IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received 31 May 2023, accepted 30 June 2023, date of publication 5 July 2023, date of current version 12 July 2023.

Digital Object Identifier 10.1109/ACCESS.2023.3292407

RESEARCH ARTICLE

Toward Generalizable Facial Presentation Attack Detection Based on the Analysis of Facial Regions

LAZARO JANIER GONZALEZ-SOLER[1], MARTA GOMEZ-BARRERO[2], (Member, IEEE), AND CHRISTOPH BUSCH[1], (Senior Member, IEEE)

[1]da/sec—Biometrics and Security Research Group, Hochschule Darmstadt, 64295 Darmstadt, Germany
[2]Hochschule Ansbach, 91522 Ansbach, Germany

Corresponding author: Lazaro Janier Gonzalez-Soler (lazaro-janier.gonzalez-soler@h-da.de)

This work was supported in part by the DFG-ANR RESPECT Project 406880674; and in part by the German Federal Ministry of Education and Research and the Hessian Ministry of Higher Education, Research, Science and the Arts within their joint support of the ATHENE National Research Center for Applied Cybersecurity.

ABSTRACT In the last decade, breakthroughs in the field of deep learning have led to the development of powerful presentation attack detection (PAD) algorithms which reported reliable performance across different realistic scenarios. Typically, most of these techniques analyse the full face to detect attack presentations (APs), ignoring that the attributes or artefacts produced in the fabrication of the attacks vary their location on the face depending on the presentation attack instruments (PAI) species, subject and environmental conditions. In addition, they still fail to categorise bona fide subjects who inadvertently occlude their face with accessories such as glasses, scarves or masks to prevent respiratory infections. To mitigate these issues, this paper explores the utility of using different facial regions for PAD. In this context, a new metric, Face Region Utility, is proposed, which indicates the usefulness of a particular test region to spot an attack attempt based on another training region. A thorough evaluation in challenging scenarios on well-known databases shows which face regions can successfully substitute the full face to detect APs in scenarios where pristine subjects use some of the mentioned accessories: up to a 67.73% of detection performance improvement is yielded by applying our proposed analysis when pristine subjects wear masks to prevent respiratory infections.

INDEX TERMS Biometrics, presentation attack detection, face, facial regions, facial region utility.

I. INTRODUCTION

The large variety of commercial and legal requests together with the availability of the relevant technologies (e.g., smartphones, digital cameras, GPUs) have led to the deployment of numerous face recognition (FR) systems in the last decade [1], [2]. In contrast to other biometric characteristics such as fingerprints and irises, the human face not only can be used to identify a person but can also inform about mood, intention and attention. On the other hand, a representation thereof (i.e. a facial image) can be easily copied or replicated by any unauthorised subject, e.g. through photos or videos stemming from social media. Thus, those impostors can get access to various applications such as financial transaction authentication, device unlocking, and automated cars, where FR systems are commonly deployed. In particular, unattended applications (e.g., remote authentication for automated payment - pay-by-face [3]) which do not require direct monitoring are the target of malicious subjects launching attack presentations. Therefore, the next generation of FR systems must include an efficient mechanism for the detection of attack presentations.

To that end, a large number of facial presentation attack detection (PAD) approaches have been recently proposed [4]. Their goal is to determine whether a face sample stems from a real subject (i.e., it is a bona fide presentation - BP) or is an artificial replica (i.e., it is an attack presentation - AP). The great success of deep learning and its application to several pattern recognition tasks has led as well to the development of powerful PAD methods. Those schemes are mostly based on

The associate editor coordinating the review of this manuscript and approving it for publication was Weizhi Meng.

FIGURE 1. Examples of web-collected facial images occluded by different accessories such as masks, glasses, hands, paper, and tattoos.

supervised learning where the input sample is classified into one of two categories (i.e., BP or AP). They have reported a remarkable detection performance when the samples for training and testing are created using the same set of Presentation Attack Instrument (PAI) species (e.g., a printed photo of a face). However, those algorithms drop their accuracy when the test samples are fabricated with an unknown set of PAI species.

A peculiarity of most PAD approaches in FR systems is that they detect AP attempts by analysing the full face, thus ignoring that the attributes separating a BP from an AP vary their location on the face depending on the PAI species, the subject and the environmental conditions. To address this shortcoming, some PAD subsystems spot APs based on the analysis of several local patches extracted around the full face [5], [6], [7], [8]. Despite the improvement achieved by those methods, they still fail when pristine subjects inadvertently occlude their face with different accessories, as shown in Fig. 1. In particular, the use of accessories such as masks to prevent respiratory infection, glasses, or traditional clothes have resulted in a detection performance deterioration of most PAD algorithms that analyse the full face or local patches [9].

To fill this gap in the literature, in this work, we present an in-depth study of various facial regions to determine which are most appropriate for PAD in different scenarios. In contrast to our previous research [10], we analyse to what extent the use of a given facial region can produce similar or superior results to those obtained with the full facial image which might contain some of the above occlusions (see Fig. 1). In summary, the main contributions of this study are:

- An in-depth analysis compliant with the metrics defined in the international standard ISO/IEC 30107-3 [11] for biometric PAD of several facial regions.
- A study on the impact of the facial region resize on the PAD performance.
- A comprehensive analysis of the impact of wearing glasses for PAD.
- The definition of a metric, *Face Region Utility*, which combines correlation between facial regions and algorithm's detection performance to determine the most useful regions for PAD. The empirical results computed by the proposed metric can be successfully employed to replace the full face in those applications whose pristine subjects partially occlude their face with some kind of accessory without the need to retrain the base PAD algorithm.
- A benchmark of the best performing facial regions according to the newly defined utility metric in challenging scenarios including unknown environmental conditions, unknown PAI species, and cross-database.
- A benchmark of state-of-the-art PAD approaches for a particular use case where individuals wore masks to prevent SARS-CoV-2 coronavirus. We evaluate to which extent the central region of the faces could outperform the results achieved by the full face on a masked database.

The remainder of this manuscript is organised as follows: a review of facial PAD methods is included in Sect. II. Sect. III presents general concepts and the definition of facial regions studied in our work. The experimental setup is explained in Sect. IV. The experimental results of the analysis of facial regions for PAD are discussed in Sect V. Finally, conclusions and future work directions are presented in Sect. VI.

II. RELATED WORK

To mitigate the threats posed by attack presentations and thus increase the security of biometric systems, PAD has been one of the most studied topics in the last decade. Numerous literature reviews [4], [12], [13] classify PAD approaches into two broad categories: hardware and software. PAD algorithms belonging to the former detect living characteristics of a human body such as intrinsic properties (e.g., reflectance [14], [15]), involuntary signals (e.g., thermal radiation [16]), or responses to external stimuli (e.g., motion estimation [17]) by adding an extra sensor to the capture device. Those hardware-based methods are usually tailored for a particular PAI species, thereby reporting a high detection performance for it. However, they decrease their accuracy when the PAI species used in the evaluation are unknown [12]. Furthermore, an extra sensor combined with the capture device can significantly increase their production cost (e.g., a thermal sensor for an iPhone exceeds 250 EUR[1]).

In contrast, software-based methods are interoperable not only for face systems. Some of these schemes explore certain involuntary gestures of a face or head e.g. eye-blinking [18], lips movement, nodding, smiling, and looking in different directions [19], [20]. Other algorithms analyse texture properties in the images through e.g. Fourier Spectrum [21], Gaussian filters [22], statistical models [23], or traditional

[1] https://amz.run/44Mp

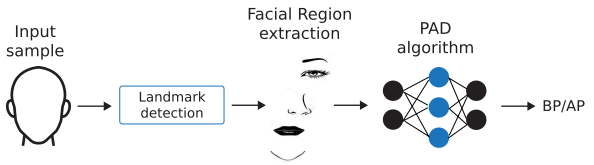


FIGURE 2. Proposed framework to evaluate PAD approaches.

texture descriptors - Local Binary Patterns (LBP) [24], [25], Histogram of Oriented Gradients (HOG) [26], Binarized Statistical Image Features (BSIF) [27], Local Phase Quantization (LPQ) [28]. In addition, some methods combine textural descriptors with generative models to improve the generalisation capabilities to unknown attacks [29], [30].

The advances experienced by deep learning techniques and their great success in several pattern recognition tasks have led to the development of powerful architectures for PAD [31], [32], [33], which outperform the aforementioned handcrafted-based methods. In 2014, Yang et al. [34] fine-tuned ImageNet pre-trained CaffeNet [35] and VGG-face [36] models to distinguish a BP from an AP. Following this idea, Xu et al. [37] combined Long Short-Term Memory (LSTM) units with Convolutional Neural Networks (CNNs) to learn temporal features from face videos. To improve the lack of generalisation capability of PAD subsystems, Sanghvi et al. [38] combined three CNN sub-architectures to spot the three common face PAI species i.e., print, replay, and mask attacks. Other techniques [39], [40] have also proposed CNNs to analyse properties in 3D mask attacks based on the fact that 2D face PAD algorithms suffer a significant detection performance degradation on this type of PAI species. Given that acquisition properties such as facial appearance, pose, illumination, capture devices, PAI species, and even subjects vary between datasets, several important face PAD approaches have recently explored domain adaptation (DA) to align the features of two different domains [41], [42], [43], [44]. This would improve the detection of previously unseen PAIs.

