V-MAD: VIDEO-BASED MORPHING ATTACK DETECTION IN OPERATIONAL SCENARIOS

Guido Borghi, Annalisa Franco, Nicolò Di Domenico, Matteo Ferrara, Davide Maltoni
Department of Computer Science and Engineering
University of Bologna
{name.surname}@unibo.it

ABSTRACT

In response to the rising threat of the face morphing attack, this paper introduces and explores the potential of Video-based Morphing Attack Detection (V-MAD) systems in real-world operational scenarios. While current morphing attack detection methods primarily focus on a single or a pair of images, V-MAD is based on video sequences, exploiting the video streams often acquired by face verification tools available, for instance, at airport gates. Through this study, we show for the first time the advantages that the availability of multiple probe frames can bring to the morphing attack detection task, especially in scenarios where the quality of probe images is varied and might be affected, for instance, by pose or illumination variations. Experimental results on a real operational database demonstrate that video sequences represent valuable information for increasing the robustness and performance of morphing attack detection systems.

1 Introduction

In the last decades, the wide diffusion of Facial Recognition Systems (FRSs) [1] has significantly increased the demand for robust security measures to counter emerging threats, including those associated with the face morphing attack [2, 3] through which it is possible to create a sort of hybrid face with a double identity.

Current methods to counter this kind of attack are referred to as Morphing Attack Detection (MAD) systems [4, 5] and predominantly are focused on the analysis of the single document image, *i.e.,* Single-image Morphing Attack Detection (S-MAD) [6, 7] or pairs of images (the document and the live acquired ones), *i.e.,* Differential Morphing Attack Detection (D-MAD) [8, 9]. However, in real-world operational scenarios such as Automated Border Control (ABC) gates in international airports [10], many commercial FRS technologies often acquire video streams, providing a continuous sequence of frames [11]. This operational mode is indeed considered in the evaluation of FRS vulnerability to morphing attacks: the Morphing Attack Potential [12] metric is defined considering multiple verification attempts. On one hand, the presence of multiple frames could strengthen the morphing attack, increasing the probability of success, since a single match for a frame of the sequence might be sufficient to pass the verification check [13]. On the other hand, we believe that the use of multiple frames could be advantageous from the MAD task perspective and must be considered.

Therefore, in this paper, we introduce the **Video-based Morphing Attack Detection** (V-MAD) task, as an effective solution for adapting MAD algorithms to real-world operational scenarios, such as ABC gates in international airports, by leveraging multiple frames (see Fig. 1). Indeed, we consider the ability to exploit multiple frames is an opportunity to design more accurate and robust MAD systems, enabling for instance the possibility of discarding low-quality frames affected by uneven illumination or non-frontal pose that might harm traditional D-MAD approaches.

From a practical point of view, this study specifically focuses on investigating the effectiveness of providing multiple input frames in current state-of-the-art D-MAD algorithms, including frame-based quality scores and machine learning techniques. By analyzing potential advantages derived from the use of multiple frames over the conventional D-MAD approach, we aim to explore the feasibility and benefits of such an approach in practical scenarios.

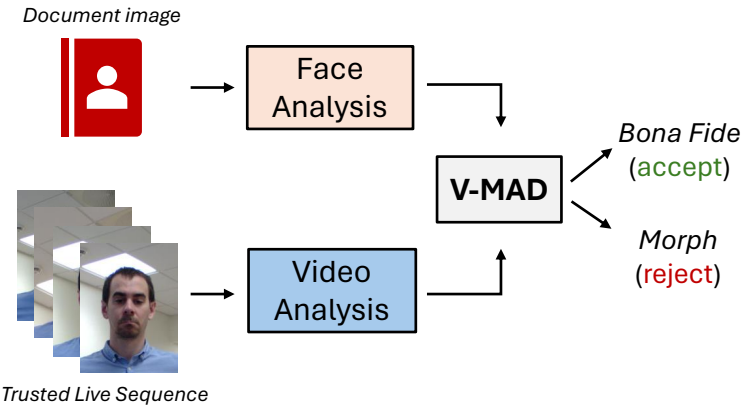


Figure 1: In operational scenarios, V-MAD represents a viable paradigm when a single probe document image is compared with an input sequence to detect whether the image is morphed or not. V-MAD differs from currently available literature solutions, based only on a single (S-MAD) or a pair (D-MAD) of images.

Summarizing, our investigation is organized in three sequential steps: i) we primarily focus on widely-used score-level fusion strategies across individual frames provided as input to different D-MAD systems [14, 8, 9]; then, ii) we analyze the usefulness of face image quality tools [15, 16, 17] as an additional input for V-MAD, assuming that image quality metrics can contribute to identifying the most reliable frame for analysis; finally, (iii) we exploit machine learning techniques to investigate the potential of artificial intelligence in this task.

By examining these strategies, our goal is to establish a foundation for understanding the potential benefits of leveraging video information in the context of the MAD task, particularly when compared to the classical D-MAD literature approaches.

2 Related Work

2.1 Face Morphing

Face Morphing is a method of manipulating images whereby one image gradually transforms into another. Within the context of electronic Machine-Readable Travel Documents (eMRTDs), this technique enables the creation of facial images that exhibit a double identity. Studies in the literature [2] indicate that morphed images have the capability to bypass both Commercial-Off-The-Shelf (COTS) FRSs and human controls, rendering face morphing a significant security threat. Furthermore, the proliferation of generative Artificial Intelligence techniques, such as Diffusion Models [18], Variational Autoencoders [19] and Generative Adversarial Networks (GANs) [20], greatly exacerbates this threat by simplifying the process for potential malicious actors. Additionally, morphed images can be enhanced through either manual or automated retouching procedures [21, 22], effectively eliminating both detectable and undetectable artifacts. Consequently, there is an urgent need to develop novel MAD systems capable of counterattacking new morphing algorithms and retouching methods. In this scenario, the introduction of the V-MAD paradigm can further improve the accuracy of existing MAD algorithms.

