Evaluating the Integration of Morph Attack Detection in Automated Face Recognition Systems

Andrea Panzino Simone Maurizio la Cava Giulia Orrù
Gian Luca Marcialis
University of Cagliari, Piazza d'Armi I - 09123 Cagliari (Italy)
{andrea.panzino,simonem.lac,giulia.orru,marcialis}@unica.it

Abstract

Due to the possibility of automatically verifying an individual's identity by comparing his/her face with that present in a personal identification document, systems providing identification must be equipped with digital manipulation detectors. Morphed facial images can be considered a threat among other manipulations because they are visually indistinguishable from authentic facial photos. They can have characteristics of many possible subjects due to the nature of the attack. Thus, morphing attack detection methods (MADs) must be integrated into automated face recognition. Following the recent advances in MADs, we investigate their effectiveness by proposing an integrated system simulator of real application contexts, moving from known to never-seen-before attacks.

1. Introduction

The phenomenon of face spoofing has become increasingly important in biometrics and cybersecurity over the years. Advances in technology, especially in image and video manipulation, have seen the birth of phenomena such as deepfakes, face synthesis, and morphing techniques. The latter consists of gradually transforming a face image into another (Figure 1) and can be used for malicious purposes, for example, to deceive a face recognition system (FRS). In fact, it is possible to obtain false faces containing the characteristics of multiple real faces. The result of this operation can be maliciously exploited to share an identity document [8], as the face resulting from a well-made morphing process can be associated with all the contributing identities by a human operator and an automatic FRS [23]. The problem is even more evident if we consider that such a document can also be used by a terrorist to evade border control.

To stem this problem, the research community has recently increased the effort to create Morph Attack Detectors (MAD). These systems aim to detect whether facial images

Figure 1. Example of morphing between two image belonging to FRCG dataset [18].

are morphs and represent a potential aid in application contexts when paired with Face Recognition Systems (FRS), such as border controls. However, to our knowledge, although there are several platforms for analyzing the performances of the individual MADs [33] and FRSs [26], none of them extensively addresses the problem from the point of view of integrating these systems in a single device. To help in this analysis, a solution may refer to what is proposed for evaluating the integration between matching and presentation attack detection systems with other biometric traits [15]. That paper showed that it is possible to depict several application scenarios by only relying on the ROC curves of the individual systems when sequentially combined, as the process is normally intended: the authenticity of the digital data is verified before the personal verification stage [8]. Adapting the system reported in [15], we propose the first analysis concerning morphing and facial recognition by testing the embedded MAD-FRS systems obtained from four MADs and two FRSs. The results provide useful elements that can be exploited in the design phase of a robust integrated system, *i.e.*, with a low rate of morphing samples accepted and genuine samples rejected.

The paper is organized as follows. Section 2 reviews the current literature on morphing creation, its malicious usage, and its detection to motivate our contribution. Section 3 describes the model employed for the integration between FRSs and MADs. Section 4 describes the protocol used to

conduct our evaluation, while section 5 reports the obtained results. Finally, conclusions are drawn in section 6.

2. Morphing as Presentation Attack and Its Detection

To our knowledge, Ferrara et al. proposed the first work that academically explores the danger of the morphing process concerning the possible creation of presentation attacks (PAs) [8]. Specifically, they describe how the morphing process can be exploited to create a PA relating to automatic face recognition systems (FRS), also known as a morphing attack (MA). The authors suggest morphing on a travel document to evade automatic border controls (ABCs) [8]. This attack, which involves an accomplice, exploits the possibility of sharing facial information by morphing two faces to deceive both automatic FRSs and border control operators. The process can, therefore, be described as follows. In the first instance, the accomplice initially submits the request to obtain a travel document using personal information. Subsequently, when the competent authorities request a valid passport photo, the accomplice sends a photo obtained by morphing his face with the suspect's. If the process is successful, the offices issue the accomplice a travel document with a photo given by the morph.

Moreover, due to morphs' distinctive feature lies in associating a face with each of the subjects morphed in the final image, an identity document containing a morphed photo could allow it to be shared between the accomplice and the suspect. In particular, it is possible that, during identity control, the accomplice's identity could be mistakenly associated with that of the suspect. This can happen whether the check is carried out by an automatic facial recognition system or a human operator since even a well-trained operator struggles to recognize a morph from a genuine photo, especially if the latter appears to be of good quality [23].

The issue of MAs also exists in other application scenarios, such as unauthorized entrance into a territory [27] and identity checks not strictly related to traveling cases. For example, one could use an identity document embedding morphed images to impersonate another individual during routine identity checks.

In parallel with the development of systems to counter this attack, the research community has proposed methods and models for creating increasingly higher-quality morphs, aiming to train ever better systems capable of recognizing them. Therefore, in addition to the detection methods described in Subsection 2.2, we provide a brief overview of the morph generation methods in Subsection 2.1, to facilitate the contextualization of the analysis proposed in this paper.

2.1. Morph generation techniques

Although the research community proposed various techniques for generating facial morphs, most of them could be summarized into two main approaches, namely approaches based on landmarks and those based on deep learning.

In particular, the first class of techniques combines facial landmark detection [4] with geometric transformations to combine the source faces. The identification of landmarks allows the extraction of spatial and geometric facial information, which can then be used for an alignment phase between the contributing faces, implemented through appropriate scaling and rotation operations. The alignment minimizes the relative distance between the corresponding facial landmarks, thus facilitating subsequent operations. After the alignment, it is possible to proceed with the actual deformation phase, which further minimizes the relative distance between the corresponding facial landmarks. Various methods have been proposed to perform such a deformation, such as interpolation [24]. Finally, the colors of the starting images are blended at the single pixel level through a generally linear combination of the intensities of the relative pixels of the images obtained in the previous step, taking into account the influence of each image in the color blending process [22]. Generally, this process is followed by a post-processing phase, which aims to eliminate or limit artifacts on the morphed image, not present on the contributing images, mainly due to the color deformation and blending processes. A possible solution is to apply morphing only in the area of one of the contributing faces and then use the related background through blending.FaceMorpher [1] and WebMorph [2] are two examples of this technique.

