Quality-driven Adaptive Morphing Attack Detection in Operational Scenarios via Online Learning

Quality-driven Adaptive Morphing Attack Detection in Operational Scenarios via Online Learning

Nicolò Di Domenico[1], Annalisa Franco[1], Guido Borghi[2], Davide Maltoni[1]
[1] Department of Computer Science and Engineering, University of Bologna, Italy
[2] Department of Education and Humanities, University of Modena and Reggio Emilia, Italy

***Abstract*— Morphing Attack Detection (MAD) systems often suffer from performance degradation when deployed in operational environments, such as airports, that differ from the training domain. We propose an adaptive differential MAD framework that continuously refines a pre-trained detector using live bona fide samples acquired at the gate. The system is memoryless, so no samples are stored in memory to mitigate privacy concerns about the collection of personal data. To prevent the loss of discriminative power caused by bona fide-only adaptation, the method generates synthetic morph samples on-the-fly by combining the current operational subject with identities from an external public or synthetic face dataset. The adaptation process further relies on a quality-aware bona fide selection strategy and a controlled balancing mechanism for synthetic morph generation. Experimental results show that the proposed method improves target-domain specialization while maintaining robustness against morph attacks.**

I. INTRODUCTION

Face recognition systems are increasingly used in security-critical applications such as border control and Automated Border Crossing (ABC) gates. In these scenarios, the reliability of biometric verification is essential, as identity decisions directly impact security and access control. However, Face Recognition systems (FRs) are known to be vulnerable to several types of presentation attacks. Among these, face morphing attacks represent a particularly challenging threat [30].

A morphing attack consists of generating a facial image that combines the identity and appearance of two different individuals [11]. When such an image is successfully enrolled as a reference image in a document, both contributing subjects may later be able to authenticate using the same biometric identity. This capability makes morphing attacks especially dangerous in applications such as passport issuance and border control, where the reference image is typically acquired only once during the enrollment process.

To address this issue, Morphing Attack Detection (MAD) methods have been proposed to distinguish bona fide facial images from morphed ones [25]. Existing approaches can generally be divided into three main categories [10]: single-image MAD (S-MAD), which analyzes a single facial image to determine whether it is morphed, and differential MAD (D-MAD), which exploits the comparison between two images of the same identity, typically the reference image stored in the document and the probe image captured during verification. Recently, the Video MAD (V-MAD), based on an input sequence, has also been introduced [2]. In operational scenarios such as ABC gates, D-MAD and V-MAD approaches are particularly attractive since both images, or video frames acquired by FRs, are naturally available during the verification process.

This project received funding from the European Union's Horizon Europe research and innovation program under Grant Agreement No.101121280. This text reflects only the author's views, and the commission is not liable for any use that may be made of the information contained therein.

(a) Document (b) Gate

Fig. 1: The differences between the photos in the document and the images captured live in operational scenarios, such as airport gates, can be significant, for instance, in terms of pose and lighting. These variations negatively impact the performance of Face Recognition systems.

In recent years, Deep Learning-based techniques have significantly improved the performance of MAD systems [1]. Models trained on large datasets containing both bona fide and morphed images have demonstrated promising results in controlled evaluation settings. Nevertheless, a key limitation remains: MAD models trained offline often experience a noticeable drop in performance when deployed in real operational environments.

This issue is mainly caused by the domain shift between the training data and the images captured in operational conditions. Indeed, training datasets are typically collected under controlled or semi-controlled conditions, while operational environments introduce a wide range of variations, including different acquisition devices, illumination conditions, subject poses, and image quality levels, as shown in Figure 1. As a result, models trained offline may not fully capture the characteristics of the target deployment scenario. The consequence is that the system generates too many false alarms (bona fide documents wrongly classified as morphing attempts), requiring frequent intervention from human operators and making it unsuitable for practical deployment.

One possible solution is to adapt the MAD model during

deployment using data collected directly in the operational environment. However, this setting raises several practical challenges. A straightforward approach would consist in capturing images at the gate, exporting them, and performing offline fine-tuning of the model. However, this solution is hardly applicable in practice: exporting biometric data is often prohibited due to privacy regulations, and manual intervention on the gate is typically not feasible, as it may both violate regulatory constraints and introduce overhead or negatively impact system usability. Therefore, model adaptation must take place directly on the gate device, requiring solutions that are both effective and computationally efficient. However, this setting introduces additional constraints: even maintaining a small set of reference images (e.g., morphed pairs) to support the update process can be problematic due to storage limitations and privacy concerns. As a result, adaptation strategies must avoid relying on external data storage or replay buffers, and instead operate within the strict resource of the deployment environment. Finally, relying solely on operational samples for adaptation is inherently risky. In real-world scenarios, morphing attack attempts are expected to represent only a very small fraction of the overall traffic, resulting in a strong imbalance toward bona fide samples. Consequently, updating the model using only operational data, largely assumed to be genuine, may progressively bias it toward the bona fide class, ultimately reducing its ability to effectively detect morphing attacks.