2.2 Morphing Attack Detection (MAD)

Since the introduction of the face morphing attack, several approaches have been proposed as potential countermeasures in the literature. A recent review of the existing methods is given in [23] and shows that the research is mainly focused on two different categories of approaches. The first one, named Single-image Morphing Attack Detection (S-MAD), relies on a single image, which is analyzed to point out any trace of a possible morphing process. These MAD systems are mainly designed to be exploited during the document enrollment stage, where the ID photo is analyzed for possible inclusion in the eMRTD. The second category is Differential Morphing Attack Detection (D-MAD) and includes systems that are supposed to be applied at the face verification stage, such as at airport ABC gates, where the ID photo stored in the document is compared to a trusted live capture acquired at the gate. In this case, two images are available and can be compared for MAD. It is worth noting that only D-MAD approaches can be included in the V-MAD scenario and therefore are introduced and analyzed in the following.

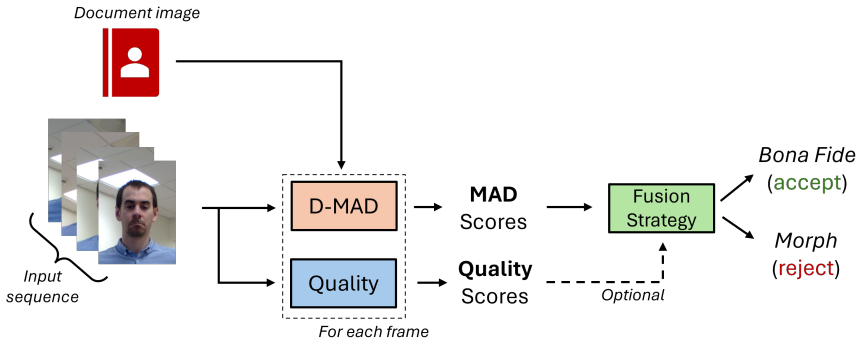


Figure 2: Practical implementation of the V-MAD task. Each frame of the input sequence is analyzed by the same D-MAD model, which receives also the document image as additional input and by a quality tool. Both output MAD and quality scores are then combined through a specific fusion strategy to produce in output the final single score.

D-MAD methods are based either on traditional computer vision and machine learning techniques or on deep-learning solutions. One of the current most accurate solutions is proposed in [8], in which two embeddings extracted through the ArcFace model [24] are classified by an SVM to decide if the input image is morphed or not. Since the embeddings are obtained using a model trained for the face recognition task, it is possible to infer that the classifier learns to detect the presence of morphing only using information about the subjects' identities.

The idea of exploiting not only information based on identity is proposed in [9]: specifically, the identity features are combined with features related to the presence of visible or not visible morphing traces (artifacts) in the document image. In this way, even if two similar subjects are provided as input, the morphing detection ability is preserved.

Other literature methods are not based on learning procedures. In particular, in [14] a reverse morphing procedure, referred to as demorphing, is used to unveil the genuine identity concealed within the morphed image. This approach is based on COTS FRSs; a key challenge arises from the non-linearity inherent in the morphing process, contrary to the linear combination assumed by the authors. Furthermore, the success of the entire process hinges on the accurate estimation of facial landmark positions, with even minor localization errors potentially compromising the efficacy of the entire pipeline.

3 Video-based MAD (V-MAD)

The typical identity verification process at border gates consists of comparing an ID photo d stored into an eMRTD to a sequence of n trusted live-captured frames $\mathbf{F} = (f_1, f_2, ..., f_n)$, acquired for face verification. Therefore, a theoretical V-MAD system $V(d, \mathbf{F})$ should analyze in input the whole sequence $\mathbf{F}$ and compare it to the document image d to provide in output a single score to indicate the morphing probability of the document image.

Being aware that this paper is a seminal work and no V-MAD methods are currently available in the literature, we focus our investigation on adapting D-MAD methods to the V-MAD task, as represented in Figure 2, to establish a foundation and guidelines for future V-MAD works.

Therefore, in our V-MAD implementation, we have a D-MAD system D able to compute a morphing score $D(d, f_i)$, representing the probability that the document image d is morphed, based on its comparison with a specific frame f_i. This can be repeated for each frame in the sequence of gate images $\mathbf{F}$, thus producing a sequence $S(d, \mathbf{F})$ of morphing scores:

$$S(d, \mathbf{F}) = (D(d, f_i), \ i = 1, \ldots, n) \tag{1}$$

For the V-MAD task, we consider a MAD system V able to compute a morphing score $V(d, \mathbf{F})$ based on the comparison of the document image d with the whole sequence of frames $\mathbf{F}$. In its simplest form, a V-MAD system can combine, through a function ϕ, the sequence $S(d, \mathbf{F})$ of D-MAD scores (see Eq. 1) computed on the individual frames $f_i \in \mathbf{F}$:

$$V(d, \mathbf{F}) = \phi(S(d, \mathbf{F})) \tag{2}$$

where ϕ is a function $\phi : \mathbb{R}^n \rightarrow \mathbb{R}$, *i.e.,* a function that takes as input a vector of n scores and produces as output a single score for a given sequence.