The second type of attack involves techniques based on deep learning for extracting facial information and morph synthesis. In particular, these techniques are generally based on Generative Adversarial Networks (GANs) [9]. To our knowledge, the first network used in the literature for generating morphs is the MorGAN [5], which adapts the BiGAN architecture [7] to generate low-resolution facial morphs. A significant improvement in the quality of the morphs was achieved through MIPGAN and MIPGAN-II [12], based on the StyleGAN and StyleGAN 2 architectures, respectively, which include a modification to the network architecture and the introduction of a new loss function, making the outputs dangerous for commercial recognition systems and those based on deep learning [35]. Despite the generally higher graphic quality of the images compared to landmark-based techniques and the absence of artefacts caused by their use, deep learning techniques can still suffer from visual deformations unique to these types of generative techniques, such as nose and eye distortions or unreal-

[1] https://github.com/alyssaq/face_morpher
[2] https://debruine.github.io/project/webmorph/

istic colour rendering.

2.2. Morph detection techniques

Techniques for Morphing Attack Detection (MAD) can be divided into two main categories: Single-Image MAD (S-MAD) and Differential MAD (D-MAD) [33].

The first group comprises all MAD approaches that do not employ reference images to determine if an image is morphed or bona-fide, but identifies specific artefacts within the image created by the morphing process. The S-MAD category may be further divided into several methodologies for artefact detection [33].

Textural algorithm analysis uses textual descriptors such as BSIF, LPB, or LPQ [20] to extract features from face images and use them for classification. The main disadvantage of this approach is related to the lack of robustness to the noise introduced in the image, for example, due to the scan process of a printed image, as in the case of the typical document check-in border control [26].

Algorithms that analyze the quality aim to detect image degradation due to possible morphing, for example, by identifying JPEG compression artefacts [33]. Another approach involves the variation analysis of the noise pattern, aiming to detect the spectral components' alteration caused by the morphing process [6]. However, the effectiveness of this technique is strictly dependent on the post-processing technique used. Some works remove the noise pattern from the image and analyze it for artifacts that may indicate a possible morphing process, such as discontinuities at the pixel level [32], while deep learning-based methods use neural networks as feature extractors [21]. Finally, hybrid methods use different feature extractors to classify morphed and bona-fide images to obtain better performance at the cost of a greater computational load [33].

Despite the evolution of the S-MAD techniques in terms of performance, the lack of a reference image represents a limitation for this approach, as the artifacts can either be caused by other processes independent from the morphing one [13] or be mitigated in the post-processing phase. This problem is partially overcome by using D-MAD techniques, which also involve a reference image in the classification process to compare with the image that must be analyzed directly. Therefore, these techniques are based on the difference between the features extracted from the two images. For example, these features can be extracted and compared using Siamese network architectures [30].

While strategies and systems to identify this attack constantly evolve, new methods for producing more realistic morphs to defeat automatic systems are also being developed. However, in this arms race, we often forget the big picture: if attacks aim to overcome MAD systems while preserving their ability to overcome facial recognition systems, what will be the effect of integrating them in a specific scenario? Furthermore, will the best MAD system in a given context also provide the best system integrated with a facial recognizer? Or could a system created by combining it with a slightly less performing MAD be better overall?

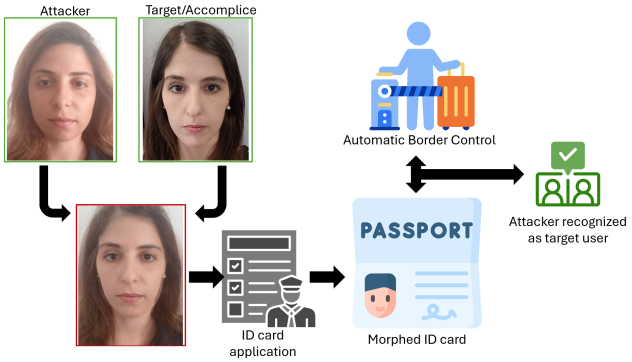

Figure 2. Use case of an integrated system between MAD and FRS: at ABCs the presented document ID may not correspond to the user (zero-effort impostor) or contain a morphed image.

In this context, we propose using a framework that ensures that designers can answer these questions to allow an optimal choice of the MAD and FRS pair concerning the target application scenario.

3. Simulation of MAD embedding in a FRS

The main problem of integration between an FRS and a MAD lies in the difficulty of empirically testing and evaluating these systems under various conditions, especially when dealing with sophisticated morphing attacks that are not well-represented in SOTA datasets or entirely unknown. This motivates the need for a specific protocol that is able to provide the performance of embedded systems without the practical difficulties stated above.

Based on the findings of Micheletto et al. [15], who proposed a simulator, called BIOWISE, that allows the study and design of the fusion of Presentation Attack (PA) detectors into fingerprint verification systems, we explore whether this type of simulation can be transposed to the field of morphing detection. This simulator for the sequential combination of two non-zero error-free systems is a significant step forward in biometric security. The key advantages of BIOWISE include the ability to simulate integrated error rates, system design flexibility, and meta-design process support. Its application in the field of morphing detection would depict the potential effects of the sequential fusion on the overall performance while considering specific morphs and an estimated probability of being attacked.

When considering the application of this methodology to the MAD/FRS integration, it's crucial to recognize the distinct context and operational dynamics of morphing attacks compared to other PA types. The BIOWISE simu-

lation framework can be adapted to explore the fusion of MAD into verification systems, but it will need a few key adjustments and considerations. In the fingerprint case, the probes (bona-fide or attack presentations) are compared with a gallery of genuine templates. In the case of MAD, however, morphing is usually embedded in an identity document (Figure 2) to bypass an automatic FRS such as those present in ABCs.

