-the-art methods that are evaluated in [24]. For the verification task, IJB-B provides 12,115 templates with 10,270 genuine matches and 8M impostor matches. In the case of verification, we compare the TAR at FAR values of 0.01% (1e-4) and 0.1% (1e-3), while for 1:N identification, we compare the Rank-1 and Rank-5 accuracy for closed-set protocol and IET@FPIR=0.1% for open-set protocol [40]. As we can appreciate, for both verification and identification, masked-based models also achieve worse results than their unmasked versions. The greater difference in performance are shown for VarGFaceNet-Mask and ShuffleFaceNet-Mask, being MobileFaceNet-Mask which presents the lowest drops. Derived from the results presented in Table 14, in unmasked environments it is still a better option to use models learned with images without masks. Unsurprisingly, current models are not able to fully generalize over masked and unmasked faces. Thus, for applications where both, masked and unmasked faces can be present, an alternative to consider is to introduce a previous stage where the face masks are detected.

D. NETWORK VISUALIZATION

In order to qualitatively analyze and interpret the experimental results, Gradient-weighted Class Activation Mapping (Grad-CAM) [33] technique is adopted to localize the discriminative areas. Grad-CAM uses the gradient information flowing into the last convolutional layer of the CNN to assign importance values to each neuron for a particular decision of interest, without any modification in the network architecture.

The generated Grad-CAM maps of the lightweight masked and periocular models are shown in Figure 7. In addition, we included the Grad-CAMs of R100-ArcFace model. In the figure, the first two columns correspond to the visualizations over masked face images from the Simulated Masked CASIA-WebFace, and the last two columns over periocular images from the Periocular CASIA-WebFace. As we can see, the Grad-CAM maps greatly vary for different models. However, we can appreciate that all the masked models assign a very low weight to the mask region. In general, the regions activated in the features maps for MobileFaceNet models are bigger than those of the other models. On the other hand, if we compare the maps for the masked models against those of the periocular ones, we found that in the case of the masked models, they are more focused over the the area around the eyes.

E. EFFICIENCY ASSESSMENT

Table 15 presents the computational requirements for the lightweight face models used in this study, and the comparison with some state-of-the-art deep face architectures that have been employed to address the masked face recognition problem. Specifically, we compare the storage space (model size) in Megabytes (MB), the compactness (number of parameters) and the Floating Point Operations Per Second (FLOPs) of the models. We can observe that, lightweight

TABLE 15: Comparison of storage space (model size), compactness (number of parameters) and Float-Operating-Points (FLOPs) of the lightweight face architectures with state-of-the-art face recognition models.

Method	Model size(MB)	#Param.(M)	GFLOPs
AlexNet	244	61	729
FaceNet	95	7.5	500
VGG-Face	526	138	15
VGG-Face2	165	25.6	4.0
R100-ArcFace	249	65.2	24.2
LCNN29	125	12.6	3.9
MobileFaceNet	**8.2**	**2.0**	0.9
ShuffleFaceNet	10.5	2.6	**0.6**
VarGFaceNet	20.0	5.0	1.0