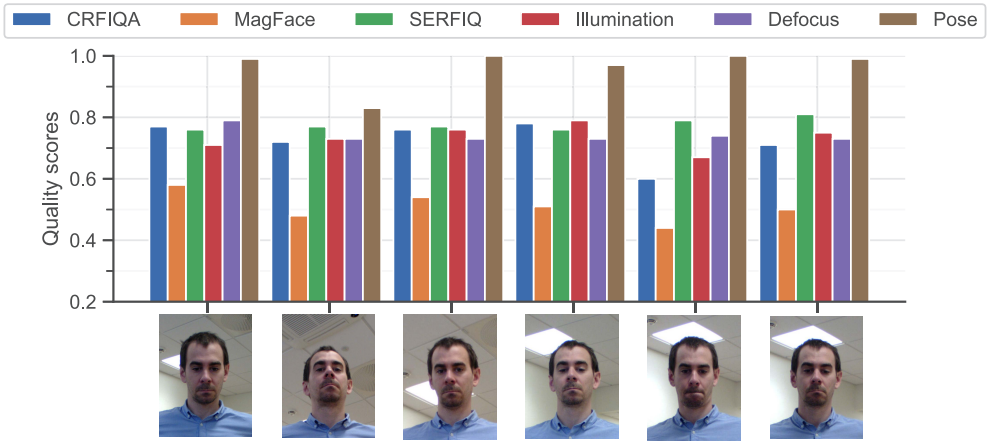


Figure 3: An example of the quality scores computed on a sequence of frames. As reported in Section 3.2, the first three methods (CRFIQA [17], MagFace [15] and SERFIQ [16]) are able to compute an overall quality score on the whole image, while the last three are based on specific aspects of the image.

The function ϕ, and then the V-MAD task, can be easily generalized to the case where multiple scores are available for each frame f_i:

$$V(d, \mathbf{F}) = \phi(S_k(d, \mathbf{F}),\ \ k = 1, \ldots, K) \tag{3}$$

where each S_k is a set of scores computed starting from the document image and the sequence of probe frames $\mathbf{F}$. Therefore, in this case, the application domain is $\phi : \mathbb{R}^{n \times K} \rightarrow \mathbb{R}$, *i.e.,* K scores available for each of the n frames are condensed in a single output value.

3.1 D-MAD Score fusion

Let's focus first on the case where the V-MAD score $V(d, \mathbf{F})$ for a given document image d and a sequence of frames $\mathbf{F}$ is defined as $S_D(d, \mathbf{F})$, applying a function ϕ to the D-MAD scores computed for every single frame $f_i \in \mathbf{F}$.

Then, we can define a variety of ϕ functions to produce in output a single score as follows:

- **Avg**: the average D-MAD score of the sequence $\mathbf{F}$

$$V\left(d, \mathbf{F}\right) = \frac{1}{n} \sum_{i=1}^{n} D\left(d, f_i\right) \tag{4}$$

- **Med**: the median D-MAD score of the sequence $\mathbf{F}$

$$V\left(d, \mathbf{F}\right) = \underset{i=1,..,n}{\text{med}}\ D\left(d, f_i\right) \tag{5}$$

- **Vote**: a voting system based on the computed D-MAD scores is defined as follows:

$$V\left(d, \mathbf{F}\right) = \frac{1}{n} \sum_{i=1}^{n} m\left(D\left(d, f_i\right)\right) \tag{6}$$

where

$$m(D(d, f_i)) = \begin{cases} 1 & \text{if } D(d, f_i) > thr \\ 0 & \text{otherwise} \end{cases} \tag{7}$$

in which thr is a decision threshold.

3.2 Incorporating Face Image Quality

A further contribution to the V-MAD task could come from the possibility of exploiting image quality metrics able to assign, for each frame of the probe image sequence $\mathbf{F}$, a quality score.

In this case, the input to the V-MAD model consists of two sets of scores, *i.e.* the D-MAD scores over the single frames ($S_D(d, \mathbf{F})$) and the set of quality scores of the probe frames in the sequence ($S_Q(\mathbf{F})$):

$$V(d, \mathbf{F}) = \phi(S_D(d, \mathbf{F}), S_Q(\mathbf{F})) \tag{8}$$

The document image could be considered as well, but since ID document images have to fulfill strict quality requirements [25], we expect its quality to be high and will not have a noticeable impact on MAD then we focus on gate images only. Two possible ϕ functions are considered to combine the single D-MAD scores $S_D = \{D(d, f_i),\ \forall f_i \in \mathbf{F}\}$ and the corresponding quality scores $S_Q = \{Q(f_i),\ \forall f_i \in \mathbf{F}\}$:

- **Weighted average**: the final V-MAD score is computed as the average of the D-MAD scores of each frame, weighted by the corresponding quality score:

$$V(d, \mathbf{F}) = \sum_{i=1}^{n} D(d, f_i) \cdot Q(f_i) \tag{9}$$

  where $Q(f_i)$ is the quality score assigned to the frame f_i by, for instance, a Face Image Quality Assessment Algorithm (FIQAA).
- **Best quality**: the V-MAD score is the D-MAD score computed from the frame with the highest quality score:

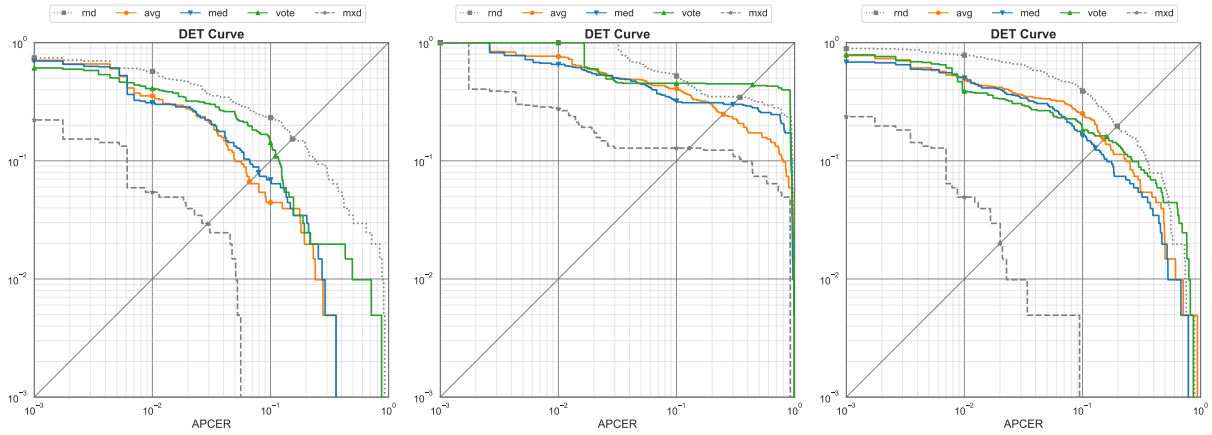
$$V(d, \mathbf{F}) = D(d, f_k) \text{ with } k = \underset{i=1,..,n}{\arg\max}\, Q(f_i) \tag{10}$$

Even in this case, several algorithms can be used for Face Image Quality Assessment (FIQA); a comprehensive review is available in the recent survey [26]. In general, face image quality can be assessed through a unified score, which takes into account different quality elements and summarizes them into a single value, or by analyzing single quality components related to specific image or face characteristics (e.g., illumination uniformity, blurring, head pose, etc.). Most of the unified FIQAAs exploit deep learning-based models and have been developed for a wide range of application scenarios where face images are typically acquired in an unconstrained environment and present therefore significant variations. From this category, we consider in particular:

- **MagFace** [15]: a framework proposed for both face recognition and FIQA. It proposes a set of loss functions that learn a universal feature embeddings capable of measuring face quality. Under this new representation, the authors show that the magnitude of the feature embeddings consistently increases for faces more likely to be recognized. MagFace also incorporates an adaptive mechanism to improve within-class feature distributions, ensuring easy samples are pulled closer to class centers while hard samples are pushed away.
- **CR-FIQA** [17]: a recent method that estimates the face image quality of a sample by learning to predict its relative classifiability, measured according to the allocation of the training sample feature representation in angular space with respect to its class center and the nearest negative class center. To predict the classifiability property of a facial image, the model is trained simultaneously with a face recognition model.
- **SER-FIQ** [16]: the quality assessment score is derived through an unsupervised methodology, relying on the comparative robustness of deeply learned embeddings of the image, rather than on predetermined ground truth acquired from human annotation or facial comparison scores, which may yield imprecise information. This approach analyzes the variability in embeddings generated from random subnetworks of a facial model to estimate the robustness of a sample's representation, and consequently, its quality.

In addition to the unified quality scores aimed at an overall evaluation of the face image quality, more specific measures can be used to analyze individual image characteristics. The ISO OFIQ [27] standard defines a number of quality components, also providing the guidelines for their computation. We selected here a set of quality components that might have a significant impact on MAD, computed through commercial tools:

- **Illumination uniformity**: it measures the difference in illumination on the left and right sides of the face. It is computed as the intersection of the normalized luminance histograms computed on the left and right parts of the face region, respectively.
- **Defocus**: it analyzes the level of sharpness. The score is computed as the difference between the face region image and the smoothed version of the same region obtained through a convolution of the image with a mean filter.
- **Pose**: it is focused on the analysis of whether the head pose is frontal. In our analysis, we take into account only the pitch angle since in real operational scenarios yaw and roll angles are mostly well controlled. Only limited variations in pitch might be observed due to the location of the acquisition device at the gate.



(a) Demorphing [14] (b) Siamese [9] (c) DFR [8]

Figure 4: V-MAD performance comparison of three D-MAD models using different MAD score fusion strategies (see Sect. 3.1). The dashed line represents the theoretical upper bound of performance, while the dotted line, based on a random choice in the score set, represents the lower bound. Metrics are expressed as errors, then lower are better.

An example of these quality scores computed on different frames of a real sequence is reported in Figure 3. A visual inspection of the frames suggests that the different quality measures are able to capture to some extent the differences in terms of quality between the images, by assigning lower quality scores when specific issues are visible.

4 Experimental evaluation

The experimental evaluation is organized in two sequential steps. Firstly, we investigate the impact of the described score fusion strategies (see Sect. 3.1) applied to scores produced by different D-MAD models. Then, we consider the scores produced by different quality tools (see Sect. 3.2) in combination with the previous D-MAD scores.

4.1 Database and evaluation protocol

Current publicly available datasets commonly used for the MAD task do not well represent the investigated operational scenario, since they do not contain probe sequences. Therefore, in this paper, a new database has been collected. Images are collected in six different locations, including two airports and four research laboratories, where images were acquired under real border control conditions using authentic ABC gates. A total of 60 different subjects have been involved in the acquisition and some of them have been acquired across multiple locations.