As mentioned in Sect. I, most state-of-the-art PAD algorithms detect AP attempts analysing the full face region, thus reporting a decrease in performance when some accessories such as glasses, masks to prevent respiratory infections, or tattoos occlude parts of the face. In fact, those PAD methods [5], [6], [7], [8], which have demonstrated the advantage of local face patches in defending against a variety of PAI species, dropped their performance in detecting BPs when pristine local patches contain some of the aforementioned accessories. Thus, these approaches might also fail to correctly separate a BP with occlusion from an intentional AP attempt. Up to now, few studies have addressed the impact of some of these occlusions for PAD [45] and most of them are focused on the analysis of facial images having masks to avoid SARS-CoV-2 coronavirus [45]. For further details on general facial PAD, the reader is referred to [4], [12], and [13].

In our research, we abstract from the fact that the input facial samples might contain some of the aforementioned occlusions and present a comprehensive analysis of the usefulness of various facial regions for PAD. These facial regions could, in turn, be used in a variety of realistic and challenging scenarios, including those in which certain occlusions are detected.

TABLE 1. Definition of facial regions by landmarks.

	Region	Enclosing landmarks
1	Full Face	The entire face region
2	Left Face	Region left of landmark 27
3	Right Face	Region Right of landmark 27
4	Central Face	[0, 33, 16, 24, 19]
5	Jaw	[1, 8, 15, 28]
6	Both Eyebrows	[17, 19, 24, 26]
7	Both Eyes	[0, 28, 16, 26, 17]
8	Left Eyebrow	[22, 24, 26]
9	Right Eyebrow	[17, 19, 21]
10	Left Eye	[42, 44, 45, 46]
11	Right Eye	[36, 38, 39, 40]
12	Mouth	[48, 50, 52, 54, 57]
13	Nose	[27, 31, 33, 35]
14	Chin	[4, 8, 11, 57]

III. PROPOSED FRAMEWORK

In our work the feasibility of using 14 facial regions for PAD purposes: both eyes, both eyebrows, central face, chin, jaw, left eye, right eye, left eyebrow, right eyebrow, mouth, nose, left face, and right face regions is explored. Fig. 2 shows the framework proposed to conduct our analysis, which is based on two main steps: *i*) the facial region is detected and extracted (see Sect. III-A), and *ii*) the facial region is the input to a PAD approach (see Sect. III-B) for BP vs. AP decision.

A. FACIAL REGIONS EXTRACTION

For facial region detection and extraction, the open-source toolbox dlib [46] which extracts 68 landmarks per face is considered. Based on such landmarks, 14 different facial regions in Tab. 1 are defined. For a comprehensive analysis, these regions are divided into two groups: single (i.e., mouth, nose, chin, left eye, right eye, left eyebrow, and right eyebrow) and composite (i.e., both eyes, both eyebrows, central face, jaw, left face, and right face, full face). Fig. 3 shows an example of those landmarks together with some facial regions. Note that the left (right) region comprises the facial portion to the left (right) of Landmark 27, bounded by the top of the forehead and Landmark 8.

B. PAD METHODS

Five state-of-the-art CNN approaches are independently evaluated:

- *AlexNet* is one of the first deep architectures developed in 2012 [48] which outperformed all traditional machine learning and computer vision approaches in the ImageNet challenge [48].
- *DenseNet* was proposed by Huang et al. [49]. The network connects each layer to every other layer in a feed-forward fashion as long as they have the same feature map size, thus reducing the vanishing gradient problem as the dense connections introduce short

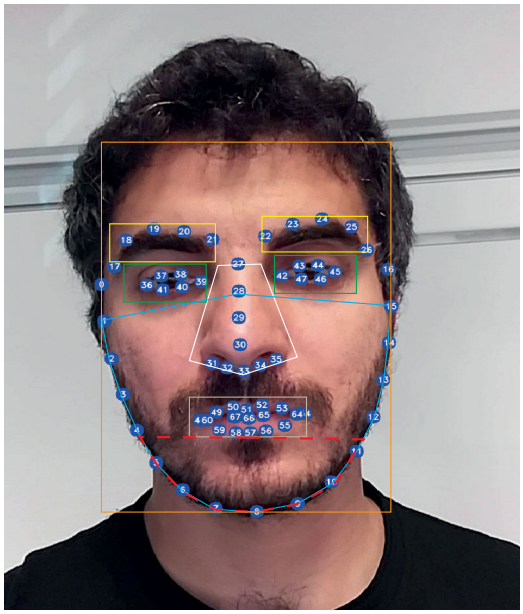


FIGURE 3. Some facial regions (i.e., mouth, nose, left and right eyes, left and right eyebrows, chin and jaw) computed from landmarks extracted from an image in REPLAY-MOBILE [47].

paths from inputs to outputs [49]. Moreover, this allows implicit deep supervision since the individual layers receive supervision from the loss function due to the shorter paths. The DenseNet model with 121 layers is utilised in the experiments.

- *ResNet* was the winner of ImageNet challenge in 2015 [50]. Like DenseNet, it does not suffer from the vanishing gradient problem [50]. ResNet introduces a residual connection strategy which improves both the training speed and accuracy. In our implementation, the version of 101 layers is adopted.
- *MobileNetV2*, in contrast to the aforementioned architectures, was developed by Sandler et al. [51] with a focus on mobile applications. This network included a novel layer, named inverted residual with linear bottleneck, which reduces the number of parameters to learn.
- *MNasNet* is a recent lightweight CNN developed by Tan et al. [52] for mobile applications. This architecture incorporates model latency into the main objective function to identify a model that achieves a good trade-off between accuracy and latency (i.e., inference time).

In our implementation, the last fully connected (FC) layer for all deep learning architectures studied is modified to a single neuron with a sigmoid activation for the BP vs. AP decision.

C. FACIAL REGION UTILITY

We define a new metric named *Facial Region Utility* which combines the correlation between facial regions and the detection performance of algorithms when they are trained using a particular facial region and evaluated on another one. This metric reports a value in the range $[0, \ldots, 1]$ which indicates the usefulness of a particular region for training to spot an attack presentation based on the other region in a probe image. Formally, the *Facial Region Utility* for a probe facial region R_P with respect to a trained region R_T is defined as follows:

$$U(R_T, R_P) = \frac{|C(R_T, R_P)| + (1 - P(R_T, R_P))}{2}, \quad (1)$$

where $C(R_T, R_P)$ is the Pearson correlation coefficient between R_T and R_P. $C(R_T, R_P)$ reports a value in the range $[-1, 1]$ indicating how correlated the features of R_P are with those of R_T. Since the direction of the Pearson correlation between R_T and R_P does not lead to any improvement, the absolute value over the coefficient is applied. $P(R_T, R_P)$ represents the normalised Detection Equal Error Rate (D-EER) when R_P is evaluated using an algorithm trained over the region R_T. Utility values close to 1 state that R_T can be employed for training whilst R_P can be successfully used for detecting an AP in the probe image. To normalise the D-EER values to the range [0, 1], the traditional Min-Max normalisation [53] is employed:

$$normalised_{\text{D-EER}} = \frac{\text{D-EER} - \mathbf{min}_{\text{D-EER}}}{\mathbf{max}_{\text{D-EER}} - \mathbf{min}_{\text{D-EER}}}, \quad (2)$$

where $\mathbf{min}_{\text{D-EER}}$ and $\mathbf{max}_{\text{D-EER}}$ are, respectively, the minimum and maximum values of the set of D-EERs computed by the proposed PAD methods (see Sect. III-B) on different training and testing configurations of the facial regions.

To make the equation 1 clear to readers, the boundary cases are outlined. Let $\mathbf{A}$ be a PAD algorithm, for the best case, we assume that the performance of $\mathbf{A}$ on two face regions (i.e., $P_{\mathbf{A}}(R_T, R_P)$) would result in a D-EER $= 0.0$, and R_T and R_P are highly correlated (i.e., $C(R_T, R_P) = 1$). Therefore, the *Facial Region Utility* between R_T and R_P would achieve the highest value (i.e., $U(R_T, R_P) = 1$). On the contrary, for the worst case, $P_{\mathbf{A}}(R_T, R_P) = 1, 0$ and $C(R_T, R_P) = 0$, leading to a $U(R_T, R_P) = 0$.