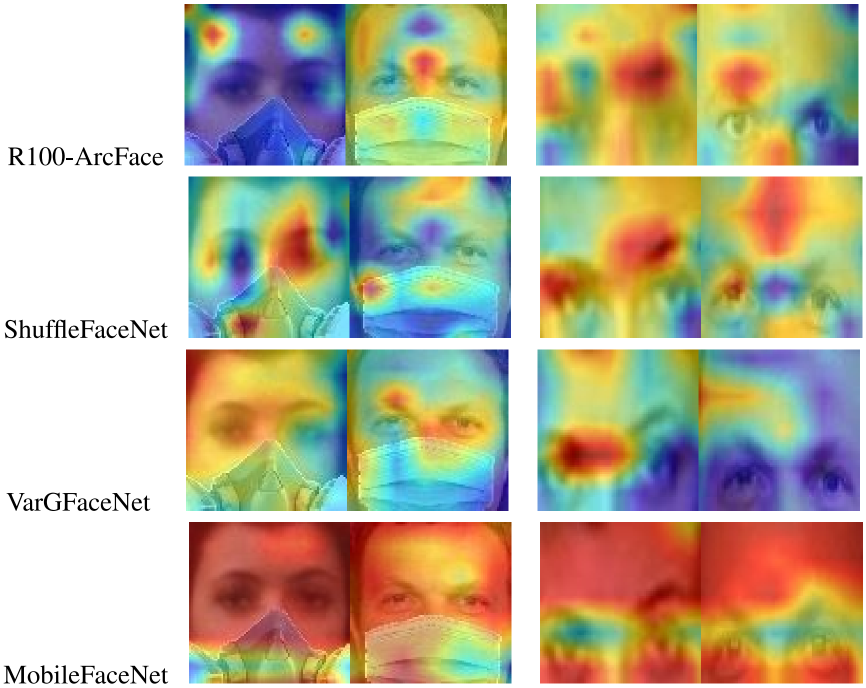

FIGURE 7: Grad-CAM visualizations of the fine-tuned lightweight face networks and the state-of-the-art R100-ArcFace model for face images from the Simulated Masked and the Periocular CASIA-WebFace training datasets. First two columns correspond to visualizations over masked face images and the last two ones over periocular images.

models improve remarkably the efficiency of the considered state-of-the-art models in all the requirements measured. The size of the biggest lightweight model (VarGFaceNet) is 10 times smaller than that of the well-known R100-ArcFace model, while the number of parameters is 13 times lower. The results indicate that the lightweight models have the best deployment capacity, which make potentially suitable and practical for using in embedding and low computational power devices.

VI. SUMMARY AND CONCLUSIONS

In this paper, we have presented a comprehensive study and evaluation of the performance of three state-of-the-art lightweight face models in order to address the effect of wearing masks on face recognition scenarios. To this end, two different approaches are investigated: on the one hand, the models are fine-tuned with several masked face images and on the other hand, the periocular information is considered.

Due to the lack of public datasets containing real masked face images, we created simulated masked datasets by placing synthetic masks over the face images from the CASIA-WebFace dataset for training, and from well-established face benchmarks for testing. Moreover, we test the trained models on some real masked dataset and also collect a database, named RMFR-CEN, of 100 subjects of our laboratory with real masked face images The proposed dataset is part of an ongoing effort to gather a larger scale database with realistic variations and will be available upon request. In order to compare the two considered approaches under the same conditions, periocular versions of these datasets is also constructed for training and evaluation.

From the experimental evaluation, our study pointed out the significant drop in the performance of the exiting face recognition solutions when considering masked face probes, especially in realistic scenarios. We found that by fine-tuning the models on masked faces, we are able to achieve better results than by using the periocular region. Moreover, we corroborate that models obtain a higher accuracy when matching masked vs. unmasked images is performed, which is an important aspect in the development of real applications. Compared with existing solutions for addressing the masked face recognition problem, which are based on more heavier deep networks, the considered lightweight models shown a very competitive performance. This indicates that utilizing a larger and deeper deep learning models does not necessarily and solely lead to higher recognition performance.

In addition, we observed that the masked-based models can recognize unmasked faces on general unconstrained scenarios. However, there is still a margin of improving the performance when there are more drastic appearance variations in the faces such as those caused by aging and larger variations in poses. The aforementioned conclusions open opportunities to propose new methods, algorithms, architectures, and/or loss functions that allow obtaining models able to generalize better in the presence of facial artifacts such masks, in order to provide existing face recognition systems with greater robustness to people wearing of such a necessary accessory in times of COVID-19. Also, it is a real necessity to detect masks as an additional functionality.

Regarding to the efficiency of the lightweight face architectures employed in this study, we assess to their computational requirements and compare them with some of the state-of-the-art methods that have been used in the literature for recognizing faces wearing masks. As results, we show that the lightweight models are potentially suitable for being employed in embedding and low computational power devices.

As future work, we plan to continuous collecting more real masked images from our laboratory in order to enrich the proposed RMFR-CEN database. Moreover, although masked-based models allow us to obtain higher recognition performance than periocular-based models, we think that combining both approaches could improve the performance of current masked face recognition solutions.

REFERENCES

[1] K. Ahuja, R. Islam, F. A. Barbhuiya, and K. Dey. Convolutional neural networks for ocular smartphone-based biometrics. Pattern Recognition Letters, 91:17–26, 2017.

[2] A. Anwar and A. Raychowdhury. Masked face recognition for secure authentication. arXiv preprint arXiv:2008.11104, pages 1–8, 2020.

[3] S. Bakshi, S. Kumari, R. Raman, and P. K. Sa. Evaluation of periocular over face biometric: A case study. Procedia engineering, 38:1628–1633, 2012.

[4] F. Boutros, N. Damer, J. N. Kolf, K. Raja, F. Kirchbuchner, R. Ramachandra, A. Kuijper, P. Fang, C. Zhang, F. Wang, et al. Mfr 2021: Masked face recognition competition. In 2021 IEEE International Joint Conference on Biometrics (IJCB), pages 1–10. IEEE, 2021.