Summarizing, the database contains: i) 205 bona fide document images acquired in a capture setup, which meets the requirements for a document image in a passport application, ii) 612 gate images acquired live with real ABC gates, and iii) 1142 morphed document images created starting from the bona fide document images using 12 morphing algorithms and various morphing factors.

For the D-MAD task, bona fide document images are compared against gate images of the same subject (for a total of 2187 bona fide attempts) and morphed document images are compared against gate images of both contributing subjects (for a total of 34698 morphed attempts).

For the V-MAD task, bona fide document images are compared against gate sequences of the same subject (for a total of 125 bona fide attempts) and morphed document images are compared against gate sequences of both contributing subjects (for a total of 1145 morphed attempts).

4.2 Metrics

In the evaluation of the effectiveness of the V-MAD models, we utilize the common error-based metrics tailored to the MAD task as outlined in [28]. Specifically, we compute the Bona Fide Presentation Classification Error Rate (BPCER), that quantifies the ratio of authentic images erroneously classified as morphed. Conversely, the Attack Presentation Classification Error Rate (APCER) expresses the ratio of morphed images inaccurately identified as genuine. In the literature, BPCER is often analyzed in conjunction with a predetermined APCER threshold: in our experimental setup,

we investigate BPCER_{10} ($\mathbf{B}_{10}$), BPCER_{20} ($\mathbf{B}_{20}$) and BPCER_{100} ($\mathbf{B}_{100}$), denoting the lowest BPCER achievable at APCER values not exceeding 10%, 5% and 1%, respectively. It is noteworthy that the latter metric poses a particularly stringent benchmark and conventionally represents the standard operational point for facial verification systems in real-world applications [29]. These values are also visually represented through the Detection Error Trade-off (DET) curve, useful to directly compare different solutions at a glance.

Method	EER	$\mathbf{B}_{10}$	$\mathbf{B}_{20}$	$\mathbf{B}_{100}$
Siamese [9]	.392	.455	.575	1.00
DFR [8]	.221	.361	.486	.691
Demorphing [14]	.150	.205	.293	.501

Table 1: Performance of the three different models, on the database used for our evaluation protocol, for the D-MAD task.

4.3 Tested D-MAD models

Following the considerations reported in Section 3.1, we test different D-MAD methods to produce a MAD score for each frame of a sequence. Specifically, we test three recent D-MAD methods: our implementation of DFR [8] and Siamese [9], and the official implementation of the method Demorphing [14]. For each method, further details are reported in Section 2.2.

For the sake of completeness, we report in Table 1 the performance of each model on the evaluation database for the D-MAD task. It is important to note that these results are not directly comparable with the ones obtained in the V-MAD task, but they are useful to understand the performance of the single methods in the simple D-MAD task. From a general point of view, we observe that the Demorping method exhibits great performance, followed by the DFR (current state-of-the-art method on the BOEP platform), while the effectiveness of the Siamese approach seems to be limited in this scenario.