Therefore, the morphing attack was created before a request for an identity document that a human operator validates. Subsequently, this document is used in an automatic FRS, which evaluates whether the presented document matches the presenter's true identity. Adapting the BIOWISE simulator to MAD scenarios involves redefining the probabilistic relationships to reflect the unique challenges of morphing attacks accurately. This includes modeling the probability of a morphed image bypassing the MAD system and the probability of successful authentication. In this use case, we use the image embedded in the identity document as a probe and the presenter's face as a single FRS template.

The performance metrics employed by the simulator are in accordance with the recent ISO/IEC 30107–3 standard [1]. In particular, for the MAD, we consider the Attack Presentation Classification Error Rate (APCER), that is, the portion of the morphed images incorrectly classified as bona-fide and the Bona-fide Presentation Classification Error Rate (BPCER), that is, the portion of the bona-fide images incorrectly classified as morphed. Then, in addition to the False Match Rate (FMR) and the False Non-Match Rate (FNMR), we consider the Impostor Attack Presentation Accept Rate (IAPAR), which for an integrated evaluation is defined as the rate of morph attacks that successfully pass the overall system's checks.

In this adaptation, the simulator takes the individual FRS and MAD ROCs as input to generate the ROC of their fusion, taking into account the prior probability of being attacked by morphs (w) and the specific operational point chosen for the MAD ($BPCER = p\%$ or $APCER = p\%$).

To motivate the usability of the simulator in the context of morph detection and face recognition, it is necessary to demonstrate whether the underlying algorithm can be employed in the addressed application context. Therefore, in the following subsections, we model the problem to demonstrate the validity of the choice before showing how the simulator can be implemented.

3.1. Problem Modeling

In the use case previously outlined, we can define G as the boolean event "the user is authorized", interpreted as the presented ID document really belonging to the presenter. Therefore, $\overline{G}$ indicates the opposite event.Secondly, let L be the boolean event "the input image is authentic", that is, the ID document image is not morphed, while $\overline{L}$ indicates that the image is morphed. We also indicate with $P(G)$ and $P(L)$ the corresponding probabilities so that $P(\overline{G}) = 1 - P(G)$ and $P(\overline{L}) = 1 - P(L)$.

According to this notation, there are four possible joint events:

- $\{L, G\}$ the input document ID image is not morphed and the user is authorized (bona-fide trial);
- $\{L, \overline{G}\}$ the input document ID image is not morphed and the user is unauthorized (zero-effort attack);
- $\{\overline{L}, \overline{G}\}$ the input document ID image is not morphed and the user is unauthorized (morphing attack);
- $\{\overline{L}, G\}$ the input document ID image is morphed and the user is authorized.

Note that this last event, which is not contemplated in standard presentation attacks, may be possible in the case of morphing if the attacker and target are accomplices and both use the document. However, when an authorized user uses a morphed image, he/she should still be considered unauthorized because the image is not real and, thus, does not represent a bona-fide trial.

Therefore, it is possible to observe that the relationship between L and G considered by Ref. [15] still holds in the case of morph detection. Hence, $G \subseteq L : L$ includes both bona-fide trials and zero-effort attacks. Consequently, it is possible to model the acceptance rates of a single FRS and a MAD by first defining two events driven by the outcome of such systems.

Considering the FRS access is granted to a user when the comparison score s_M between the document image and the user's face is over a given acceptance threshold s^*_M. Hence, it is possible to define this event as:

$$M = s_M > s^*_M \tag{1}$$

Similarly, the MAD gives the classification of a certain input sample as real or morph when the score s_F is over a certain threshold s^*_F:

$$F = s_F > s^*_F \tag{2}$$

Finally, on the basis of the previous definition, it is possible to represent each access trial error rate for the individual FRS, namely FNMR, FMR, and IAPAR, and for the individual MAD, namely BPCER and APCER, as:

$$FNMR(M) = 1 - P(M|G, L) \tag{3}$$

$$FMR(M) = P(M|\overline{G}, L) \tag{4}$$

$$IAPAR(M) = P(M|\overline{G}, \overline{L}) \tag{5}$$

$$BPCER(F) = 1 - P(F|L) \tag{6}$$

$$APCER(F) = P(F|\overline{L}) \quad (7)$$

3.2. The proposed evaluation approach

Given the explicit parallelism between the face morphing detection and fingerprint presentation attack detection, as well as the overlap in problem modeling between the related tasks, it is possible to adapt the BIOWISE simulator [15] for the analysis of the sequential integration between an FRS and a MAD.

Therefore, we can obtain the integrated evaluation error rates through the individual performances of the two modules and we can define the Global FMR, thus the FMR of the integrated systems, by introducing the prior probability of a presentation attack $w = P(\overline{L}, \overline{G})$:

$$GFMR(M, F) = FMR(M, F) \cdot (1 - w) + IAPAR(M, F) \cdot w \quad (8)$$

This error considers both the acceptance of zero-effort attacks and morphing attacks.

The ROC curve of the sequential system is derived by considering the individual ROCs of the MAD and the FRS. By acting on w and on the MAD's operational point $BPCER = p\%$ or $APCER = p\%$, we may depict several possible scenarios and evaluate the current state of MAD-FRS integration. Further performance metrics can be extracted from the sequential system, such as the Global Equal Error Rate (Global EER) and the Global Area Under the ROC Curve (Global AUC), obtained from the comparison between GFMR and (1-FNMR) as the acceptance threshold varies.

4. Experimental protocol

4.1. Datasets

We employed two well-known SOTA morphing datasets for the experiments: AMSL [16] and FRLL-Morphs [25, 27]. In particular, both AMSL and FRLL-Morphs are based on the Face Research Lab London set (FRLL) [3], from which frontal photos were used, both neutral and smiling poses (one photo of each type for each of the 102 subjects, respectively). Regarding AMSL, pairs of subjects were then selected to carry out a total of 2175 morphs obtained through the method depicted in [16]. In contrast, FRLL-Morphs employs four different techniques for the generation of morphs: StyleGAN 2 [11], OpenCV [2], FaceMorpher, and WebMorph. In this case, for each technique, about 1221 were finally generated.