IV. EXPERIMENTAL SETUP

The experimental evaluation goals are manifold: *i*) study the impact of image resolution for PAD across facial regions *ii*) assess the feasibility of using facial regions for PAD, *iii*) analyse the effect of wearing glasses on the detection performance of PAD techniques, *iv*) evaluate the correlation and detection performance of facial regions as well as their utility for being used on real applications where some face parts might be occluded, and *iv*) establish a benchmark of state-of-the-art using our proposed analysis. To reach our goals, we focus on three scenarios:

- *Known attacks*: Analyses of the detection performance when the same PAI species are used for training and testing. For this purpose, known-attack protocols in [22], [24], and [47] are employed.
- *Unknown PAI species*: Exploration of the detection performance when different PAI species are used for training and testing. Protocol 2 from OULU-NPU described in [54] is employed.

TABLE 2. A summary of databases considered in our experiments.

DB	#Samples	Capture device	Capture conditions	PAI species
CASIA FASD	600	Low-quality USB camera Normal-quality USB camera High-quality Sony NEX-5 camera	Natural scenes	Printed attacks, Cut photo, Video replay
REPLAY-ATTACK	1,200	Low-quality 13-inch MacBook webcam	Controlled, adverse scenes	Printed, Photo replay, Video replay
REPLAY-MOBILE	1,190	High-quality iPad Mini 2 High-quality LG G4	Controlled, adverse direct sunlight, lateral sunlight, diffuse and complex backgrounds	Printed, Photo replay, Video replay
CRMA	13,133	iPad Pro Galaxy Tab S6 Surface Pro-6	Realistic scenes	Printed and Video replay of subjects wearing masks
OULU-NPU	4,950	Samsung Galaxy S6 edge HTC Desire EYE MEIZU X5 ASUS Zenfone Selfie Sony XPERIA C5 Ultra Dual OPPO N3	Controlled and adverse scenes	Printed and Video replay

- *Cross-database*, in which the datasets employed for testing are different from the databases used for training. Both datasets contain the same PAI species to ensure that the performance degradation is due to the dataset change and not to the unknown PAI species. Protocol 3 from OULU-NPU described in [54] is adopted.

The proposed framework was implemented using PyTorch [55] and the CNNs were trained on the Nvidia Tesla M10 GPU with 16 GB DRAM. The algorithms are trained using the Adam optimiser [56]. Since the networks were initialised with the ImageNet pre-trained weights, a learning rate of 10^{-4} and a weight decay parameter of 10^{-6} are used.

A. DATABASES

The experimental evaluation is conducted with four well-established databases which are summarised in Tab. 2: CASIA Face Antispoofing database (CASIA-FASD) [22], REPLAY-ATTACK (RA) [24], REPLAY-MOBILE (RM) [47], and OULU-NPU [54]. Since most databases contain videos, a random frame per video to conduct our experiments is selected.

CASIA-FASD [22] is a small database containing 600 short videos captured from 50 different subjects under different environmental conditions. Three PAI species are included: *i*) warped photo attacks or printed attacks, in which the attackers place their face behind the hard copies of high-resolution digital photographs, *ii*) cut photo attacks, in which the face of the attacker is placed behind the hard copies of photos, where eyes have been cut out, and *iii*) video replay attacks, where attackers replay face videos using iPads. Three imaging qualities are also considered, namely the low quality, normal quality and high quality in the video acquisition.

REPLAY-ATTACK (RA) [24] has 1200 short videos of 50 different subjects. The videos were captured with a low-resolution webcam of a 13-inch MacBook Laptop under two different conditions: *i*) controlled, with uniform background and artificial lighting, and *ii*) adverse, with natural illumination and non-uniform background. Moreover, three PAI species are implemented: printed attacks, photo replay attacks (i.e., a photo is replayed by a smartphone to the capture device), and video replay attacks.

REPLAY-MOBILE (RM) [47] comprises 1190 video clips of printed attacks, photo replay attacks, and video replay attacks stemming from 40 subjects under different lighting conditions. This database focuses on the PAD evaluation over mobile scenarios, as videos were recorded with two smart capture devices: an iPad Mini2 and a LG-G4 smartphone.

Collaborative Real Mask Attack (CRMA) [9] consists of 423 BP videos and 12690 attacks of 47 subjects. The videos were acquired with three different high-definition capture devices in realistic scenarios. The PAI species are *i*) both unmasked (BM0) and masked (BM1) bona fide presentations, *ii*) printed and video replay attacks from subjects not wearing a mask (AM0), *iii*) printed and video replay attacks from subjects wearing a mask (AM1), and *iv*) partial attack where the unmasked printed/replayed faces are covered with real masks (AM2). The CRMA is challenging due to that it contains different masks to prevent SARS-CoV-2 coronavirus, multiple capture devices, and several capture distances. An example of different BP and AP samples is depicted in Fig. 4.

OULU-NPU [54] consists of 4950 high-resolution short video sequences of BPs and AP attempts stemming from 55 subjects. The BP samples were acquired in three different sessions under different illumination conditions and background scenes. The PAI species are printed attacks and video replay attacks which were recorded using the frontal cameras of six mobile phones. This database defines four different protocols as follows:

- Protocol 1 focuses on the generalisation ability of PAD techniques across different environmental conditions (i.e., illumination and background scenes). The settings

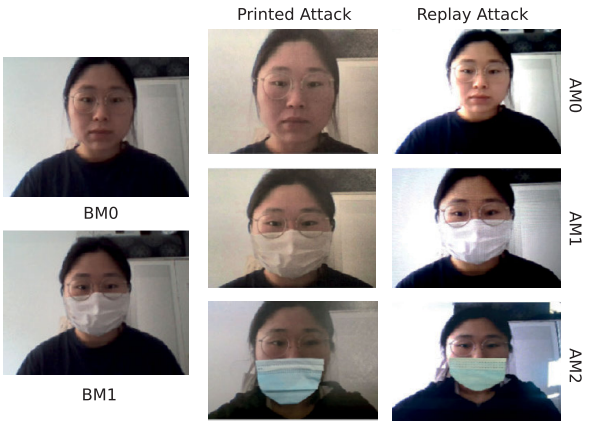


FIGURE 4. Example of BPs and APs in the CRMA database taken from [9].

of the environmental conditions used for the capture of the tested samples are different from those for the acquisition of the training images.

- Protocol 2 is designed to evaluate the PAD generalisation capability when the tested PAI species remain unknown from the training set (i.e., Unknown PAI species scenario).
- Protocol 3 analyses the capture device interoperability following a Leave One Camera Out (LOCO) protocol, where samples recorded by five smartphones are used for training whilst instances captured by the sixth mobile device are used for testing. In essence, this protocol evaluates cross-database scenarios.
- Protocol 4 is the most challenging scenario, as it combines all described protocols. In particular, the generalisation ability of PAD approaches across previously unknown illumination conditions, background scenes, PAI species, and capture devices is simultaneously evaluated.

B. EVALUATION METRICS

The experimental results are analysed and reported in compliance with the metrics defined in the international standard ISO/IEC 30107-3 [11] for biometric PAD:

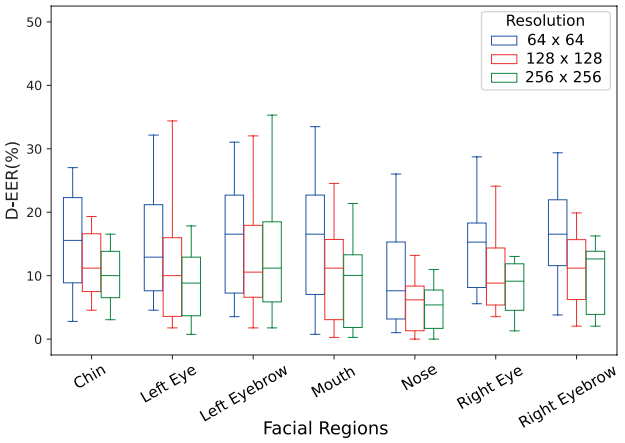
- Attack Presentation Classification Error Rate (APCER), which computes the proportion of attack presentations wrongly classified as bona fide presentations.
- Bona Fide Presentation Classification Error Rate (BPCER), which is defined as the proportion of bona fide presentations misclassified as attack presentations.

Based on these metrics, we report *i*) the BPCER observed at an APCER value or security threshold of 10% (BPCER10); and *ii*) the Detection Equal Error Rate (D-EER), which is defined as the error rate value at the operating point where APCER = BPCER.

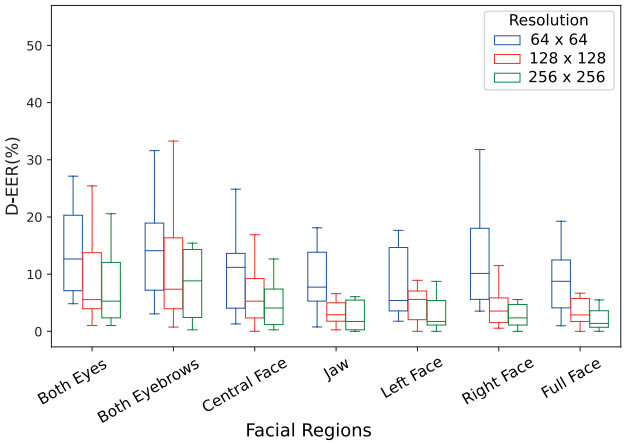
V. RESULTS AND DISCUSSION

A. KNOWN ATTACKS

In this section, we conduct several experiments aimed at evaluating the effect of varying image resolution and the use of glasses on PAD performance. For this purpose, known-attack scenarios, i.e. the PAI species used to generate the test database are known a priori in training are adopted. In this way, biases related to external variables such as PAI species, subject and environmental conditions are avoided.