In this work, we address these challenges by proposing an adaptive D-MAD framework designed for an operational gate scenario. The proposed method performs online adaptation of a pre-trained D-MAD model using live facial images acquired during system operation, and could be easily extended to V-MAD approaches. The adaptation process relies on the practical assumption that the vast majority of document images correspond to bona fide users (indeed, the number of attacks is significantly lower than the number of non-counterfeit documents), allowing them to be used as unlabeled training samples. To prevent the loss of discriminative capability that may arise when adapting the model using only bona fide samples, the proposed framework introduces a mechanism for generating synthetic morph samples on the fly. Specifically, morph images are generated by combining the facial image of the subject currently interacting with the gate with identities drawn from an external dataset of facial images. These synthetic morphs act as negative samples that help maintain the model's sensitivity to morphing attacks during the adaptation process.

In addition, the proposed method incorporates a selection strategy for live samples based on quality-related criteria. Not all frames acquired during the interaction with the gate are equally suitable for model updates, as some may be affected by motion blurring, extreme poses, or poor illumination. Therefore, the impact of image quality on the adaptation process needs to be investigated.

Finally, the framework controls the number of synthetic morphs generated during adaptation, allowing the training process to maintain a balanced contribution between bona fide operational samples and synthetic attack examples.

Overall, the proposed approach enables the MAD detector to progressively specialize to the characteristics of the target operational environment while avoiding the need to store past samples or rely on labeled data. By combining live bona fide images with on-the-fly synthetic morph generation, the framework supports continuous adaptation while preserving the ability to detect morphing attacks.

The main contributions of this work can be summarized as follows:

- **Memoryless adaptive D-MAD framework**: an online adaptation strategy that refines a pre-trained morphing attack detector using facial images acquired during system operation at an automated gate, without storing past samples.
- **On-the-fly synthetic morph generation**: to avoid the loss of discriminative power caused by bona fide-only adaptation, the proposed method generates morph samples online by combining the current subject with identities drawn from an external face dataset.
- **Quality-aware selection of live samples**: a strategy to select reliable frames based on quality-related criteria, improving the stability of the adaptation process.
- **Balanced online adaptation**: an analysis of the number of generated morph samples to maintain a balanced contribution between bona fide live samples and synthetic morph examples during training.

II. Related works

A. Face Morphing Attacks

Morphed facial images have demonstrated the capacity to successfully deceive both Commercial-Off-The-Shelf (COTS) Face Recognition Systems and trained human examiners [11], thereby establishing face morphing as a substantial threat to border security infrastructure. In addition, the rapid advancement and widespread accessibility of generative AI frameworks, including Diffusion Models [26], Variational Autoencoders (VAEs) [9], and Generative Adversarial Networks (GANs) [14], have considerably lowered the technical barrier for malicious exploitation, amplifying the scale and sophistication of potential attacks. Compounding this issue, morphed facial composites are amenable to both manual and automated post-processing retouching pipelines [7], [3], which systematically suppress visible and statistical artifacts, rendering detection substantially more challenging. These converging factors underscore the critical necessity for the development of robust, next-generation MAD systems endowed with the capacity to generalize across different training and operational domains.

B. Morphing Attack Detection

As mentioned, Single-image MAD (S-MAD) methods operate exclusively on a single facial image and are primarily deployed during the enrollment phase, where submitted identity photographs are screened for morphing artifacts prior to their inclusion in an eMRTD. The second approach, *i.e.* the Differential MAD (D-MAD), is instead designed for

the verification stage: for instance, at ABC gates where the document photo is directly compared against a live facial capture acquired on-site. It is important to note that, by design, only D-MAD approaches are compatible with the scenario in which sequences are available as input (V-MAD).