4.2. Systems employed

Regarding MAD systems, the tests were carried out mainly on four different algorithms: texture- and deep-learning-based. Regarding the texture-based algorithms, two different SVM classifiers with a radial basis function kernel were employed, one based on BSIF features (3×3 filter size; code length equal to 8 bit) [22] and the other on LBP features (3 points; radius equal to 5). We used the Platt scaling [19] to the output scores from the SVM classifier.

[3] https://doi.org/10.6084/m9.figshare.5047666.v5

In addition to the texture-based methods, we also used two deep learning-based methods: AlexNet [14] and VGG19 [29] architectures, which are adapted to our morph detection task through the employment of transfer learning. The last layer is then trained using the Adam optimizer based on Cross-Entropy Loss for the AlexNet model and using the SGD optimizer for the VGG19 model.

As FRSs, we employed two common architectures: FaceNet [28] and VGGFace [17]. FaceNet aims to map a face image into a multidimensional Euclidean space. In this case, we employed a FaceNet model based on Inception-ResNet V1 architecture [31], pre-trained on the VGGFace2 dataset [3] and with a 512-element vector. VGGFace is an architectural model based on VGG16 [29] and pre-trained on two face recognition datasets: LFW [34] and YouTube Faces [10].

4.3. Evaluation protocol

The tests were carried out in two steps: single-system and integrated-system evaluation. The purpose of single system evaluation is to assess the performances of MAD and FRS systems operating individually before employing the proposed integration evaluation. In this regard, we divided the users of the datasets into a pool for training the MAD and another pool for testing the systems.

Concerning the MAD training procedure, we select a pool of 62 subjects out of the 102 available. For each selected subject, both neutral and smiling images were chosen as bona-fide samples, resulting in 124 samples. The morphs were obtained by randomly sampling 128 morphed images composed only of the subjects in the training pool.

A different pool of smiling photos from 20 subjects was selected as the FRS gallery. In our use case, the FRS gallery corresponds to the set of users who physically present themselves at the border control.

Finally, the test set comprises the 20 users present in the gallery and other 20 users.

In fact, to test the integrated system, we need three different types of comparisons: bona-fide, zero-effort impostor, and morph trails. The first consists of the 20 neutral images of the same subjects used for the gallery. The second set represents the zero-effort impostors and includes the remaining 20 subjects selected neither for training nor for the gallery. In this case, only the neutral images were selected. Finally, the last set consists of morphs obtained between genuine and zero-effort impostors. All evaluations, both on single and integrated systems, were carried out using a k-fold cross-validation with a k-value equal to 10. In summary, for each fold, we have the following comparisons:

- 20 bona-fide comparisons, used to calculate the BPCER of the MAD and the FNMR of the integrated system;
- 400 zero-effort comparisons, used to calculate the BPCER of the MAD and the FMR of the integrated system;
- about 4000 morph comparisons, used to calculate the APCER of the MAD and the IAPAR of the integrated system (the number of morphs varies based on the subjects included in the test for the specific fold).

The morph comparisons for the FRLL-Morphs dataset are separated according to the kind of alterations, which enables us to assess how well each type breaches the MAD and the integrated

FRS	Dataset	FMR at FNMR=1%	FNMR at FMR=1%	IAPAR at FNMR=1%	FNMR at IAPAR=1%
FaceNet	FRLL	0.03	0.00	1.92	2.75
	AMSL	0.03	0.00	5.14	4.50
VGGFace	FRLL	0.52	0.02	2.72	3.50
	AMSL	0.50	0.00	3.64	8.00

Table 1. FRSs baseline performance on FRLL and AMSL datasets.

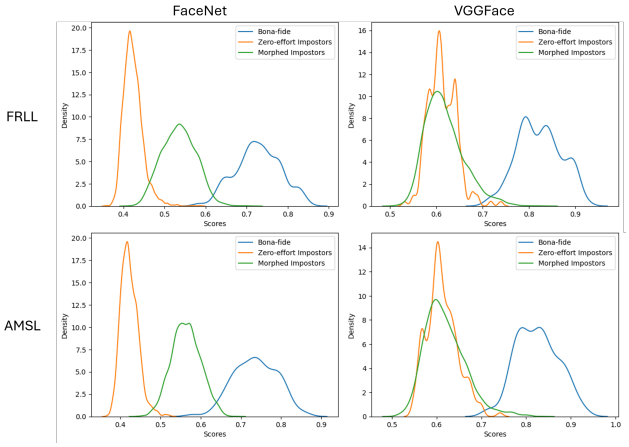

Figure 3. FaceNet and VggFace (FRSs) score distributions.

system. This also allows us to implement two types of protocols: (i) intra-manipulation, where the MAD is trained and tested on all types of manipulations (ii) cross-manipulation, where the MAD is trained on one type and tested on the others.

5. Results

In this section, we present and discuss the results obtained from the previously described experiments. In particular, in Subsection 5.1, we briefly report the results obtained by the FRSs. In Subsection 5.2, we analyze the capabilities of MADs in recognizing morph attacks. Finally, in Subsection 5.3, we discuss the effects of integration between FRSs and MADs in overall performance.

5.1. Face Recognition Performance

As it is possible to observe from the results of the FRS modules (Table 1), the introduction of morphs leads to a decay in the performance of the analyzed FRSs with respect to the recognition scenario with only zero-effort impostors. In particular, this is caused by a tendency of the systems to provide higher scores in the comparisons between bona-fide users and morphs created by including their identity in the morphing process or to a greater dispersion of the related distribution with respect to the distribution of scores related to comparisons with zero-effort impostors (Figure 3). Hence, these results confirm that morphs could represent an issue in face recognition and, thus, that sensible application scenarios like border control could potentially benefit from the introduction of MADs in automated face recognition processes.