(a) Impact on the single regions



(b) Impact on the composite regions

FIGURE 5. Impact of image resolution on different facial regions.

1) EFFECTS OF IMAGE RESOLUTION FOR PAD

Since the size of facial regions can vary across images, the impact of image resolution for PAD is investigated. To that end, the D-EER per facial region and algorithm defined in Sect. III-B over three databases is computed: CASIA, RM, and RA. Fig. 5 reports the boxplots per facial region over three resolutions i.e., 64×64, 128×128, 256×256: greater resolution configurations might result in a performance deterioration due to pixel value interpolation for the smallest regions. Note that the D-EER improves with the image resolution, thus yielding the best detection performance for an image size of 256×256 pixels. Observe that those regions having a large image size (e.g., full face, right face, left face, and jaw) report a low standard deviation (std) for an image resize greater or equal than 128×128 pixels (see red and

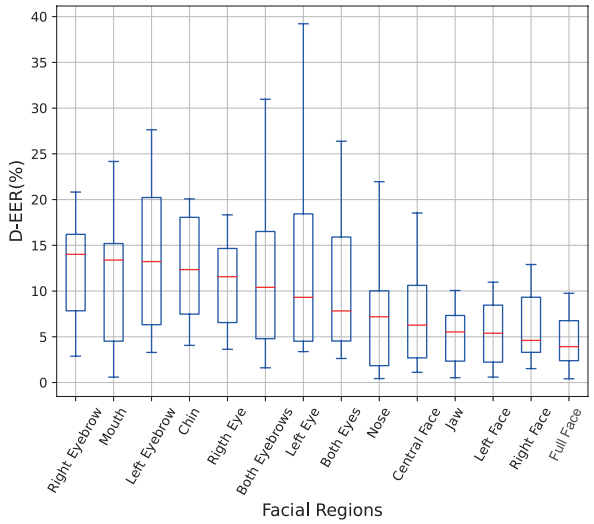


FIGURE 6. Best performing facial regions for known attacks.

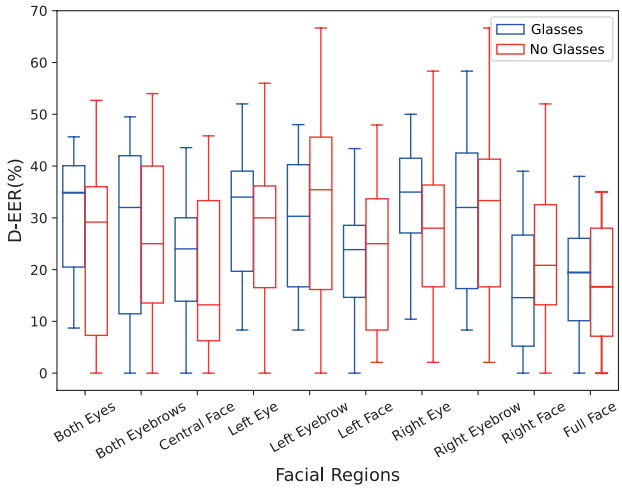


FIGURE 7. Detection performance for images containing glasses (blue boxes) and no glasses (red boxes).

green boxes in Fig. 5-b): the mean std is approximately 6.99. In contrast, their standard deviation increases when a small size of 64 × 64 is used (see blue boxes in Fig. 5-b): the mean std is approximately 10.25.

Following the above observations, it can be also seen that the pixel value estimation for the smallest facial regions (i.e., left and right eyes, left and right eyebrows, both eyebrows, both eyes, mouth, nose, and chin) during the resize significantly affects the algorithm's detection performance, thus resulting in a high standard deviation in the ranges [6.72, . . . , 13.62]. These resolution results confirm the findings in [57]: the up-sampling or down-sampling step performed by the deep learning approaches to adjust the size of a given image in the input layer leads to an information loss of artefacts for the smallest or largest sizes, respectively.

Since most facial regions report on average their best detection performance for a resize configuration of 256 × 256 pixels, we select it for further experiments.

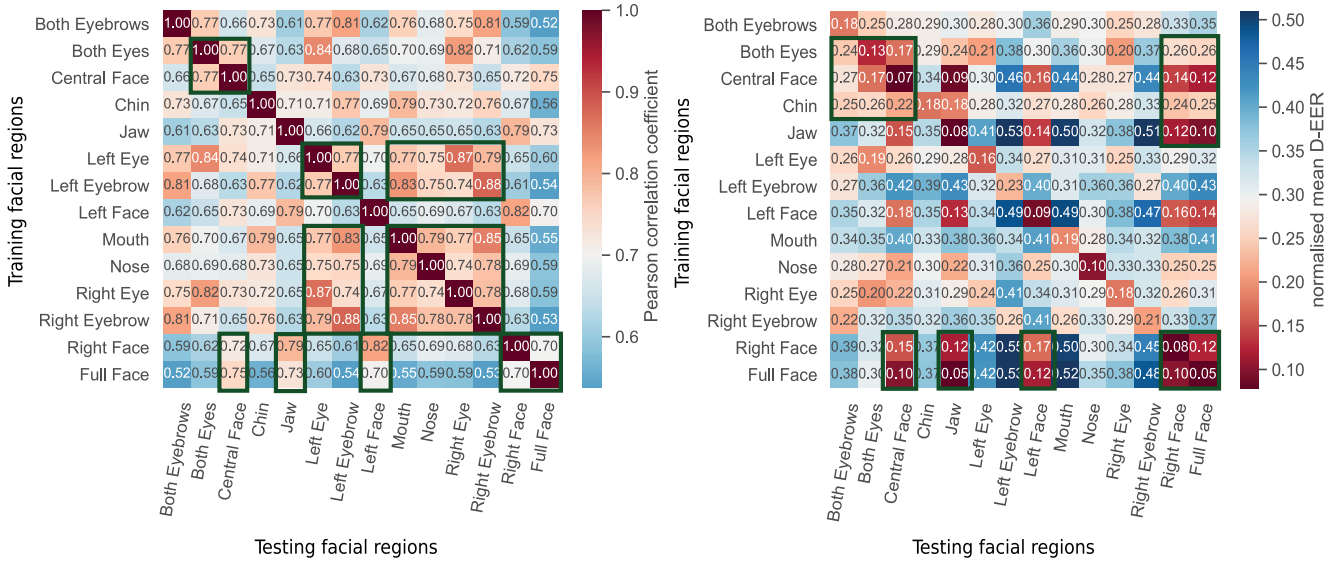
2) DETECTION PERFORMANCE OF FACIAL REGIONS

In the second set of experiments, the PAD performance for each facial region over CASIA, RM, and RA databases following their corresponding known-attack protocols is evaluated. Similar to the above experiment, the D-EER per facial region and algorithm defined in Sect. III-B is computed - their detection performance is reported as boxplots in Fig. 6. As it may be noted, the training and evaluation of selected approaches using the full face attain on average the best D-EER: a median D-EER of 3.92% (indicated by the central blue mark in the boxplots) outperforms the remaining facial regions. Regarding composite regions, it can be observed that they report the best performances e.g., right face (median D-EER = 4.61%), left face (median D-EER = 5.38%), jaw (median D-EER = 5.53%), and central face (median D-EER = 6.28%). Furthermore, the error rates of these composite regions tend to their median values, thus resulting in a low standard deviation with respect to the mean values: their standard deviation is in the ranges [5.88, . . . , 7.97]. Among the single regions, the nose achieves the best detection performance, yielding a median D-EER of 7.19% with a standard deviation of 7.06. Even, this outperforms the performance attained by both eyes (median D-EER = 7.83%). Whereas 75% of the D-EER values of the nose region are below their median, only 25% of the error rates of both eyes are below their median, hence indicating that the nose is more suitable for PAD than both eyes. Since the nose is a flat region composed mostly of skin, we think that any variation in quality, colour, or texture can lead to an improvement in the detection of APs.

Note that the worst regions are the right and left eyebrows and mouth which report median D-EERs above 13% and standard deviations in the ranges [10.28, . . . , 12.08]. Observe that the union between both regions (i.e., both eyebrows) improves their individual errors by three percentage points (i.e., 10.41% for both eyebrows vs. 14.01% for the right eyebrow). This is because the region comprising both eyebrows includes flat skin in between which allows algorithms to detect APs. Similar behaviour can be also perceived in the results achieved for both eyes.