In the D-MAD category, a range of methodological strategies has been explored. A prominent solution proposed in [31] employs a dual-embedding architecture based on the ArcFace model [6], wherein the resulting feature vectors are fed into an SVM classifier to determine whether a given image has been subjected to morphing manipulation. Subsequent works have investigated the integration of complementary feature sources. In [2], identity embeddings are augmented with artifact-sensitive features explicitly encoding morphing traces, thereby preserving detection capability even when input subjects exhibit strong physiological resemblance. Ensemble-based strategies [34], [4] combine multiple classifiers and heterogeneous feature extractors to achieve more robust and generalized detection performance.

The V-MAD category, originally established in [2], is still largely underexplored in the literature. Some works [32], [33] explore the use of multi-frame inputs sourced from four distinct camera systems for the D-MAD task, though this approach operates under a different operational configuration than the one examined in the present work, in which frames are acquired by a single device in a frontal position with respect to the human. Despite these preliminary contributions, the V-MAD domain remains substantially underexplored, with critical open challenges pertaining to dataset availability, generalization across unseen morphing algorithms, and robustness to video compression artifacts.

III. Proposed Approach

We propose an adaptive, memoryless D-MAD framework operating directly at the acquisition gate. The proposed D-MAD method continuously updates the model using only currently observed data, combining selected bona fide samples from live video with on-the-fly generated morphs to preserve robustness against morphing attacks. All operations are performed in real-time to meet the requirements of operational scenarios.

An overview of the pipeline is shown in Fig. 2. From a general view, the pipeline can be summarized in four different steps:

1) **Video acquisition and frame selection:** a video of the subject is captured and decomposed into a sequence of frames. Each frame is evaluated through an image quality assessment module, and only those satisfying predefined criteria are retained. The selected frames form a set of trusted (bona fide) live samples, representative of the current acquisition conditions.
2) **Bona fide pairs construction:** each selected live frame is paired with the corresponding document image to form bona fide comparison pairs, used to adapt the model to the current domain.
3) **On-the-fly morph pairs generation:** morphed images are generated by combining the document image with identities sampled from a reference face gallery, producing morphed versions of the document image. The generated morphed document images are paired with the same selected live frames to produce the morphed pairs used for model updating.
4) **Model update:** both bona fide and synthetic attack pairs are used to update the D-MAD model using only current data, without storing past samples.

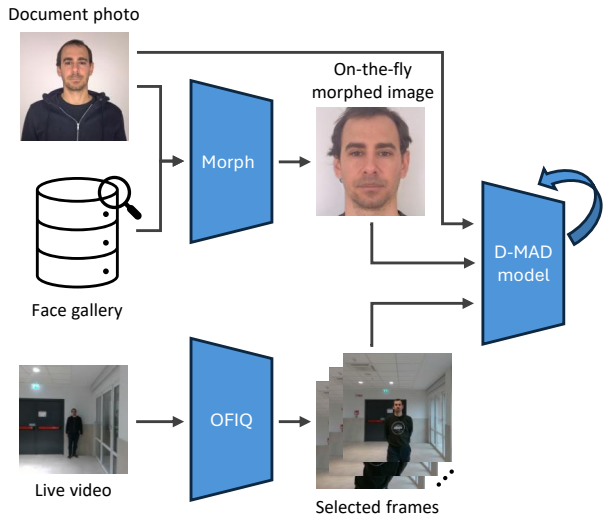


Fig. 2: Architecture of the proposed system. The proposed D-MAD model is continuously adapted using bona fide images, obtained through a quality selection of the live video frames, and on-the-fly morph pairs. In this manner, the D-MAD does not require a memory containing past samples, alleviating privacy concerns.

The approach assumes that document images are predominantly bona fide, with only a negligible number of potential morph outliers that do not significantly affect the update. Extending the framework with a preliminary MAD stage to explicitly verify document authenticity is left for future work.

A. Problem Setting

We address the problem of online adaptation of a differential morphing attack detection model at the acquisition gate. Rather than the proposal of a new MAD model, the paper is focused on defining an effective update strategy that leverages streaming data acquired during operation, an element not investigated in the literature.

At each time step t, the system captures a live video of the subject, represented as a set of frames $\mathcal{V}_t = \{x^i_{l,t}\}_{i=1}^{N_t}$, together with an associated document image x_d. Here, i indexes the individual frames within the video, and N_t denotes the total number of frames acquired at time t. Since the live acquisition is performed under controlled conditions at the gate, all frames in $\mathcal{V}_t$ are considered trusted bona fide samples. However, not all frames are equally informative, since the acquisition conditions (in terms of lighting, blurring, and pose) can vary significantly, as shown

in Figure 3. Therefore, a subset $\mathcal{S}_t \subseteq \mathcal{V}_t$ is selected based on quality criteria and used to construct comparison pairs.