5.2. MAD performance

The analyzed MADs show relevant differences in performance depending on the employed morph attack (Table 2). According to the analyzed performance metrics, it is possible to observe that, on average, the analyzed MADs show the worst detection capability on the AMSL dataset. This result seems reasonable as morphs in the AMSL dataset have a high visual quality and low presence of artifacts. Moreover, it is clear from the intra-manipulation results presented in Table 2 that handcrafted approaches perform significantly worse than deep learning-based methods. However, none of the MADs is the best on all the morph attacks, highlighting a certain degree of complementarity between them.

This complementarity is also confirmed through the cross-manipulation protocol, highlighting different generalization capabilities between the analyzed MADs (Figure 4). In particular, the investigation of strict operational points (i.e., APCER at BPCER=1% and BPCER at APCER=1%) highlights that the two most performing MADs are BSIF and VGG19, on average. Interestingly, considering the single cases, it is possible to observe that the first detector is less robust whenever morphs generated through OpenCV are included either in the training set or in the test set, coherently with the intra-manipulation results. A less clear pattern is instead shown in the case of results on VGG19, still suggesting less robustness in cross-manipulation involving StyleGAN in either training or test set and WebMorph in the only test set.

5.3. Performance after integration

The following analysis aims to point out the BIOWISE simulation reliability in predicting a real MAD/FRS sequential system and subsequently evaluate, considering the margin of error introduced by the simulation, how the designer should interpret it. For all the chosen FRSs and MADs, we investigated the difference between the IAPAR obtained from the integrated systems through the standard design approach (i.e., standard IAPAR) and that estimated after the simulation (i.e., estimated IAPAR) at APCER=1% (Figure 5). In particular, it is possible to observe that the simulation tends to underestimate the value of the IAPAR, with a more marked difference for lower values of FNMR. Still, this ΔIAPAR (between estimated and standard values) tends to reduce for higher values of FNMR until the overlap between the standard IAPAR curve and the estimated one is reached. Hence, these observations suggest that the simulation could be considered more reliable with higher values of the FNMR. This is expected since, as the FNMR increases, the system becomes more conservative in accepting samples by increasing the acceptance threshold. This increased threshold, resulting in a higher rejection rate, naturally limits the opportunities for impostor attempts to be accepted, potentially making the simulation's IAPAR estimates more accurate. However, in real application scenarios, the FNMR value is kept low so if the simulator is used it is necessary to consider the FNMR range between 0 and 20% maximum which corresponds to an average ΔIAPAR of 6.99% (for the intra-manipulation protocol). Furthermore, comparing the results in the two datasets, it is possible to observe that the results tend to be less reliable in the AMSL dataset. However, even in the case of AMSL, the simulation permits the comparison of the various combinations since the trend in the difference between the standard IAPAR and the estimated one is comparable for the different combinations of integrated systems (Figure 6).

The BIOWISE simulation results are reported in Tables 3,4 and 5. Considering the Global EER at various operating conditions and attack probability, the results from the intra-manipulation protocol

	FRLL-Morphs OpenCV (cv)		FRLL-Morphs FaceMorpher (fm)		FRLL-Morphs StyleGan2 (sg)		FRLL-Morphs WebMorph (wm)		AMSL	
	EER [%]	**AUC [%]**	**EER [%]**	**AUC [%]**	**EER [%]**	**AUC [%]**	**EER [%]**	**AUC [%]**	**EER [%]**	**AUC [%]**
AlexNet	11.99	90.63	11.50	90.65	14.97	89.19	13.37	87.62	13.15	87.79
LBP	27.55	81.73	23.24	86.75	11.62	95.10	34.62	68.57	27.94	79.56
BSIF	11.93	94.29	0.57	99.97	2.69	99.50	0.12	99.98	26.84	80.13
VGG19	2.82	99.48	3.14	99.36	2.19	99.71	2.55	99.76	10.31	94.17

Table 2. Results obtained from MADs on the FRLL and AMSL datasets with the intra-manipulation protocol.

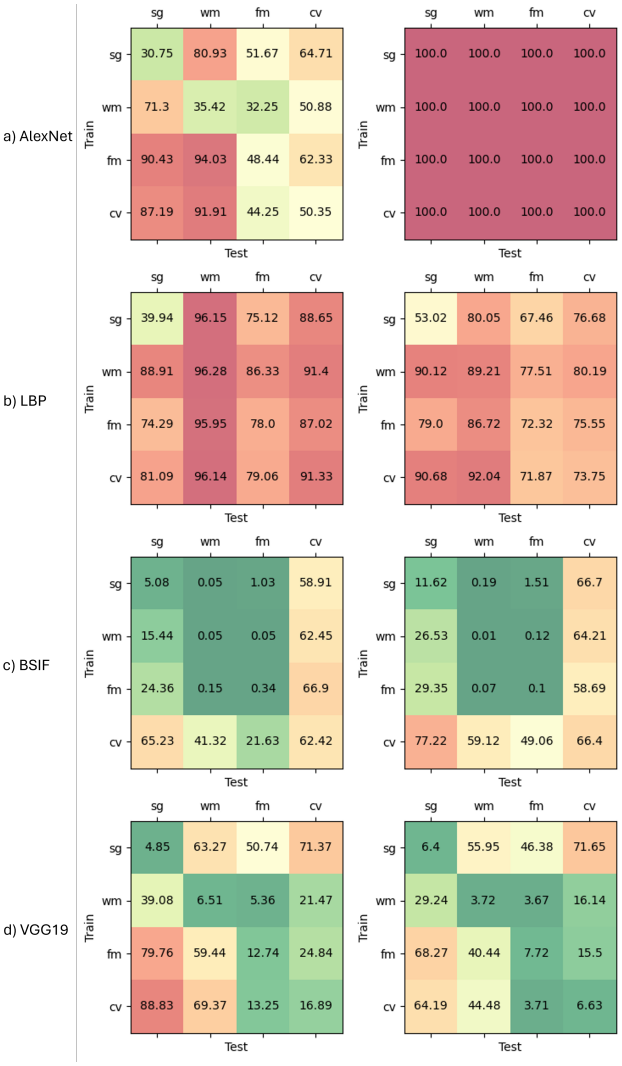

Figure 4. *APCER at BPCER=1% (first column) and BPCER at APCER=1% (second column) obtained with the cross-manipulation protocol by employing AlexNet (a), LBP (b), BSIF (c), and VGG19 (d) as MADs.*

on the FRLL dataset revealed that the FRSs could benefit more from the integration of BSIF and VGG19 than other MADS, especially when considering extremely conservative operational points in terms of admissible unrecognized attacks. Therefore, this outcome confirms the previous observations from the analysis of the