3) IMPACT OF WEARING GLASSES ON PAD

Note in Fig. 6 that most regions around the eyes (i.e., left and right eyes and left and right eyebrows) report a high-performance deterioration, thus yielding std values in the ranges [10, . . . , 12]. Based on this observation, the effect of wearing glasses in those regions that might contain such an accessory is investigated. To that end, we follow the same experimental evaluation used in Sect. V-A2 and split the training and evaluation sets from the CASIA, RM and RA databases into two balanced sets each containing faces with glasses and faces without glasses. The boxplots representing D-EERs computed by the methods defined in Sect. III-B, per facial region in the above databases are shown in Fig. 7. Note



(a) Absolute correlation values between facial regions

(b) Detection performance normalised between facial regions

FIGURE 8. Correlation and detection performance between facial regions. The green rectangles highlight some examples of facial region configurations which report high correlations and detection performances.

that *i*) wearing glasses affects the detection performance of approaches evaluated when trained using either the full face or the central face, *ii*) right and left faces are not affected by wearing glasses, thus yielding a better detection performance when faces contain glasses, *iii*) wearing glasses impact the PAD performance for both left and right eyes along their fusion (i.e., both eyes), and *iv*) whereas the performance for left and right eyebrows is not highly affected by wearing glasses, the fusion region (i.e., both eyebrows) is. The latter is due to the accuracy of the region extraction algorithm: it includes part of the glasses in the final images. These findings complement the study conducted in [58]: wearing glasses also has a negative impact on iris segmentation and thus on iris recognition. A possible solution to improve the detection performance of PAD techniques against subjects wearing glasses would therefore be to first focus on detecting the glasses and then use the jaw area, which reports high performance in Fig. 6 to reject AP attempts.

4) CROSS-DETECTION PERFORMANCE, CORRELATION, AND UTILITY

Now the correlation between facial regions is explored. For this purpose, the PAD approaches for each facial region over the CASIA, RM, and RA databases are first trained. On the evaluation sets, latent vectors from the last FC layer before the final decision layer are extracted and averaged them. Fig. 8-a shows the average Pearson correlation coefficient between facial regions. It should be noted that the features representing the facial region combination share at least 50% of their characteristics with each other. Facial regions that are highly correlated with each other are highlighted with a green rectangle. Specifically, the latent vectors of facial regions such as the left and right eyes, left and right eyebrows, mouth and nose report as expected a high Pearson correlation ranging from 0.74 to 1.00. As expected, the right and left regions of the faces share 82% of their characteristics. Therefore, they can be interchangeably used for PAD. Finally, the full face is highly correlated with the central part of the face, followed by the jaw and the left and right regions of the face.

Following the above idea, the detection performance between facial regions (as it is illustrated in Fig. 8-b) is computed. In this experiment, the architectures are trained using one facial region (depicted by the rows) and evaluated on the remaining regions (shown by the columns) over three databases (i.e., CASIA, RM, and RA). Then, the mean D-EER between facial region combinations is reported in Fig. 8-b. It is important to point out that error rates are normalised following the eq. 2. It should be observed that the evaluation of facial regions such as the jaw, central face, and left and right regions achieves the best detection when the algorithms are trained using the full face. In fact, the jaw yields the same D-EER as the one attained by the full face (i.e., D-EER = 0.05). Subsequently, the central face and left and right face regions depict similar detection performance (i.e., 0.10 vs. 0.12). As a consequence of these results, we perceive that the full face can be used to spot an AP attempt in the probe image when PAD algorithms are trained on either jaw, left face, right face, or central face regions.

Based on Fig. 8, the *Facial Region Utility* is computed using the eq. 1 and reported in Fig. 9. As was mentioned in Sect. III-C, this metric indicates the usefulness of a particular region for training to spot an attack presentation based on the other region in a probe image. As can be observed, the same region used simultaneously for training and testing reports

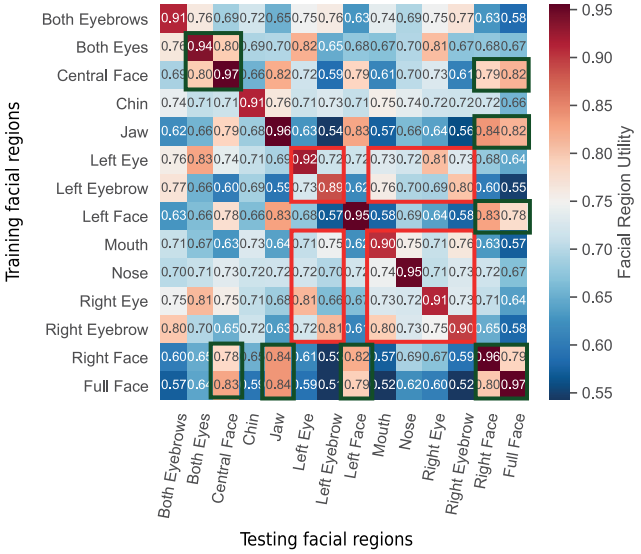


FIGURE 9. ***Facial Region Utility*** **computed from the correlation and detection performance matrices. The green rectangles highlight the combinations of facial regions with a high utility. The red rectangles state those examples of facial region combinations whose correlation and detection performance values show a contrary trend in Fig. 8.**

the best utility (i.e., diagonal values). Note that facial regions such as the full face, left and right faces, central face, and jaw can be used the create a reliable train and test configuration as they report high *Facial Region Utility* values. In particular, the training of a PAD approach over the full face (i.e, red rectangle at the bottom) allows the successful evaluation of regions such as jaw ($U(\cdot) = 0.84$), central face ($U(\cdot) = 0.83$), right face ($U(\cdot) = 0.80$), and left face ($U(\cdot) = 0.79$). It should be noted that the *Facial Region Utility* highly depends both on the correlation and the algorithm's detection performance. Those train-test facial regions which drop their *Facial Region Utility* due to contrary trends depicted in Fig. 8 are highlighted with red rectangles. Whereas mouth, nose, right and left eyes and left and right eyebrows pose a high correlation with each other (see Fig. 8-a), the detection performance between them decreases considerably (see Fig. 8-b). Therefore, they are not suitable for a PAD train-test configuration.

B. ANALYSIS OF THE FACIAL REGION UTILITY ON CHALLENGING SCENARIOS

To verify the usefulness of the *Facial Region Utility*, several state-of-the-art PAD techniques are trained using the full face. The best-utility regions (i.e., jaw, central face, right face, and left face) are then evaluated - their detection performance over the challenging protocols in the OULU-NPU database [54] is reported in Tab. 3. Note that the results depicted in Tab. 3 might differ from the ones yielded by their corresponding papers. In contrast to the original pipelines which use all video frames to make the final decision, these algorithms were trained and assessed using a random video frame. Observe that the D-EERs improve with the utility of facial regions independently of the evaluated protocol. Specifically,

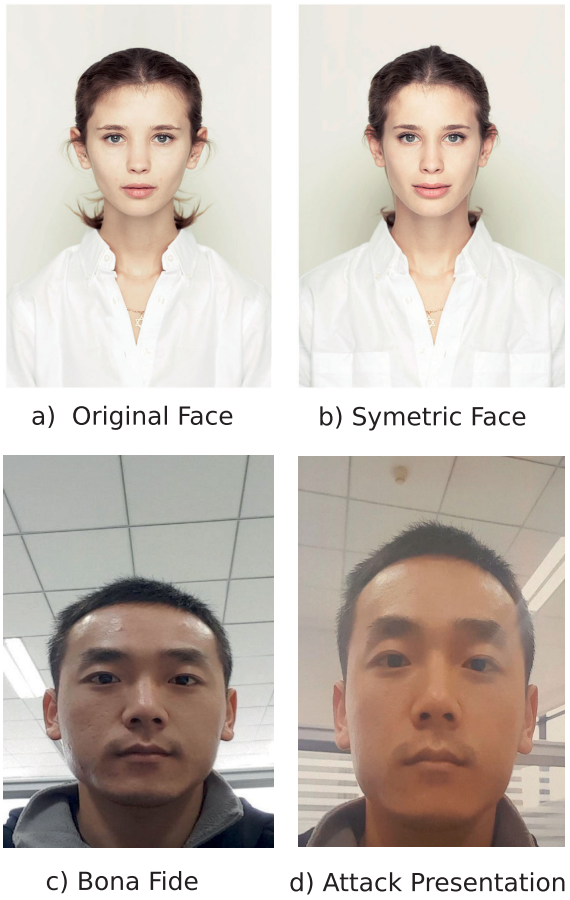


FIGURE 10. **Some images show why the detection performance between left and right faces is different. *a*) and *b*) represent the visual differences between a perfect symmetrical face (i.e., *b*) and its original face (i.e., *a*) [60]. *c* and *d* are examples of BP and AP in OULU-NPU whose artificial light configurations differ from each other.**

the best detection performance is yielded by the full face, followed by the jaw. According to the *Facial Region Utility*, the central face is the third best region to spot an AP attempt in a probe image after the full face and jaw. However, this region reports a detection performance decrease with respect to the results achieved by the right and left faces. This behaviour mostly happens due to the sensitivity of this region to the use of glasses (see Fig. 7). On the other hand, the right face outperforms the left face in all experiments. This is mainly due to variables such as the asymmetry of the face and the artificial light positions used in the BP and AP acquisition. The latter causes most of the characteristics separating a BP from an AP to be detected in the right region of the face (see Fig. 10).