The D-MAD model operates on pairs of images of the form (x_d, x_l), where x_d is the document image and x_l is a trusted live capture. The model is updated sequentially over time: starting from an initial model D_0, at each step t the model D_{t-1} is updated using the currently observed data to obtain D_t.

In this initial formulation, we assume that document images are predominantly bona fide, with only a negligible number of potential morph outliers that do not significantly affect the update process. The goal is to update the model using only the current data, improving adaptation to the operational domain while preserving robustness to morphing attacks through the inclusion of synthetically generated samples. The update is performed in a streaming fashion, without retaining past samples.

As D-MAD model, we take inspiration from one of the best-performing methods in the literature [31]. Specifically, we first obtain the ArcFace [6] embeddings of both the document image x_d and the trusted live capture image x_l, obtaining respectively e_d and e_l; then, differently from the original paper, we train a Multi-Layer Perceptron to classify whether x_d is morphed or not by feeding it the difference $e_d - e_l$ of the two identity embeddings.

B. Online update

The live acquisition at time t of a given subject is represented as a set of frames $\mathcal{V}_t = \{x^i_{l,t}\}$ extracted from the captured video. Since not all frames are equally informative, a selection step is performed using a quality function $Q(\cdot)$:

$$\mathcal{S}_t = \{x \in \mathcal{V}_t \mid Q(x) \geq \tau\}. \tag{1}$$

Specifically, we used the OFIQ framework[1] in this work, which is the reference implementation of the ISO/IEC 29794-5 [16] standard for face image quality assessment. In particular, in this work we consider the pose-related quality score, which is one of the most critical factors for D-MAD performance.

All the selected frames are then paired with the document image x_d to form bona fide pairs:

$$\mathcal{P}^{bf}_t = \{(x_d, x_{l,t}) \; \forall x_{l,t} \in \mathcal{S}_t\}. \tag{2}$$

To simulate morphing attacks, synthetic morphed document images are generated by exploiting a morphing algorithm M to combine the document image x_d with identities sampled from a face gallery $\mathcal{G}$:

$$\mathcal{M}_t = \{M(x_d, x_g) \mid x_g \in \mathcal{G}\}. \tag{3}$$

Each generated morph is paired with the selected live frames to simulate attack scenarios:

$$\mathcal{P}^{m}_t \subseteq \mathcal{M}_t \times \mathcal{S}_t. \tag{4}$$

where $\mathcal{M}_t$ denotes the set of synthetically generated morphed document images.

[1] https://github.com/BSI-OFIQ/OFIQ-Project

Fig. 3: Sample of live video frames in which the image quality (pose, lighting, and blurring) significantly varies.

At each time step t, the model D_{t-1} is updated using the current sets of bona fide and attack pairs: bona fide pairs $\mathcal{P}^{bf}_t$, and morph pairs $\mathcal{P}^{m}_t$.

The update is performed as:

$$D_t = \mathcal{U}(D_{t-1}, \mathcal{P}^{bf}_t, \mathcal{P}^{m}_t), \tag{5}$$

where $\mathcal{U}(\cdot)$ denotes the update procedure.

Importantly, after the update, all samples are discarded. Therefore, the method follows a memoryless learning paradigm that can alleviate privacy concerns.

Taking inspiration from [21], the update procedure $\mathcal{U}(\cdot)$ is implemented using a continual learning strategy based on Learning without Forgetting (LwF [18]). This choice is motivated by the need to adapt the model to newly observed data while preserving its ability to detect morphing attacks learned from previous updates. In particular, LwF introduces a regularization mechanism that constrains the updated model D_t to remain consistent with the predictions of the previous model D_{t-1}, thus mitigating catastrophic forgetting. This is particularly relevant in the proposed setting, where the update relies only on current data and no past samples are retained. From a practical perspective, the use of LwF allows performing incremental updates using the current sets of bona fide and synthetic morph pairs, without requiring access to previously observed data.

In the experimental evaluation, the proposed update strategy is compared against two alternative approaches: (i) standard fine-tuning, which updates the model using only current data without explicit regularization, and (ii) full retraining from scratch using all available data, which serves as an upper bound on performance. This comparison highlights the effectiveness of the proposed continual learning approach in balancing adaptation and robustness.