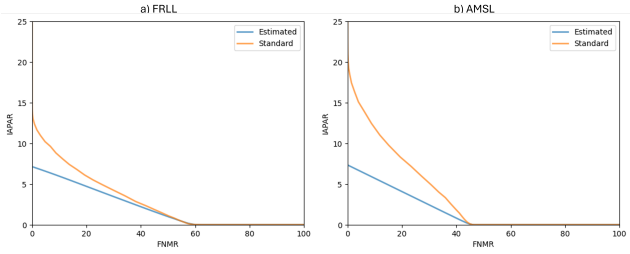

Figure 5. *Standard and estimated average IAPAR in intra-manipulation experiments from the possible combination of FRSs and MADs in FRLL (a) and AMSL (b).*

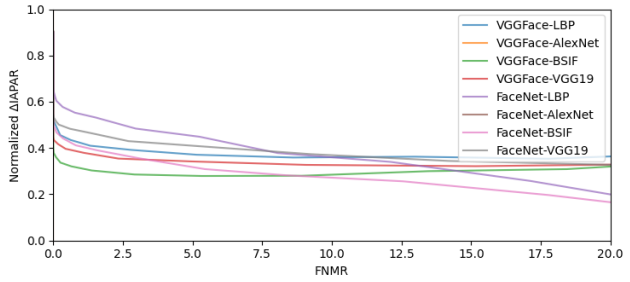

Figure 6. *Normalized difference between standard and estimated IAPAR ($\frac{Standard\,IAPAR - Estimated\,IAPAR}{Standard\,IAPAR}$) in intra-manipulation experiments from the possible combination of FRSs and MADs in AMSL.*

	APCER	AlexNet	BSIF	LBP	VGG19
FaceNet	1%	16.00 ± 0.00	3.95 ± 0.04	15.98 ± 0.36	4.90 ± 0.06
	10%	16.60 ± 0.00	1.97 ± 0.08	6.93 ± 0.04	1.97 ± 0.07
	20%	3.05 ± 0.05	1.94 ± 0.06	5.01 ± 0.06	1.96 ± 0.06
VGGFace	1%	16.60 ± 0.00	15.16 ± 0.39	16.44 ± 0.09	15.45 ± 0.48
	10%	16.60 ± 0.00	14.98 ± 0.38	15.73 ± 0.20	15.03 ± 0.35
	20%	15.39 ± 0.30	14.99 ± 0.38	15.52 ± 0.26	15.09 ± 0.32

Table 3. Global Equal Error Rate [%] obtained from integration on the FRLL dataset with the intra-manipulation protocol.

single MADs in intra-manipulation scenarios.

Considering the two FRSs, the integration of FaceNet highlighted the best performance. In particular, taking into account the integration with the most performing MADs, it provides similar results on the less challenging operational points (i.e., APCER=10% and APCER=20%) while revealing more significant differences at APCER=1%, with a difference of almost 1% of EER with respect to the integration with VGG19. Another interesting aspect is that these two combinations highlighted an EER that varies little with

	APCER	AlexNet	BSIF	LBP	VGG19
FaceNet	1%	50.00 ± 0.00	49.85 ± 0.08	49.82 ± 0.11	49.54 ± 0.27
	10%	50.00 ± 0.00	47.96 ± 1.18	48.06 ± 1.13	13.24 ± 0.12
	20%	6.00 ± 0.12	38.56 ± 0.09	30.93 ± 0.10	6.78 ± 0.11
VGGFace	1%	50.00 ± 0.00	49.85 ± 0.08	49.82 ± 0.11	49.54 ± 0.27
	10%	50.00 ± 0.00	47.97 ± 1.18	48.06 ± 1.13	46.51 ± 0.90
	20%	45.02 ± 0.78	48.46 ± 1.27	48.82 ± 1.09	45.19 ± 0.77

Table 4. Global Equal Error Rate [%] obtained from integration on the AMSL dataset with the intra-manipulation protocol.

	APCER	AlexNet	BSIF	LBP	VGG19
FaceNet	1%	4.55 ± 0.00	0.53 ± 0.01	4.49 ± 0.03	3.54 ± 0.00
	10%	4.55 ± 0.00	0.31 ± 0.01	4.41 ± 0.08	1.15 ± 0.00
	20%	2.25 ± 0.00	0.24 ± 0.01	4.23 ± 0.00	0.53 ± 0.00
VGGFace	1%	4.55 ± 0.00	4.27 ± 0.10	4.89 ± 0.03	4.44 ± 0.15
	10%	4.55 ± 0.00	4.27 ± 0.11	4.42 ± 0.08	4.44 ± 0.11
	20%	4.50 ± 0.12	4.26 ± 0.11	4.45 ± 0.13	4.36 ± 0.10

Table 5. Global Equal Error Rate [%] obtained from integration on the FRLL dataset with the cross-manipulation protocol.

respect to the attack probability. From a designer's perspective, this could be advantageous whenever a system is required to be employed in application scenarios where it is difficult to estimate the actual attack probability.

The results on AMSL (Table 4) revealed instead that none of the investigated combinations are able to provide satisfactory performance, with an EER lower than 10% only at APCER=1% when combining FaceNet with either AlexNet or VGG19. This highlights that if single MADs lead to a high EER, integration with an FRS is not recommended.