[5] S. Chen, Y. Liu, X. Gao, and Z. Han. Mobilefacenets: Efficient cnns for accurate real-time face verification on mobile devices. In J. Zhou, Y. Wang, Z. Sun, Z. Jia, J. Feng, S. Shan, K. Ubul, and Z. Guo, editors, Biometric Recognition, pages 428–438, 2018.

[6] W. Chen and Y. Gao. Recognizing partially occluded faces from a single sample per class using string-based matching. In European Conference on Computer Vision, pages 496–509. Springer, 2010.

[7] N. Damer, J. H. Grebe, C. Chen, F. Boutros, F. Kirchbuchner, and A. Kuijper. The effect of wearing a mask on face recognition performance: an exploratory study. In 2020 International Conference of the Biometrics Special Interest Group (BIOSIG), pages 1–6. IEEE, 2020.

[8] J. Deng, J. Guo, X. An, Z. Zhu, and S. Zafeiriou. Masked face recognition challenge: The insightface track report. arXiv preprint arXiv:2108.08191, pages 1–8, 2021.

[9] J. Deng, J. Guo, E. Ververas, I. Kotsia, and S. Zafeiriou. Retinaface: Single-shot multi-level face localisation in the wild. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pages 5203–5212, 2020.

[10] J. Deng, J. Guo, N. Xue, and S. Zafeiriou. Arcface: Additive angular margin loss for deep face recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pages 4690–4699, 2019.

[11] Y. Guo, L. Zhang, Y. Hu, X. He, and J. Gao. Ms-celeb-1m: A dataset and benchmark for large-scale face recognition. In European conference on computer vision, pages 87–102. Springer, 2016.

[12] G. B. Huang, M. Ramesh, T. Berg, and E. Learned-miller. Labeled faces in the wild: A database for studying face recognition in unconstrained environments, 2007.

[13] Y. G. Jung, C. Y. Low, J. Park, and A. B. J. Teoh. Periocular recognition in the wild with generalized label smoothing regularization. IEEE Signal Processing Letters, 27:1455–1459, 2020.

[14] Y. G. Jung, J. Park, C. Y. Low, L. C. O. Tiong, and A. B. J. Teoh. Periocular in the wild embedding learning with cross-modal consistent knowledge distillation. arXiv preprint arXiv:2012.06746, pages 1–30, 2020.

[15] D. E. King. Dlib-ml: A machine learning toolkit. J. Mach. Learn. Res., 10:1755–1758, 2009.

[16] P. Kumari and K. Seeja. Periocular biometrics: A survey. Journal of King Saud University-Computer and Information Sciences, pages 1–12, 2019.

[17] P. Kumari and K. Seeja. Periocular biometrics for non-ideal images: with off-the-shelf deep cnn & transfer learning approach. Procedia Computer Science, 167:344–352, 2020.

[18] P. Kumari and K. Seeja. A novel periocular biometrics solution for authentication during covid-19 pandemic situation. Journal of Ambient Intelligence and Humanized Computing, pages 1–17, 2021.

[19] C. Li, S. Ge, D. Zhang, and J. Li. Look through masks: Towards masked face recognition with de-occlusion distillation. In Proceedings of the 28th ACM International Conference on Multimedia, pages 3016–3024, 2020.

[20] Y. Li, K. Guo, Y. Lu, and L. Liu. Cropping and attention based approach for masked face recognition. Applied Intelligence, 51(5):3012–3025, 2021.

[21] W. Liu, Y. Wen, Z. Yu, M. Li, B. Raj, and L. Song. Sphereface: Deep hypersphere embedding for face recognition. In Proceedings of the IEEE conference on computer vision and pattern recognition, pages 212–220, 2017.

[22] Y. Martindez-Diaz, L. S. Luevano, H. Mendez-Vazquez, M. Nicolas-Diaz, L. Chang, and M. Gonzalez-Mendoza. Shufflefacenet: A lightweight face architecture for efficient and highly-accurate face recognition. In The IEEE International Conference on Computer Vision (ICCV) Workshops, pages 1–8, Oct 2019.

[23] A. Martinez and R. Benavente. The AR face database. Tech. Rep. 24 CVC Technical Report, pages 1–8, 01 1998.

[24] Y. Martinez-Diaz, M. Nicolas-Diaz, H. Mendez-Vazquez, L. S. Luevano, L. Chang, M. Gonzalez-Mendoza, and L. E. Sucar. Benchmarking lightweight face architectures on specific face recognition scenarios. Artificial Intelligence Review, pages 1–44, 2021.