IV. Experimental Setting

A. Datasets

For our experiments we rely on two distinct datasets that are employed at different times: the first, Idiap Morph, is used to train the initial model D_0 and then to verify its retention of morphing attack detection capabilities; the second, VideoID-V, is used to update the model in the following timesteps $t > 0$, as well as to evaluate the variation in performance with respect to the new scenario.

More in detail, the Idiap Morph [27], [28] dataset is a publicly available set of several datasets common in the MAD literature, namely Feret [23], FRGC [22], and Face Research Lab London Set [5]. These datasets are then morphed

with five different algorithms, in particular OpenCV [29], FaceMorpher [24], StyleGAN [17], WebMorph [5], and AMSL [20]. In total, this dataset offers 13478 pairs of document and live images; 80% of these are reserved for training the initial model D_0, while the remaining 20% is never seen during training and is reserved for evaluating the model's performance on detecting morphing attacks carried out with several algorithms.

The dataset reserved for $t > 0$, namely VideoID-V, is composed of videos of 65 subjects with varying age, ethnicity, and gender, and is acquired in two different indoor settings: one similar to an airport gate, and one in a darker room, providing a more challenging scenario.

For every scenario, each subject walks toward a stationary camera fixed on a tripod, but different acquisition conditions concerning head orientation and garments are considered. More specifically, the following three scenarios are represented:

- **Frontal gaze**: depicts the subject walking towards the camera while continuously looking straight at it;
- **Looking around**: the subject is asked to walk towards the camera while performing slow neck rotations and not fixing their gaze towards a specific point;
- **With occlusions**: alternative version of the second one, but the subject is now wearing either a hat, a scarf, or both, with the aim of partially occluding the subject's face, providing the most challenging scenario.

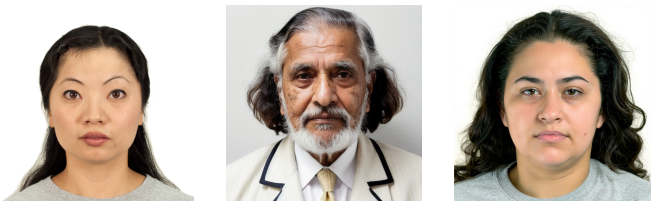
All video sequences are acquired at a resolution of 1280×720 pixels at 30 frames per second, with a mean duration of 14.3 ± 3.6 seconds, employing an Intel RealSense D435i camera. In addition, for each subject, an ICAO-compliant image simulating the document photo is available. This photo is acquired at a resolution of 3024×5376 pixels and its compliance with ICAO specifications is verified with the Correlance ccEngine ICAO[2] commercial SDK. In total, this dataset provides 65 high-quality ICAO-compliant images and 384 videos that simulate a gate transit.

For our purposes, we reserve half of the subjects for training, and the other half for testing. The morphed images used to construct the attack attempts in the test split are generated with the UBO [12], [13] face morphing algorithm with a morphing factor of 0.5, sampling only from subjects belonging to the test split, to avoid any identity leakage between the two sets. Criminal-accomplice pairs are chosen by selecting the most similar other subjects of the test set for each individual, according to the average similarity scores obtained with three commercial SDKs, to maximize the attack potential of the generated morphed images.

B. Morphed Images Generation

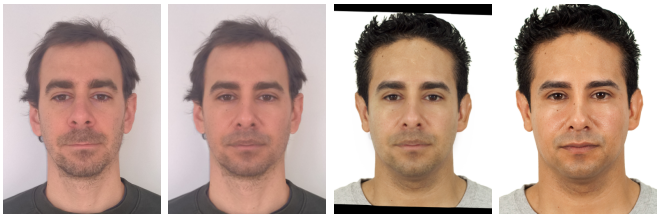
To generate the synthetic morphed document images, we combine a given document image with identities sampled from a face gallery $\mathcal{G}$ of ICAO-compliant images. For our experiments, we populate this gallery in two alternative ways, one with real identities, thus obtaining $\mathcal{G}_R$, and one with purely synthetic ones, named $\mathcal{G}_S$. More specifically, $\mathcal{G}_R$ contains 831 subjects of varying ethnicities sourced from the Chicago Faces Dataset [19] (also referred to as CFD); in contrast, $\mathcal{G}_S$ is composed of a total of 7592 distinct identities, 7281 of which belong to the FLUXSynID [15] dataset, while the remaining 311 are obtained from the ONOT [8] dataset. A sample of all three datasets is depicted in Figure 4.