Note that the detection performance attained by the handcrafted-based technique (i.e., FV-GMM) shows for the jaw an improvement regarding the remaining regions. Unlike deep learning approaches evaluated, this algorithm derives a kernel from the parameters learned by a generative model (i.e., Gaussian Mixture Models (GMM) [61]) to characterise how the distribution of a set of unknown local descriptors

TABLE 3. Benchmark of the state-of-the-art algorithms trained on the full face and evaluated on the regions with the best *Facial Region Utility* in terms of D-EER(%) using the OULU-NPU database [54].

P	Approach	Facial Regions				
		Full Face	Jaw	Central Face	Right Face	Left Face
1	FV-GMM [30]	8.19 ± 1.19	7.59 ± 0.80	13.71 ± 2.39	8.54 ± 1.05	12.06 ± 1.07
	DeepPixelBis [31]	4.17	6.67	6.67	4.48	10.83
	CDCN [59]	4.48	10.83	20.94	15.83	16.68
2	FV-GMM [30]	8.30 ± 1.75	7.23 ± 1.37	11.71 ± 2.59	9.53 ± 2.20	10.74 ± 2.56
	DeepPixelBis [31]	2.78	3.83	7.25	8.01	10.52
	CDCN [59]	3.96	11.11	18.89	16.94	17.78
3	FV-GMM [30]	8.29 ± 6.75	8.27 ± 6.54	12.59 ± 6.42	10.22 ± 6.50	14.92 ± 8.41
	DeepPixelBis [31]	1.25 ± 1.23	4.10 ± 2.80	6.33 ± 5.49	5.53 ± 5.47	7.62 ± 5.08
	CDCN [59]	1.88 ± 0.93	13.40 ± 3.38	22.78 ± 3.75	13.47 ± 5.30	14.27 ± 5.41
4	FV-GMM [30]	24.86 ± 8.47	19.88 ± 10.17	27.66 ± 7.62	20.56 ± 13.12	28.53 ± 6.19
	DeepPixelBis [31]	10.42 ± 10.51	12.71 ± 5.83	16.67 ± 5.74	16.04 ± 13.66	19.58 ± 11.64
	CDCN [59]	13.54 ± 4.21	22.29 ± 7.56	28.33 ± 6.83	22.71 ± 14.88	22.71 ± 12.31

TABLE 4. The detection performance of the DeepPixelBis algorithm on the CRMA database. The PAD decision threshold employed in the APCER, BPCER, and ACER computation is the one yielded at a BPCER10 (i.e., BPCER@APCER = 10%) on only unmasked data in the development set.

Approach	BPCER (%)		APCER (%) - Print			APCER (%) - Replay			ACER (%)
	BM0	BM1	AM0	AM1	AM1	AM0	AM1	AM1	
DeepPixelBis (Full Face) [31]	63.16	64.04	0.00	0.00	0.00	0.00	0.64	0.00	29.47
DeepPixelBis$_{\mathbf{RW}}$ (Full Face) [45]	35.09	41.23	0.00	0.10	1.17	0.19	1.95	0.58	18.58
DeepPixelBis$_{\mathbf{PAL}}$ (Full Face) [45]	42.11	51.75	0.00	0.00	0.00	0.00	0.44	0.00	23.35
DeepPixelBis$_{\mathbf{RW-PAL}}$ (Full Face) [45]	26.32	29.82	0.00	0.19	1.17	0.00	1.32	0.29	14.81
DeepPixelBis (Central Face)	7.02	15.79	3.12	2.83	7.02	8.48	6.84	9.06	**9.51**

differs from the distribution of known features. Therefore, this does not require the probe image to be similar in terms of shape to the trained samples. It can be observed that deep learning schemes suffer from a detection performance deterioration when the object in the probe image (e.g., the jaw) is different from the one used for training (i.e., the full face). We think that deep learning solutions focused on local patches could improve the above limitation.

C. BENCHMARK WITH THE STATE OF THE ART

The usefulness of the *Facial Region Utility* for a real application where subjects wore masks to prevent respiratory infections is also evaluated. For this purpose, the best performing algorithm in Tab. 3 (i.e., DeepPixelBis) is selected and a benchmark with the state-of-the-art techniques in Tab. 4 over the CRMA database is established. To build a realistic analysis where the behaviour of the PAD on facial images containing masks to prevent SARS-CoV-2 coronavirus is still unknown, we follow the experimental setup in [45] and report the APCER and BPCER values by using the threshold BPCER10 that is computed on only unmasked data in the development. In this experiment, the algorithms are trained on the full faces and evaluated either on the full face (i.e., the four first rows) or the central face (i.e., the last row). As the jaw which is the second region with the highest utility (see Fig. 9) is fully occluded by the masks, i.e. a large drop in the detection performance is expected, the central face is directly evaluated. The Average Classification Error Rate (ACER) is computed due to the lack of a proper evaluation of the state-of-the-art compliant with the ISO/IEC 30107-3 [11] for biometric PAD.

Note in Tab. 4 that the evaluation of the algorithms using the full face leads to a significant detection performance deterioration. In particular, the BPCER values for BM0 and BM1 are considerably high (i.e., first row, BPCER $\geq$ 63.16%), thereby confirming our initial hypothesis: PAD algorithms misclassify BPs as an intentional AP when subjects wear some accessories e.g., masks. In fact, the Regional Weight (RW) and Partial Attack Label (PAL) methodologies proposed in [45] to mitigate such attacks build a secure (APCER $\leq$ 1.95%) but not convenient (BPCER $\geq$ 26.32) PAD subsystem. In contrast, the evaluation of the central region alongside the earlier detection of a facial mask to avoid SARS-CoV-2 coronavirus resulted in an overall improvement of the detection performance: an ACER of 9.51% which outperforms the DeepPixelBis [31] and DeepPixelBis$_{\mathbf{RW-PAL}}$ [45] method by relative improvements of 67.73% and 17.98%, respectively, allows the building of a secure and convenient system. Finally, it is worth noting that the subjects only used glasses in 16% of the images in the

CRMA database unlike OULU-NPU, whose subjects wore glasses in 50% of the images. Therefore, these results are not fully biased by this type of accessory.

VI. CONCLUSION

In this work, the feasibility of using different facial regions for PAD was explored. In particular, 14 regions including single and composite regions were evaluated in compliance with the metrics defined in the international standard ISO/IEC 30107-3 [11] for biometric PAD. The experimental evaluation conducted over the freely available databases such as CASIA, REPLAY-MOBILE, REPLAY-ATTACK, CRMA, and OULU-NPU depicted that the composite regions achieved the best detection performances. In particular, the full face yielded a median D-EER of 3.92%, followed by the right (median D-EER = 4.61%) and left (median D-EER = 5.38%) faces, jaw (median D-EER = 5.53%), and central face (median D-EER = 6.28%). As expected, there exists a correlation between representing the left and right regions of the face as well as both eyes and eyebrows. In addition, the proposed *Facial Region Utility* metric indicated those regions capable of being used in unattended applications where subjects have some common accessories. In fact, these facial regions with a high *Facial Region Utility* (i.e., jaw, central face, left and right face) can be also combined to improve the particular result reported by the use of the full face on those applications.

In our work, we also showed the usefulness of the *Facial Region Utility* for a particular use case where individuals wore masks to prevent respiratory infections: the use of the central face over the full face yielded an ACER of 9.51% which outperforms the state-of-the-art methods by a relative improvement up to 67.73%. We noted that the BPCER values yielded by the state-of-the-art were decreased down to BM0 = 7.02% and BM1 = 15.79% for pristine subjects. These results allow therefore the building of a secure and convenient PAD module.

The use of other accessories such as glasses also impacts the detection performance of algorithms when either the full face, eyes, or central face is used to detect AP attempts. We observed that increasing the size of facial regions also affects the detection performance of the analysed algorithms: 256×256 pixels reported the best results for all regions. The pixel value estimation for the smallest facial regions such as left and right eyes, left and right eyebrows, both eyebrows, both eyes, mouth, nose, and chin during the resize considerably affects the algorithm's detection performance, thus resulting in high standard deviation values.

Finally, a disadvantage of the proposed analysis relies on the detection of facial landmarks and thus on the extraction of regions in images whose pristine subjects have some kind of disease, e.g. paralysis.[2] AP attempts launched by these individuals should be appropriately addressed in future work.