Unexpectedly, the cross-manipulation results (Table 5) highlighted an improvement in average performance concerning the corresponding intra-manipulation results. This behaviour of the integrated systems can be explained by the analysis of the individual systems: while the MADs often have a higher misclassification rate for more realistic morphs (e.g. without artefacts, as for the GAN-generated morphs), FRSs tend to make fewer errors with this type of sample. This discrepancy suggests that for a morph to appear more realistic and thus more challenging for MAD systems to detect, it must incorporate changes that make it diverge from the original contributing identities. For this reason, the integration of the two systems tends to work better in a cross-manipulation context: poorly elaborated morphs that maintain more information relating to individuals are blocked by the MAD (both if trained on GAN-generated samples and landmark-based) while morphs that are realistic but divergent from the original identities are blocked by the FRS.

From this analysis, some guidelines for designers of integrated systems can be extracted:

- It is critical to include a wide range of samples in the MAD training process, even those that may appear visually unrealistic. Although the human operator has no difficulty in identifying them as fakes, the FRS may be more affected by these types of attacks which keep the information of the contributors unchanged and it is therefore essential that the MAD blocks this type of attacks.
- The most effective integrated system may not necessarily be comprised of the highest-performing MAD and FRS when considered separately. Instead, the emphasis should be on how effectively these components work together. Simulations, such as those provided by BIOWISE, can be useful in determining which combinations of MAD and FRS provide the most compatibility and complementing advantages.
- The results revealed that some MADs are not ready to be integrated with an FRS due to their high APCERs. This highlights the urgent need for continued research to develop MAD techniques that succeed at generalization across never-seen-before attacks.

6. Conclusions

In this paper, we proposed a novel approach for evaluating the effectiveness of the employment of MADs in face recognition applications, allowing an in-depth analysis of the performance obtained after integrating a MAD and an FRS.

Although the simulation leads to an underestimation of the errors of the integrated system, its primary utility lies not in exact precision but in identifying potential trends. Extensive experiments on two FRSs and four MADs and the evaluation of the effectiveness of the related integrated systems confirmed the validity of the estimate, emphasizing that the best among the integrated systems is not necessarily composed of the best MAD and the best FRS. Similarly, the results highlighted that the choices between the available MADs and FRSs should be driven by the required operating conditions, both in terms of attack probability and tolerated frequency of misclassifications on morphs or bona-fide images.

It must be remarked that the proposed evaluation approach is related to a sequential integration between the involved modules, although parallel integrations could be suitable as well, as highlighted by other biometric traits dealing with presentation attacks. However, the sequential approach represents a simple and flexible way of performing such a combination, and, to our knowledge, the integrated evaluation approach is still not present in the literature related to morphs.

Future studies should extend the analysis to other combinations between FRSs and MADS, also extending the investigation of further and more numerous datasets. Similarly, the complementarity between MAD approaches, revealed by the morph detection capability on the single morphing processes, should also be further investigated, introducing new morph generation processes and analyzing new MADs. Despite these limitations, through this contribution, we try to add a piece for a future and effective introduction of MADs in facial recognition processes to improve security in sensitive application scenarios like border control and searching for wanted persons.

Acknowledgment

This work is supported by the European Union – Next Generation EU under the PRIN 2022 PNRR project "BullyBuster 2 – the ongoing fight against bullying and cyberbullying with the help of artificial intelligence for the human wellbeing" (CUP: P2022K39K8).

References

[1] ISO/IEC 30107-3:2023(en), Information technology — Biometric presentation attack detection — Part 3: Testing and reporting. 4

[2] G. Bradski. The OpenCV Library. *Dr. Dobb's Journal of Software Tools*, 2000. 5

[3] Qiong Cao, Li Shen, Weidi Xie, Omkar M. Parkhi, and Andrew Zisserman. Vggface2: A dataset for recognising faces across pose and age. In *2018 13th IEEE International Conference on Automatic Face Gesture Recognition (FG 2018)*, pages 67–74, 2018. 5

[4] Oya Çeliktutan, Sezer Ulukaya, and Bülent Sankur. A comparative study of face landmarking techniques. *EURASIP Journal on Image and Video Processing*, 2013(1):13, 2013. 2

[5] Naser Damer, Alexandra Moseguí Saladié, Andreas Braun, and Arjan Kuijper. MorGAN: Recognition Vulnerability and Attack Detectability of Face Morphing Attacks Created by Generative Adversarial Network. In *2018 IEEE 9th International Conference on Biometrics Theory, Applications and Systems (BTAS)*, pages 1–10, 2018. 2

[6] Luca Debiasi, Ulrich Scherhag, Christian Rathgeb, Andreas Uhl, and Christoph Busch. PRNU-based detection of morphed face images. In *2018 International Workshop on Biometrics and Forensics (IWBF)*, pages 1–7, Sassari, 2018. IEEE. 3

[7] Jeff Donahue, Philipp Krähenbühl, and Trevor Darrell. Adversarial feature learning. In *International Conference on Learning Representations*, 2017. 2

[8] Matteo Ferrara, Annalisa Franco, and Davide Maltoni. The magic passport. In *IEEE International Joint Conference on Biometrics*, pages 1–7, 2014. 1, 2

[9] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative Adversarial Nets. In *Advances in Neural Information Processing Systems*. Curran Associates, Inc., 2014. 2

[10] Gary B. Huang, Marwan Mattar, Tamara Berg, and Eric Learned-Miller. Labeled Faces in the Wild: A Database forStudying Face Recognition in Unconstrained Environments. In *Workshop on Faces in 'Real-Life' Images: Detection, Alignment, and Recognition*, Marseille, France, 2008. Erik Learned-Miller and Andras Ferencz and Frédéric Jurie. 5

[11] Tero Karras, Samuli Laine, and Timo Aila. A Style-Based Generator Architecture for Generative Adversarial Networks. In *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 4396–4405, 2019. 5