[2] https://www.correlance.com/cms/it/ccEngineICAO



(a) CFD (b) ONOT (c) FLUXSynID

Fig. 4: Sample images of the three datasets used to populate the two galleries – $\mathcal{G}_R$ and $\mathcal{G}_S$ – to construct the synthetic morphed images.



(a) Bona fide (b) Accomplice (c) Criminal (d) Gallery

Fig. 5: Sample images of document photos belonging to the same subject. The bona fide photo (a) is sourced directly from the traveler's document once scanned; to generate the morphed images on the fly, the subject is paired with the most similar gallery photo (d), and then two morphed images are generated, where the original subject respectively takes the role of the accomplice (b) and the criminal (c).

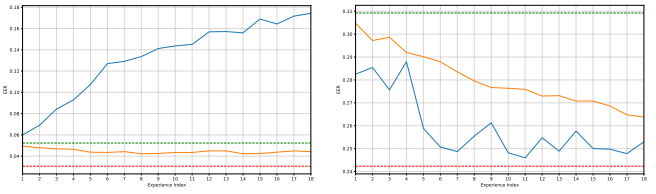
More specifically, for each subject we identify the single most similar face x_g within the gallery $\mathcal{G}$, *i.e.*, the one that obtains the highest cosine similarity between the ArcFace [6] embeddings of the original subject and those of all gallery images. Once this match is identified, we generate two morphed images: in the first, the original subject is assigned the role of the criminal and the chosen gallery face that of the accomplice; in the second, these roles are reversed. In both cases, the morphing factor is set to 0.5, which means that both subjects contribute equally to the resulting morphed image; however, the identity that is immediately visible to the naked eye always belongs to the accomplice, thus obtaining two distinct images. Subsequently, for both bona fide and newly generated morphed document images, we construct the corresponding attempts by pairing each document image with a selection of frames $\mathcal{S}_t$ from all videos belonging to the original subject. To avoid faces that are too in profile, we set $\tau = 85$ as our threshold for the quality function $Q(\cdot)$. A sample of document images and frames present in an attempt is shown in Figure 5.

TABLE I: Comparison of different adaptation strategies across different types of acquired sequence in terms of relative EER reduction with respect to the baseline error rate.

		Baseline	Joint	F.T.	LwF [18]
Initial	EER	0.052	-42%	+233%	-16%
	$B_{0.1}$	0.028	-63%	+774%	-26%
	$B_{0.05}$	0.061	-69%	+502%	-43%
	$B_{0.01}$	0.176	-57%	+235%	+5.8%
Frontal gaze	EER	0.265	-18%	-16%	-14%
	$B_{0.1}$	0.510	-35%	-16%	-27%
	$B_{0.05}$	0.671	-37%	-1.5%	-22%
	$B_{0.01}$	0.883	-6.3%	+5.2%	-14%
With occlusions	EER	0.346	-22%	-19%	-8.8%
	$B_{0.1}$	0.644	-29%	-12%	-11%
	$B_{0.05}$	0.781	-21%	-4.2%	-14%
	$B_{0.01}$	0.954	-5.2%	-0.9%	-5.6%
Looking around	EER	0.369	-27%	-25%	-15%
	$B_{0.1}$	0.746	-35%	-34%	-17%
	$B_{0.05}$	0.829	-23%	-19%	-13%
	$B_{0.01}$	0.965	-14%	-4.1%	-9.0%

TABLE II: Comparison of real vs. synthetic gallery for on-the-fly morphing generation.

		Real			Synthetic		
		Joint	F.T.	LwF	Joint	F.T.	LwF
Initial	EER	-42%	+233%	-16%	-25%	+183%	-16%
	$B_{0.1}$	-63%	+774%	-26%	-63%	+700%	-11%
	$B_{0.05}$	-69%	+502%	-43%	-43%	+460%	-33%
	$B_{0.01}$	-57%	+235%	+6%	-44%	+298%	-21%
Adapt.	EER	-22%	-23%	-15%	-12%	-4.5%	-14%
	$B_{0.1}$	-36%	-27%	-20%	-12%	+9.9%	-17%
	$B_{0.05}$	-33%	-14%	-16%	-0.2%	+6.1%	-12%
	$B_{0.01}$	-4.6%	-1.6%	-10%	+0.5%	-2.7%	-6.0%



(a) Initial test set (b) Adaptation test set

Fig. 6: Evolution of the EER of model D_t on the initial and adaptation test sets for two update strategies: blue lines show the model's performance when using fine tuning, while orange lines show its performance with LwF. The green dashed line indicates the baseline before adaptation, and the red dashed line indicates the performance achieved with the joint approach, which serves as an upper bound.