[2]https://www.facialparalysisinstitute.com/photo-gallery/

REFERENCES

[1] J. Deng, J. Guo, N. Xue, and S. Zafeiriou, "ArcFace: Additive angular margin loss for deep face recognition," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 4685–4694.
[2] F. Boutros, N. Damer, F. Kirchbuchner, and A. Kuijper, "ElasticFace: Elastic margin loss for deep face recognition," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Jun. 2022, pp. 1577–1586.
[3] Finextra. (Jan. 2020). *Worldline Takes Pay-by-Face System on the Road*. Accessed: Feb. 2, 2023. [Online]. Available: https://www.finextra.com/newsarticle/35009/worldline-takes-pay-by-face-system-on-the-road/
[4] Y. S. El-Din, M. N. Moustafa, and H. Mahdi, "Deep convolutional neural networks for face and iris presentation attack detection: Survey and case study," *IET Biometrics*, vol. 9, no. 5, pp. 179–193, Sep. 2020.
[5] X. Yang, W. Luo, L. Bao, Y. Gao, D. Gong, S. Zheng, Z. Li, and W. Liu, "Face anti-spoofing: Model matters, so does data," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 3502–3511.
[6] T. Shen, Y. Huang, and Z. Tong, "FaceBagNet: Bag-of-local-features model for multi-modal face anti-spoofing," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Jun. 2019, pp. 1611–1616.
[7] R. Cai, H. Li, S. Wang, C. Chen, and A. C. Kot, "DRL-FAS: A novel framework based on deep reinforcement learning for face anti-spoofing," *IEEE Trans. Inf. Forensics Security*, vol. 16, pp. 937–951, 2021.
[8] A. Kantarci, H. Dertli, and H. K. Ekenel, "Shuffled patch-wise supervision for presentation attack detection," in *Proc. Int. Conf. Biometrics Special Interest Group (BIOSIG)*, Sep. 2021, pp. 1–5.
[9] M. Fang, N. Damer, F. Kirchbuchner, and A. Kuijper, "Real masks and spoof faces: On the masked face presentation attack detection," *Pattern Recognit.*, vol. 123, Mar. 2022, Art. no. 108398.
[10] L. J. González-Soler, "Generalisable presentation attack detection for multiple types of biometric characteristics," Doctoral thesis, Dept. Comput. Sci., Hochschule Darmstadt, Darmstadt, Germany, 2023.
[11] *ISO/IEC 30107-3. Information Technology—Biometric Presentation Attack Detection—Part 3: Testing and Reporting*, International Organization for Standardization, Standard ISO/IEC JTC1 SC37 Biometrics, 2017.
[12] J. Galbally, S. Marcel, and J. Fierrez, "Biometric antispoofing methods: A survey in face recognition," *IEEE Access*, vol. 2, pp. 1530–1552, 2014.
[13] R. Ramachandra and C. Busch, "Presentation attack detection methods for face recognition systems: A comprehensive survey," *ACM Comput. Surveys*, vol. 50, no. 1, pp. 1–37, Jan. 2018.
[14] N. Kose and J. Dugelay, "Reflectance analysis based countermeasure technique to detect face mask attacks," in *Proc. 18th Int. Conf. Digit. Signal Process. (DSP)*, Jul. 2013, pp. 1–6.
[15] Y. Wang, X. Hao, Y. Hou, and C. Guo, "A new multispectral method for face liveness detection," in *Proc. 2nd IAPR Asian Conf. Pattern Recognit.*, Nov. 2013, pp. 922–926.
[16] L. Sun, W. Huang, and M. Wu, "TIR/VIS correlation for liveness detection in face recognition," in *Proc. Int. Conf. Comput. Anal. Images Patterns*. Berlin, Germany: Springer, 2011, pp. 114–121.
[17] K. Kollreider, H. Fronthaler, and J. Bigun, "Non-intrusive liveness detection by face images," *Image Vis. Comput.*, vol. 27, no. 3, pp. 233–244, Feb. 2009.
[18] K. Patel, H. Han, and A. K. Jain, "Cross-database face antispoofing with robust feature representation," in *Proc. Chin. Conf. Biometric Recognit.*, 2016, pp. 611–619.
[19] J. Bigun, H. Fronthaler, and K. Kollreider, "Assuring liveness in biometric identity authentication by real-time face tracking," in *Proc. IEEE Int. Conf. Comput. Intell. Homeland Secur. Pers. Saf.*, Jul. 2004, pp. 104–111.
[20] S. Tirunagari, N. Poh, D. Windridge, A. Iorliam, N. Suki, and A. T. S. Ho, "Detection of face spoofing using visual dynamics," *IEEE Trans. Inf. Forensics Security*, vol. 10, no. 4, pp. 762–777, Apr. 2015.
[21] J. Li, Y. Wang, T. Tan, and A. K. Jain, "Live face detection based on the analysis of Fourier spectra," *Proc. SPIE*, vol. 5404, pp. 296–303, Aug. 2004.
[22] Z. Zhang, J. Yan, S. Liu, Z. Lei, D. Yi, and S. Z. Li, "A face antispoofing database with diverse attacks," in *Proc. 5th IAPR Int. Conf. Biometrics (ICB)*, Mar. 2012, pp. 26–31.
[23] H. P. Nguyen, A. Delahaies, F. Retraint, and F. Morain-Nicolier, "Face presentation attack detection based on a statistical model of image noise," *IEEE Access*, vol. 7, pp. 175429–175442, 2019.