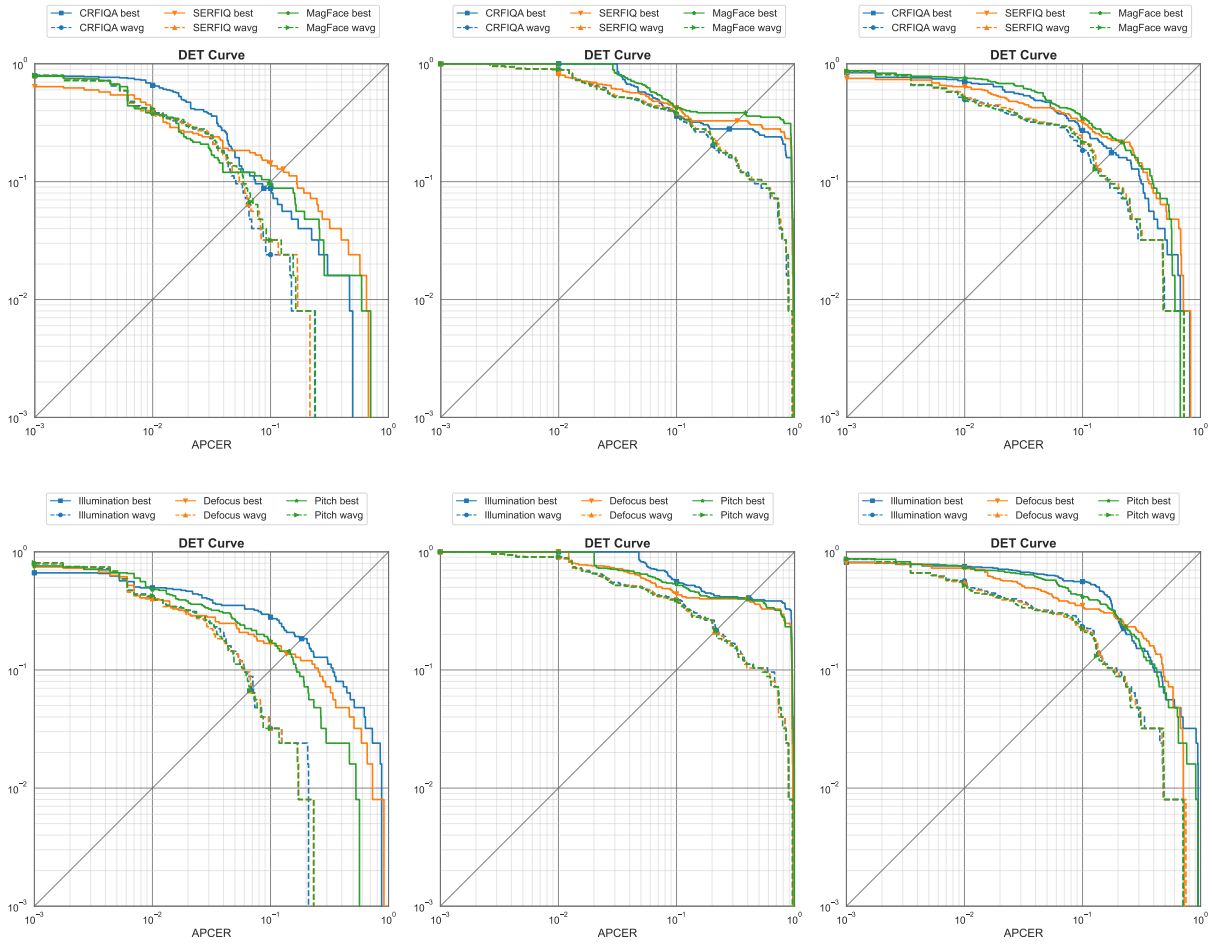
4.4 Experimental results

4.4.1 Evaluation of D-MAD score fusion strategies

The results obtained by applying different score fusion strategies to the scores produced by Demorphing [14], Siamese [9], and DFR [24] are reported in Figure 4. To better understand the range of the performance, we compute two different baselines. The first one, here referred to as "rnd" (gray dotted line) is based on the random choice of a single D-MAD score among those available in a given sequence: in this manner, we can compare the performance of each fusion strategy with the corresponding D-MAD approach since we have the same amount of scores for the V-MAD task. The second baseline, referred to as "mxd" (gray dashed line), simulates an oracle system able to choose for each attempt (either bona fide or morphed) the best possible score. Specifically, we select the minimum or the maximum scores in the given sequence relying on the ground truth annotation: the minimum for the bona fide sequences, and the maximum for the morphed ones. Then, this baseline reveals the theoretical best performance achievable with the scores produced by a specific D-MAD algorithm.

The analysis of the results highlights some important findings. Firstly, the main observation is that even a V-MAD system consisting of simple score fusion strategies outperforms the tested D-MAD approaches in most cases. In other words, in a real scenario, merging the D-MAD scores of multiple frames is better than computing the MAD score on a random frame of the acquired sequence. In particular, we note that the fusion strategies based on the average or median functions achieve great performance, while the voting system is negatively influenced by the different thresholds to be adopted to compute the votes. Indeed, the "avg" strategy allows to achieve an EER of 0.216, 0.136 and 0.066 for Siamese, DFR and Demorphing, respectively. Even if a direct comparison is not possible due to the different number of attempts, these results give us a clear perception of improvement over the D-MAD results given in Table 1.

A second important finding is that the "mxd" result reveals that theoretically, the V-MAD approach can significantly improve the MAD performance, reaching impressively low EER: 0.048 for Siamese, 0.016 for DFR, and 0.008 for Demorphing.



(a) Demorphing [14] (b) Siamese [9] (c) DFR [8]

Figure 5: DET curves for the different V-MAD approaches obtained by exploiting image quality estimated as either unified (first row) or specific (second row) quality measures.

4.4.2 Evaluation of the impact of image quality

Assuming that MAD scores could be estimated more reliably when the quality of the gate image is good [21], the second investigation analyzes the impact of combining MAD and quality scores obtained using the approaches described in Section 3.2. The results of this analysis, reported also in this case through the DET curves, is given in Figure 5.

For all the D-MAD methods tested, the results reveal a substantial homogeneity across the different quality models, in particular with respect to the EER. Moreover, it is possible to note that the weighted average fusion strategy, referred to as "wavg" (see Eq. 9) outperforms, even with a limited margin, the best quality strategy (see Eq. 10). To better appreciate the possible advantages deriving from the use of image quality scores, Table 2 compares the best results obtained with D-MAD morphing scores only to the best results obtained combining D-MAD and quality scores. For all the D-MAD systems, the introduction of quality has a positive effect since it allows to some extent to reduce the error rates. We only consider here the unified quality scores obtained with MagFace, SER-FIQ and CR-FIQA since they are more effective than single quality components according to the results of Figure 5. Among the three FIQAAs, CR-FIQA achieves the best results even if a comparable improvement is observed for the other FIQAAs as well.

4.4.3 Evaluation of the impact of Machine Learning

A further experiment has been carried out to evaluate if a machine learning approach can be effectively exploited as a fusion strategy. In other words, we investigate if a regressor that receives as input a sequence of MAD and quality scores is able to output a single V-MAD score.

Method	Quality	EER	B_{10}	B_{20}	B_{100}
Demorphing [14]	-	.066	.032	.120	.432
	CR-FIQA [17]	.065	**.024**	**.104**	.400
	SER-FIQ [16]	**.064**	.032	.120	**.392**
	MagFace [15]	.067	.032	.136	.416
Siamese [9]	-	.216	.392	.504	.904
	CR-FIQA [17]	**.203**	**.368**	**.488**	.896
	SER-FIQ [16]	.216	.384	**.488**	**.888**
	MagFace [15]	.210	.384	**.488**	.896
DFR [8]	-	.136	.224	.312	.520
	CR-FIQA [17]	**.127**	**.184**	**.304**	**.496**
	SER-FIQ [16]	.132	.216	.312	.520
	MagFace [15]	.128	.216	.312	.512

Table 2: Comparison of the V-MAD results without (-) and with the contribution of quality scores using the weighted average (wavg) method.