[25] N. McLaughlin, J. Ming, and D. Crookes. Largest matching areas for illumination and occlusion robust face recognition. IEEE transactions on cybernetics, 47(3):796–808, 2016.

[26] D. Montero, M. Nieto, P. Leskovsky, and N. Aginako. Boosting masked face recognition with multi-task arcface. arXiv preprint arXiv:2104.09874, pages 1–6, 2021.

[27] S. Moschoglou, A. Papaioannou, C. Sagonas, J. Deng, I. Kotsia, and S. Zafeiriou. Agedb: the first manually collected, in-the-wild age database. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, pages 51–59, 2017.

[28] I. Q. Mundial, M. S. Ul Hassan, M. I. Tiwana, W. S. Qureshi, and E. Alanazi. Towards facial recognition problem in covid-19 pandemic. In 2020 4rd International Conference on Electrical, Telecommunication and Computer Engineering (ELTICOM), pages 210–214, 2020.

[29] M. Ngan, P. Grother, and K. Hanaoka. Ongoing face recognition vendor test (frvt) part 6b: Face recognition accuracy with face masks using post-covid-19 algorithms, november 2020.

[30] U. Park, R. R. Jillela, A. Ross, and A. K. Jain. Periocular biometrics in the visible spectrum. IEEE Transactions on Information Forensics and Security, 6(1):96–106, 2010.

[31] U. Park, A. Ross, and A. K. Jain. Periocular biometrics in the visible spectrum: A feasibility study. In 2009 IEEE 3rd international conference on biometrics: theory, applications, and systems, pages 1–6. IEEE, 2009.

[32] P. J. Phillips, P. Grother, and R. Micheals. Evaluation methods in face recognition. In Handbook of Face Recognition, pages 551–574. 2011.

[33] R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, and D. Batra. Grad-cam: Visual explanations from deep networks via gradient-based localization. In 2017 IEEE International Conference on Computer Vision (ICCV), pages 618–626, 2017.

[34] 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, pages 773–782, 2019.

[35] J. Tapia, M. Gomez-Barrero, R. Lara, A. Valenzuela, and C. Busch. Selfie periocular verification using an efficient super-resolution approach. arXiv preprint arXiv:2102.08449, pages 1–13, 2021.

[36] L. Tiong, Y. Lee, and A. Teoh. Periocular recognition in the wild: Implementation of rgb-oclbcp dual-stream cnn. Applied Sciences (Switzerland), 9(13):1–17, 2019.

[37] H. N. Vu, M. H. Nguyen, and C. Pham. Masked face recognition with convolutional neural networks and local binary patterns. Applied Intelligence, pages 1–16, 2021.

[38] Z. Wang, G. Wang, B. Huang, Z. Xiong, Q. Hong, H. Wu, P. Yi, K. Jiang, N. Wang, Y. Pei, et al. Masked face recognition dataset and application. arXiv preprint arXiv:2003.09093, pages 1–3, 2020.

[39] R. Weng, J. Lu, and Y.-P. Tan. Robust point set matching for partial face recognition. IEEE transactions on image processing, 25(3):1163–1176, 2016.

[40] C. Whitelam, E. Taborsky, A. Blanton, B. Maze, J. Adams, T. Miller,

N. Kalka, A. K. Jain, J. A. Duncan, K. Allen, et al. Iarpa janus benchmark-b face dataset. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, pages 90–98, 2017.

[41] M. Yan, M. Zhao, Z. Xu, Q. Zhang, G. Wang, and Z. Su. Vargfacenet: An efficient variable group convolutional neural network for lightweight face recognition. In The IEEE International Conference on Computer Vision (ICCV) Workshops, pages 1–8, Oct 2019.

[42] D. Yi, Z. Lei, S. Liao, and S. Z. Li. Learning face representation from scratch. arXiv preprint arXiv:1411.7923, pages 1–9, 2014.

[43] D. Zeng, R. Veldhuis, and L. Spreeuwers. A survey of face recognition techniques under occlusion. IET Biometrics, pages 1–23, 2020.

[44] Z. Zhao and A. Kumar. Accurate periocular recognition under less constrained environment using semantics-assisted convolutional neural network. IEEE Transactions on Information Forensics and Security, 12(5):1017–1030, 2016.

[45] Z. Zhao and A. Kumar. Improving periocular recognition by explicit attention to critical regions in deep neural network. IEEE Transactions on Information Forensics and Security, 13(12):2937–2952, 2018.

[46] T. Zheng, W. Deng, and J. Hu. Cross-age lfw: A database for studying cross-age face recognition in unconstrained environments. arXiv preprint arXiv:1708.08197, pages 1–10, 2017.