[12] Tero Karras, Samuli Laine, Miika Aittala, Janne Hellsten, Jaakko Lehtinen, and Timo Aila. Analyzing and improving the image quality of StyleGAN. In *Proc. CVPR*, 2020. 2

[13] Christian Kraetzer, Andrey Makrushin, Jana Dittmann, and Mario Hildebrandt. Potential advantages and limitations of using information fusion in media forensics—a discussion on the example of detecting face morphing attacks. *EURASIP Journal on Information Security*, 2021(1):9, 2021. 3

[14] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. ImageNet Classification with Deep Convolutional Neural Networks. In *Advances in Neural Information Processing Systems*. Curran Associates, Inc., 2012. 5

[15] Marco Micheletto, Gian Luca Marcialis, Giulia Orrù, and Fabio Roli. Fingerprint Recognition With Embedded Presentation Attacks Detection: Are We Ready? *IEEE Transactions on Information Forensics and Security*, 16:5338–5351, 2021. 1, 3, 4, 5

[16] Tom Neubert, Andrey Makrushin, Mario Hildebrandt, Christian Kraetzer, and Jana Dittmann. Extended StirTrace benchmarking of biometric and forensic qualities of morphed face images. *IET Biometrics*, 7(4):325–332, 2018. 5

[17] O. M. Parkhi, A. Vedaldi, and A. Zisserman. Deep face recognition. In *British Machine Vision Conference*, 2015. 5

[18] P.J. Phillips, P.J. Flynn, T. Scruggs, K.W. Bowyer, Jin Chang, K. Hoffman, J. Marques, Jaesik Min, and W. Worek. Overview of the face recognition grand challenge. In *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*, pages 947–954 vol. 1, 2005. 1

[19] John Platt et al. Probabilistic outputs for support vector machines and comparisons to regularized likelihood methods. *Advances in large margin classifiers*, 10(3):61–74, 1999. 5

[20] R. Raghavendra, Kiran B. Raja, and Christoph Busch. Detecting morphed face images. In *2016 IEEE 8th International Conference on Biometrics Theory, Applications and Systems (BTAS)*, pages 1–7, Niagara Falls, NY, USA, 2016. IEEE. 3

[21] R. Raghavendra, Kiran B. Raja, Sushma Venkatesh, and Christoph Busch. Transferable Deep-CNN Features for Detecting Digital and Print-Scanned Morphed Face Images. In *2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pages 1822–1830, 2017. 3

[22] Christian Rathgeb, Ruben Tolosana, Ruben Vera-Rodriguez, and Christoph Busch, editors. *Handbook of Digital Face Manipulation and Detection: From DeepFakes to Morphing Attacks*. Springer International Publishing, Cham, 2022. 2, 5

[23] David J. Robertson, Robin S. S. Kramer, and A. Mike Burton. Fraudulent ID using face morphs: Experiments on human and automatic recognition. *PLOS ONE*, 12(3): e0173319, 2017. 1, 2

[24] D. Ruprecht and H. Muller. Image warping with scattered data interpolation. *IEEE Computer Graphics and Applications*, 15(2):37–43, 1995. 2

[25] Eklavya Sarkar, Pavel Korshunov, Laurent Colbois, and Sébastien Marcel. Vulnerability analysis of face morphing attacks from landmarks and generative adversarial networks. *arXiv preprint*, 2020. 5

[26] Ulrich Scherhag, R. Raghavendra, K. B. Raja, M. Gomez-Barrero, C. Rathgeb, and C. Busch. On the vulnerability of face recognition systems towards morphed face attacks. In *2017 5th International Workshop on Biometrics and Forensics (IWBF)*, pages 1–6, 2017. 1, 3

[27] Ulrich Scherhag, Christian Rathgeb, Johannes Merkle, and Christoph Busch. Deep face representations for differential morphing attack detection. *IEEE Transactions on Information Forensics and Security*, 15:3625–3639, 2020. 2, 5

[28] Florian Schroff, Dmitry Kalenichenko, and James Philbin. Facenet: A unified embedding for face recognition and clustering. In *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 815–823, 2015. 5

[29] K Simonyan and A Zisserman. Very deep convolutional networks for large-scale image recognition. pages 1–14. Computational and Biological Learning Society, 2015. 5

[30] Sobhan Soleymani, Baaria Chaudhary, Ali Dabouei, Jeremy Dawson, and Nasser M. Nasrabadi. Differential morphed face detection using deep siamese networks. In *Pattern Recognition. ICPR International Workshops and Challenges*, pages 560–572, Cham, 2021. Springer International Publishing. 3

[31] Christian Szegedy, Sergey Ioffe, Vincent Vanhoucke, and Alex Alemi. Inception-v4, inception-resnet and the impact of residual connections on learning. *arXiv preprint*, 2016. 5

[32] Sushma Venkatesh, Raghavendra Ramachandra, Kiran Raja, Luuk Spreeuwers, Raymond Veldhuis, and Christoph Busch. Detecting Morphed Face Attacks Using Residual Noise from Deep Multi-scale Context Aggregation Network. In *2020 IEEE Winter Conference on Applications of Computer Vision (WACV)*, pages 269–278, 2020. 3

[33] Sushma Venkatesh, Raghavendra Ramachandra, Kiran Raja, and Christoph Busch. Face morphing attack generation and detection: A comprehensive survey. *IEEE Transactions on Technology and Society*, 2(3):128–145, 2021. 1, 3

[34] Lior Wolf, Tal Hassner, and Itay Maoz. Face recognition in unconstrained videos with matched background similarity. In *CVPR 2011*, pages 529–534, 2011. 5

[35] Haoyu Zhang, Sushma Venkatesh, Raghavendra Ramachandra, Kiran Raja, Naser Damer, and Christoph Busch. Mipgan—generating strong and high quality morphing attacks using identity prior driven gan. *IEEE Transactions on Biometrics, Behavior, and Identity Science*, 3(3):365–383, 2021. 2