C. Metrics

To present the experimental results, we adopt the standard error-based measures commonly used in the MAD task. In particular, we report the Bona fide Sample Classification Error Rate (BSCER), which measures the proportion of genuine samples that are incorrectly labeled as morphed. Conversely, the Morphing Attack Classification Error Rate (MACER) captures the proportion of morphed images that are erroneously classified as genuine. In the literature, BSCER is typically examined at fixed MACER operating points. In line with this practice, we consider $\text{BSCER}_{0.1}$ ($B_{0.1}$), $\text{BSCER}_{0.05}$ ($B_{0.05}$), and $\text{BSCER}_{0.01}$ ($B_{0.01}$), which correspond to the minimum BSCER achievable under the constraints that MACER remains at or below 10%, 5%, and 1%, respectively. Among these, $\text{BSCER}_{0.01}$ is a particularly demanding criterion and is widely adopted as the standard operating point for practical facial verification systems.

Additionally, to obtain a quick, global indicator of performance, we report the Equal Error Rate (EER), defined as the operating point at which the BSCER and MACER curves intersect. As these are error-focused metrics, lower values indicate better performance. Finally, to more clearly characterize performance trends, we also compute the relative change in error rates w.r.t. the baseline model's results.

V. Experimental results

This section presents the experimental evaluation of the proposed adaptive D-MAD framework. The goal is to assess how effectively the model can be updated using operational data, while preserving its ability to detect morphing attacks.

A. Evaluation of the updating procedure

This section compares the proposed MAD updating procedure based on LwF to two other updating strategies, which represent the main alternative to continual learning:

- *Joint*: all training data, from both the initial training set and the adaptation set, are supposed to be available at the same instant and to be exploited for model training. This approach is typically selected as an upper bound to the performance achievable since it represents the best possible option (not compatible with the operational conditions) where all data can be exploited at best.
- *Fine Tuning* (F.T.): the updating procedure exploits the adaptation data to update the initial model applying a simple fine tuning of the initial model.

For computational efficiency, model updates for both fine-tuning and LwF are performed every 10 acquired video sequences, rather than after each individual sequence. Frames for updating are selected exclusively according to the pose-related quality score, with a threshold set to 85. The impact of quality-based frame selection is analyzed in section V-C

The results are reported in Table I for both the initial test set and the adaptation test set. The experimental results highlight several key aspects of the proposed approach. First, the idea of updating the initial model using operational data proves to be highly effective. This is clearly demonstrated by the significant error reduction achieved by joint training, which improves performance on both the initial and the adaptation test sets. This indicates that incorporating operational data allows the model to become more robust to real-world conditions. It is worth noting the reduction of both EER and BSCER, indicating an overall lower error, as well as a reduced number of false positives.

Second, fine-tuning emerges as a highly aggressive update strategy. Although it can significantly reduce error rates on

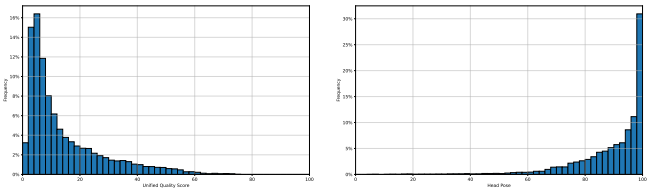
the adaptation set, this comes at the cost of substantial degradation in performance on the initial test set. This behavior confirms the well-known catastrophic forgetting effect, where the model overfits to recent data and loses its ability to generalize to previously seen distributions.

In contrast, the LwF-based approach provides a more balanced update strategy. Although the improvements are generally more moderate, they are consistent across both the initial and adaptation test sets. This demonstrates that LwF enables effective model adaptation while preserving, and in some cases even improving, performance on the original data. As such, it effectively mitigates catastrophic forgetting.