[24] I. Chingovska, A. Anjos, and S. Marcel, "On the effectiveness of local binary patterns in face anti-spoofing," in *Proc. Int. Conf. Biometrics Special Interest Group (BIOSIG)*, Sep. 2012, pp. 1–7.
[25] F. Peng, L. Qin, and M. Long, "Face presentation attack detection based on chromatic co-occurrence of local binary pattern and ensemble learning," *J. Vis. Commun. Image Represent.*, vol. 66, Jan. 2020, Art. no. 102746.
[26] A. Agarwal, D. Yadav, N. Kohli, R. Singh, M. Vatsa, and A. Noore, "Face presentation attack with latex masks in multispectral videos," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Jul. 2017, pp. 275–283.
[27] S. R. Arashloo, J. Kittler, and W. Christmas, "Face spoofing detection based on multiple descriptor fusion using multiscale dynamic binarized statistical image features," *IEEE Trans. Inf. Forensics Security*, vol. 10, no. 11, pp. 2396–2407, Nov. 2015.
[28] R. Raghavendra, S. Venkatesh, K. B. Raja, P. Wasnik, M. Stokkenes, and C. Busch, "Fusion of multi-scale local phase quantization features for face presentation attack detection," in *Proc. 21st Int. Conf. Inf. Fusion (FUSION)*, Jul. 2018, pp. 2107–2112.
[29] L. J. Gonzalez-Soler, M. Gomez-Barrero, and C. Busch, "Fisher vector encoding of dense-BSIF features for unknown face presentation attack detection," in *Proc. Int. Conf. Biometrics Special Interest Group (BIOSIG)*, Sep. 2020, pp. 1–6.
[30] L. J. González-Soler, M. Gomez-Barrero, and C. Busch, "On the generalisation capabilities of Fisher vector-based face presentation attack detection," *IET Biometrics*, vol. 10, no. 5, pp. 480–496, Sep. 2021.
[31] A. George and S. Marcel, "Deep pixel-wise binary supervision for face presentation attack detection," in *Proc. Int. Conf. Biometrics (ICB)*, Jun. 2019, pp. 1–8.
[32] A. George and S. Marcel, "On the effectiveness of vision transformers for zero-shot face anti-spoofing," in *Proc. IEEE Int. Joint Conf. Biometrics (IJCB)*, Aug. 2021, pp. 1–8.
[33] B. Yu, J. Lu, X. Li, and J. Zhou, "Salience-aware face presentation attack detection via deep reinforcement learning," *IEEE Trans. Inf. Forensics Security*, vol. 17, pp. 413–427, 2022.
[34] J. Yang, Z. Lei, and S. Z. Li, "Learn convolutional neural network for face anti-spoofing," 2014, *arXiv:1408.5601*.
[35] Y. Jia, E. Shelhamer, J. Donahue, S. Karayev, J. Long, R. Girshick, S. Guadarrama, and T. Darrell, "Caffe: Convolutional architecture for fast feature embedding," in *Proc. 22nd ACM Int. Conf. Multimedia*, Nov. 2014, pp. 675–678.
[36] O. Parkhi, A. Vedaldi, and A. Zisserman, "Deep face recognition," in *Proc. Brit. Mach. Vis. Conf. (BMVC)*. Durham, U.K.: British Machine Vision Association, 2015, pp. 1–12.
[37] Z. Xu, S. Li, and W. Deng, "Learning temporal features using LSTM-CNN architecture for face anti-spoofing," in *Proc. 3rd IAPR Asian Conf. Pattern Recognit. (ACPR)*, Nov. 2015, pp. 141–145.
[38] N. Sanghvi, S. Kumar Singh, A. Agarwal, M. Vatsa, and R. Singh, "MixNet for generalized face presentation attack detection," in *Proc. 25th Int. Conf. Pattern Recognit. (ICPR)*, Jan. 2021, pp. 5511–5518.
[39] S. Chen, T. Yao, K. Zhang, Y. Chen, K. Sun, S. Ding, J. Li, F. Huang, and R. Ji, "A dual-stream framework for 3D mask face presentation attack detection," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. Workshops (ICCVW)*, Oct. 2021, pp. 834–841.
[40] A. Liu, C. Zhao, Z. Yu, J. Wan, and A. Su, "Contrastive context-aware learning for 3D high-fidelity mask face presentation attack detection," *IEEE Trans. Inf. Forensics Security*, vol. 17, pp. 2497–2507, 2022.
[41] J. Yang, Z. Lei, D. Yi, and S. Z. Li, "Person-specific face antispoofing with subject domain adaptation," *IEEE Trans. Inf. Forensics Security*, vol. 10, no. 4, pp. 797–809, Apr. 2015.
[42] T. D. F. Pereira, "Learning how to recognize faces in heterogeneous environments," EPFL, IEL, Lausanne, Swiss, pp. 187, 2019. [Online]. Available: http://infoscience.epfl.ch/record/265913 and https://doi.org/10.5075/epfl-thesis-9366
[43] G. Wang, H. Han, S. Shan, and X. Chen, "Unsupervised adversarial domain adaptation for cross-domain face presentation attack detection," *IEEE Trans. Inf. Forensics Security*, vol. 16, pp. 56–69, 2021.
[44] Z. Li, R. Cai, H. Li, K. Lam, Y. Hu, and A. C. Kot, "One-class knowledge distillation for face presentation attack detection," *IEEE Trans. Inf. Forensics Security*, vol. 17, pp. 2137–2150, 2022.
[45] M. Fang, F. Boutros, A. Kuijper, and N. Damer, "Partial attack supervision and regional weighted inference for masked face presentation attack detection," in *Proc. 16th IEEE Int. Conf. Autom. Face Gesture Recognit. (FG)*, Dec. 2021, pp. 1–8.
[46] D. E. King, "Dlib-ML: A machine learning toolkit," *J. Mach. Learn. Res.*, vol. 10, pp. 1755–1758, Jan. 2009.
[47] A. Costa-Pazo, S. Bhattacharjee, E. Vazquez-Fernandez, and S. Marcel, "The replay-mobile face presentation-attack database," in *Proc. Int. Conf. Biometrics Special Interest Group (BIOSIG)*, Sep. 2016, pp. 1–7.
[48] A. Krizhevsky, I. Sutskever, and G. Hinton, "ImageNet classification with deep convolutional neural networks," in *Proc. Adv. Neural Inf. Process. Syst. (NeurIPS)*, vol. 25, 2012, pp. 1–9.
[49] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 2261–2269.
[50] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 770–778.
[51] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, and L. Chen, "MobileNetV2: Inverted residuals and linear bottlenecks," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 4510–4520.
[52] M. Tan, B. Chen, R. Pang, V. Vasudevan, M. Sandler, A. Howard, and Q. V. Le, "MnasNet: Platform-aware neural architecture search for mobile," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 2815–2823.
[53] L. A. Shalabi, Z. Shaaban, and B. Kasasbeh, "Data mining: A preprocessing engine," *J. Comput. Sci.*, vol. 2, no. 9, pp. 735–739, Sep. 2006.
[54] Z. Boulkenafet, J. Komulainen, L. Li, X. Feng, and A. Hadid, "OULU-NPU: A mobile face presentation attack database with real-world variations," in *Proc. 12th IEEE Int. Conf. Autom. Face Gesture Recognit. (FG)*, May 2017, pp. 612–618.
[55] A. Paszke, S. Gross, S. Chintala, G. Chanan, E. Yang, Z. DeVito, Z. Lin, A. Desmaison, L. Antiga, and A. Lerer, "Automatic differentiation in PyTorch," in *Proc. NIPS-W*, 2017, pp. 1–4.
[56] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," 2014, *arXiv:1412.6980*.
[57] L. J. Gonzalez-Soler, M. Gomez-Barrero, and C. Busch, "Evaluating the sensitivity of face presentation attack detection techniques to images of varying resolutions," in *Proc. Norwegian Inf. Secur. Conf. (NISK)*, Nov. 2020, pp. 1–11.
[58] D. O. Roig, P. Drozdowski, C. Rathgeb, A. M. González, E. Garea-Llano, and C. Busch, "Iris recognition in visible wavelength: Impact and automated detection of glasses," in *Proc. 14th Int. Conf. Signal-Image Technol. Internet-Based Syst. (SITIS)*, Nov. 2018, pp. 542–546.
[59] Z. Yu, C. Zhao, Z. Wang, Y. Qin, Z. Su, X. Li, F. Zhou, and G. Zhao, "Searching central difference convolutional networks for face anti-spoofing," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2020, pp. 5294–5304.
[60] L. Stampler. (Jun. 2014). *Here's What Faces Would Look Like if They Were Perfectly Symmetrical*. Accessed: Feb. 2, 2023. [Online]. Available: https://time.com/2848303/heres-what-faces-would-look-like-if-they-were-perfectly-symmetrical/
[61] J. Sánchez, F. Perronnin, T. Mensink, and J. Verbeek, "Image classification with the Fisher vector: Theory and practice," *Int. J. Comput. Vis.*, vol. 105, no. 3, pp. 222–245, Dec. 2013.

LAZARO JANIER GONZALEZ-SOLER received the B.Sc. degree in mathematics and computer science from the University of Havana, in 2014, and the Ph.D. degree in applied computer science from Hochschule Darmstadt, in 2022. He is currently a Postdoctoral Researcher with the National Research Center for Applied Cybersecurity (ATHENE), da/sec Group, Hochschule Darmstadt, Germany. He has served as a reviewer for different JRC journals and conferences. He has been also granted several awards, such as the Best Paper Award at the International Workshop on Information Forensics and Security (WIFS), in 2021, and the best-performing algorithm in the LivDet, in 2019 and 2021. His research interests include improving the security of biometric systems through the development of biometric presentation attack detection (PAD) techniques and the development of algorithms for biometric recognition in forensic scenarios. He has also been actively involved in European projects, such as secure and privacy-friendly mobile authentication (RESPECT).

MARTA GOMEZ-BARRERO (Member, IEEE) received the M.Sc. degree in computer science and mathematics and the Ph.D. degree in electrical engineering from Universidad Autonoma de Madrid, Spain, in 2011 and 2016, respectively. She was a Postdoctoral Researcher with the National Research Center for Applied Cybersecurity (ATHENE), Hochschule Darmstadt, Germany, between 2016 and 2020. She is currently a Research Professor for IT-Security with a focus on biometric recognition at the Hochschule Ansbach, Germany. She is also the General Chair of the BIOSIG Conference and has served for several conferences (e.g., *IJCB*, IWBF, and EUSIPCO) and journals [e.g., IEEE Transactions on Information Forensics and Security, IEEE Transactions on Pattern Analysis and Machine Intelligence, IEEE Transactions on Biometrics, Identity and Behavior, *IET BMT*, and PR (Elsevier)]. Furthermore, she is the Co-Chair of the European Association for Biometrics Academic SIG, an Associate Editor of the *EURASIP Journal on Information Security* and *EURASIP Journal on Image and Video Processing*, a member of the IARP TC4 Conference Committee and the IEEE Biometrics Council Security and Privacy Technical Committee, and represents the German Institute for Standardisation (DIN) in ISO/IEC SC37 JTC1 SC37 on biometrics. She is also coordinates the ANR-DFG Project RESPECT and was actively involved in the EU projects SOTAMD, BATL, and TABULA RASA. She has coauthored more than 90 technical publications in the field of biometrics. Her current research interests include security and privacy evaluations of biometric systems (PAD, BTP). She has also received a number of distinctions, including the EAB European Biometric Industry Award, in 2015, the Best Ph.D. Thesis Award by Universidad Autonoma de Madrid, in 2015 and 2016, the Archimedes Award for Young Researchers from Spanish MECD, the Best Paper Award at WIFS 2021, Odyssey 2018, and ICB 2015, and the Best Poster Award at ICB 2013.

CHRISTOPH BUSCH (Senior Member, IEEE) is currently a member of the Norwegian University of Science and Technology (NTNU), Norway. He holds a joint appointment with Hochschule Darmstadt (HDA), Germany. He has been lecturing on biometric systems with DTU, Denmark, since 2007. On behalf of the German BSI, he has been the Coordinator for the project series BioIS, BioFace, BioFinger, BioKeyS Pilot-DB, KBEinweg, and NFIQ2.0. He was/is a Partner of the EU projects 3D-Face, FIDELITY, TURBINE, SOTAMD, RESPECT, TReSPsS, and iMARS. He is also a Principal Investigator with the German National Research Center for Applied Cybersecurity (ATHENE) and the Co-Founder of the European Association for Biometrics (EAB). He has coauthored more than 500 technical papers and has been a speaker at international conferences. He is a member of the editorial board of the *IET Journal on Biometrics* and formerly of the IEEE Transactions on Information Forensics and Security. Furthermore, he chairs the TeleTrusT biometrics working group and the German standardization body on biometrics and the Convenor of WG3 in ISO/IEC JTC1 SC37.