Specifically, we exploit an SVM regressor, with a radial basis function (RBF) kernel, the regularization parameter $C = 1.0$, and the kernel coefficient $\gamma = 10^{-3}$. In order to have all the sequences, and then the number of the features, of the same length, we empirically set the maximum number of frames to the average sequence length (50), padding with null element shorted sequences, and clipping longer ones. For each frame of the video sequence, we compute the MAD score as reported in Section 4.4.1, and quality scores as reported in Section 4.4.2. Then, both in the training and testing phases, all quality scores are normalized in the range $[0, 1]$: the scores produced by MagFace are divided by the median value of the entire set of scores of the dataset (25.77), while the illumination, defocus and pose scores are divided by 100 since the original score is in the range $[0, 100]$. The dataset is split putting the 50% of data in training and in the testing sets: therefore, it is important to note that the results of this machine learning approach are not directly comparable with the previous ones reported in Figures 4 and 5, since they are computed on a different portion of the dataset (and then a different number of comparisons). Different types of classifiers and other combinations have been tested: in the following, we report only the best combinations found based on the SVM model.

Experimental results are reported in Table 3. In particular, for each of the D-MAD models investigated, we report three different results. In the first line, we report the performance of the best fusion strategy outlined in the first part of our experiments, *i.e.,* the average (avg) and weighted average (wavg) strategies, then without any use of a machine learning approach. In the third line, we report the results obtained providing as input to the SVM only the MAD scores, while in the last line, there are the values obtained providing as input the concatenation of the MAD scores, the quality

Method	Input	EER	B_{10}	B_{20}	B_{100}
Demorphing [14]	*avg*	*.065*	*.016*	*.127*	*.381*
	wavg	*.064*	*.016*	*.111*	***.349***
	S_D	.063	.032	.079	.524
	$S_D + S_Q$	**.033**	**.000**	**.000**	.460
Siamese [9]	*avg*	*.209*	*.317*	*.460*	*.968*
	wavg	*.206*	*.302*	*.460*	*.968*
	S_D	.190	.222	.286	.968
	$S_D + S_Q$	**.127**	**.175**	**.254**	**.651**
DFR [8]	avg	*.128*	*.190*	*.238*	*.317*
	wavg	*.118*	*.190*	*.238*	*.309*
	S_D	.111	.111	.159	.206
	$S_D + S_Q$	**.079**	**.079**	**.095**	**.175**

Table 3: V-MAD results using a machine learning approach, providing as input the MAD (S_D) and quality (S_Q) scores.

scores produced through the MagFace method and the quality scores related to the illumination uniformity. Results suggest that machine learning is a viable approach to merging different scores for the V-MAD task. In particular, SVM overcomes approaches based only on the average and weighted average fusion strategy.

5 Conclusions

This study provides significant insights into the effectiveness and potential advantages of the novel Video-based Morphing Attack Detection (V-MAD) task compared to the traditional Differential Image-based (D-MAD) Morphing Attack Detection.

Firstly, we have demonstrated that incorporating information from multiple frames can lead to substantial improvements in overall performance. Utilizing video sequences enables the development of a MAD system more robust to the inherent variability characterizing face images due to several factors like illumination, pose changes, or motion blur. Even simple score fusion strategies applied to the D-MAD scores computed for the single frames proved to be effective.

Secondly, we proved that face image quality can further contribute to the development of robust V-MAD systems. Unified quality scores as well as single quality components allow to further improve the performance, especially when exploited by means of machine learning models able to combine D-MAD and quality scores into a single effective morphing score.

In conclusion, our study confirms that V-MAD represents a significant evolution from traditional MAD approaches, offering increased effectiveness and robustness in detecting face morphing attacks. Our analysis is a preliminary study aimed at assessing the theoretic feasibility of V-MAD, and the results achieved are still quite far from those of the target oracle system, confirming the need for new and more robust V-MAD systems that, also exploiting the potentiality of deep learning, can effectively work directly on video sequences rather than on single images. The development of new V-MAD approaches will also have to address the issue related to the unavailability of datasets representative of this scenario. Our future research will be dedicated to proposing new contributions to these aspects.

References

[1] Muhtahir O Oloyede, Gerhard P Hancke, and Hermanus C Myburgh. A review on face recognition systems: recent approaches and challenges. *Multimedia Tools and Applications*, 79(37):27891–27922, 2020. 1

[2] Matteo Ferrara, Annalisa Franco, and Davide Maltoni. The magic passport. In *IEEE International Joint Conference on Biometrics, Clearwater, IJCB 2014, FL, USA, September 29 - October 2, 2014*, pages 1–7. IEEE, 2014. 1, 2

[3] Matteo Ferrara and Annalisa Franco. *Morph Creation and Vulnerability of Face Recognition Systems to Morphing*, pages 117–137. Springer International Publishing, Cham, 2022. 1

[4] Ulrich Scherhag, Christian Rathgeb, Johannes Merkle, Ralph Breithaupt, and Christoph Busch. Face recognition systems under morphing attacks: A survey. *IEEE Access*, 7:23012–23026, 2019. 1

[5] Ulrich Scherhag, Andreas Nautsch, Christian Rathgeb, Marta Gomez-Barrero, Raymond NJ Veldhuis, Luuk Spreeuwers, Maikel Schils, Davide Maltoni, Patrick Grother, Sebastien Marcel, et al. Biometric systems under morphing attacks: Assessment of morphing techniques and vulnerability reporting. In *2017 International Conference of the Biometrics Special Interest Group (BIOSIG)*, pages 1–7. IEEE, 2017. 1

[6] Muhammad Hamza, Samabia Tehsin, Hanen Karamti, and Norah Saleh Alghamdi. Generation and detection of face morphing attacks. *IEEE Access*, 10:72557–72576, 2022. 1

[7] Guido Borghi, Nicolò Di Domenico, Annalisa Franco, Matteo Ferrara, and Davide Maltoni. Revelio: A modular and effective framework for reproducible training and evaluation of morphing attack detectors. *IEEE Access*, 2023. 1

[8] 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. 1, 2, 3, 6, 7, 8, 9

[9] Guido Borghi, Emanuele Pancisi, Matteo Ferrara, and Davide Maltoni. A double siamese framework for differential morphing attack detection. *Sensors*, 21(10):3466, 2021. 1, 2, 3, 6, 7, 8, 9

[10] José Sánchez del Río, Cristina Conde, Aristeidis Tsitiridis, Jorge Raúl Gómez, Isaac Martín de Diego, and Enrique Cabello. Face-based recognition systems in the abc e-gates. In *2015 Annual IEEE Systems Conference (SysCon) Proceedings*, pages 340–346. IEEE, 2015. 1

[11] Jose Sanchez Del Rio, Daniela Moctezuma, Cristina Conde, Isaac Martin de Diego, and Enrique Cabello. Automated border control e-gates and facial recognition systems. *computers & security*, 62:49–72, 2016. 1

[12] Matteo Ferrara, Annalisa Franco, Davide Maltoni, and Christoph Busch. Morphing attack potential. In *2022 International Workshop on Biometrics and Forensics (IWBF)*, pages 1–6, 2022. 1

[13] BI Frontex. Automated biometric border crossing systems based on electronic passports and facial recognition: Rapid and smartgate. Technical report, Tech. rep., agency, European cooperation, operational borders, external . . . , 2010. 1

[14] Matteo Ferrara, Annalisa Franco, and Davide Maltoni. Face demorphing. *IEEE Transactions on Information Forensics and Security*, 13(4):1008–1017, 2017. 2, 3, 6, 7, 8, 9

[15] Qiang Meng, Shichao Zhao, Zhida Huang, and Feng Zhou. Magface: A universal representation for face recognition and quality assessment. In *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 14220–14229, 2021. 2, 4, 5, 9

[16] Philipp Terhörst, Jan Niklas Kolf, Naser Damer, Florian Kirchbuchner, and Arjan Kuijper. Ser-fiq: Unsupervised estimation of face image quality based on stochastic embedding robustness. In *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 5650–5659, 2020. 2, 4, 5, 9

[17] Fadi Boutros, Meiling Fang, Marcel Klemt, Biying Fu, and Naser Damer. Cr-fiqa: Face image quality assessment by learning sample relative classifiability. In *2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 5836–5845, 2023. 2, 4, 5, 9

[18] Robin Rombach, Andreas Blattmann, Dominik Lorenz, Patrick Esser, and Björn Ommer. High-resolution image synthesis with latent diffusion models. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 10684–10695, June 2022. 2

[19] Diederik P Kingma, Max Welling, et al. An introduction to variational autoencoders. *Foundations and Trends® in Machine Learning*, 12(4):307–392, 2019. 2

[20] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial networks. *Communications of the ACM*, 63(11):139–144, 2020. 2

[21] Guido Borghi, Annalisa Franco, Gabriele Graffieti, and Davide Maltoni. Automated artifact retouching in morphed images with attention maps. *IEEE Access*, 9:136561–136579, 2021. 2, 8

[22] Nicolò Di Domenico, Guido Borghi, Annalisa Franco, and Davide Maltoni. Face restoration for morphed images retouching. In *Proceedings of the 12th International Workshop On Biometrics And Forensics (IWBF)*, 2024. 2

[23] Ulrich Scherhag, Christian Rathgeb, and Christoph Busch. *Face Morphing Attack Detection Methods*, pages 331–349. Springer International Publishing, Cham, 2022. 2

[24] Jiankang Deng, Jia Guo, Niannan Xue, and Stefanos Zafeiriou. Arcface: Additive angular margin loss for deep face recognition. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 4690–4699, 2019. 3, 7

[25] Information technology — extensible biometric data interchange formats — part 5: Face image data. Standard, International Organization for Standardization, 2019. 5

[26] Torsten Schlett, Christian Rathgeb, Olaf Henniger, Javier Galbally, Julian Fierrez, and Christoph Busch. Face image quality assessment: A literature survey. *ACM Comput. Surv.*, 54(10s), sep 2022. 5

[27] ISO/IEC 29794-5 — Information technology — Biometric sample quality — Part 5: Face image data. Standard, International Organization for Standardization, under development. 5

[28] Kiran Raja, Matteo Ferrara, Annalisa Franco, Luuk Spreeuwers, Ilias Batskos, Florens de Wit, Marta Gomez-Barrero, Ulrich Scherhag, Daniel Fischer, Sushma Krupa Venkatesh, et al. Morphing attack detection-database, evaluation platform, and benchmarking. *IEEE transactions on information forensics and security*, 16:4336–4351, 2020. 6

[29] Frontex. Best practice operational guidelines for Automated Border Control (ABC) systems – Research and development unit. Technical report, Publications Office of the European Union, 2012. 7