Finally, it is important to consider that, in real deployment scenarios, the update process may run continuously over extended periods and leverage large amounts of incoming data, potentially collected in short time intervals. In this context, a more conservative and stable update strategy, such as LwF, is particularly well suited, as it ensures interesting performance improvements without compromising previously acquired knowledge. In Figure 6, we report the evolution of the EER computed on both the initial and adaptation test sets for model D_t after t time steps, employing either the fine-tuning strategy (in blue) or LwF (in orange). For comparison, we also include the baseline performance of D_0 (in green) and the results obtained with the joint update strategy (in red). Indeed, it is possible to note that, while fine tuning is able to sharply diminish error rates with a reduced number of time steps, at the same time we notice a significant and immediate performance penalty on the initial test set. Conversely, LwF not only is effective – though with slower convergence – in lowering error rates on the adaptation test set, but it also manages to slightly surpass the baseline performance on the initial test set. In general, the trend is much smoother for LwF, while fine-tuning is much more irregular and less stable. The good trend observed for LwF suggests that a long-term update (made possible by the application scenario) can lead to significant improvements in the model's accuracy.

An additional analysis was conducted to evaluate the effectiveness of the proposed adaptation strategy across different types of acquired sequences. In particular, the considered sequences exhibit varying levels of control, ranging from more constrained conditions to more challenging scenarios characterized by significant variability in pose, illumination, and the presence of accessories. These factors are known to have a strong impact on the performance of morphing attack detection systems.

The results, reported in Table I, confirm the effectiveness of the proposed update mechanism across all sequence types. Notably, the improvement is particularly evident in the most challenging conditions, where the initial model typically struggles the most. This suggests that the adaptation process is able to effectively specialize the model to the target operational scenario. Overall, the inclusion of samples with heterogeneous quality during the update phase contributes to improving the robustness of the system, enabling it to better handle a wide range of variations encountered in real-world conditions.



(a) OFIQ Unified quality score (b) OFIQ Head Pose score

Fig. 7: Quality score distributions for the OFIQ Unified Quality Score (a) and the Head Pose Score (b) computed over the live video frames.

B. Real vs. synthetic gallery

In order to fully adhere to a memoryless setting, the proposed framework should ideally avoid relying on a gallery of real identities for the on-the-fly generation of morph samples. In practical deployments, maintaining such a gallery may introduce privacy and data management concerns. For this reason, we investigate an alternative configuration in which the external dataset used for morph generation consists of synthetic face images, which represent a more privacy-friendly solution. To assess the impact of this choice, the same adaptation experiments were repeated by replacing the real-image gallery with the synthetic dataset described in Section IV. The results, reported in Table II, allow a direct comparison between the two configurations.

Overall, the use of synthetic identities leads to slightly lower performance compared to real data. This behavior is expected, as synthetic faces may not fully capture the variability and realism of real-world identities used during morph generation. Nevertheless, the proposed adaptation strategy still provides consistent improvements over the initial model, confirming its effectiveness also under this more constrained and privacy-compliant setting.

These findings highlight a relevant trade-off between performance and privacy. While the use of real images remains beneficial for maximizing adaptation effectiveness, synthetic data represent a viable alternative that enables memoryless operation without relying on sensitive biometric information.

C. Impact of quality-based frame selection

As a further analysis, the plots in Figure 7 show the distribution of quality scores (unified quality and head pose) observed in the live frames extracted from the videos. Regarding the unified quality score, it is evident that the overall quality is generally low, with a large proportion of values falling below 20. This is consistent with a realistic acquisition scenario, where images are often affected by blur, uncontrolled lighting, and other degradations, making them significantly different from the data typically available in public datasets, which are often collected under more controlled conditions. As for face pose, the distribution reveals some variability, particularly in the looking away sequences, as expected. However, in most cases the score remains relatively high: this is because the subject is moving towards the gate, progressively assuming a more frontal pose.

TABLE III: Relative variation of the EER after model updating, for both the initial test set and the adaptation test set, as a function of the quality threshold (applied to the unified quality score) and the head pose-related score.

		Head pose											
		Initial (baseline = 0.052)						**Adaptation** (baseline = 0.309)					
		None	**95**	**90**	**85**	**80**	**75**	**None**	**95**	**90**	**85**	**80**	**75**
Quality	**None**	-18.3%	-19.2%	-11.3%	-15.5%	-16.0%	-14.1%	-12.7%	-12.1%	-15.0%	-14.7%	-15.4%	-15.1%
	5	-16.4%	-13.6%	-14.1%	-16.4%	-14.1%	-16.9%	-12.7%	-11.6%	-11.4%	-12.5%	-11.5%	-14.3%
	10	-11.3%	-16.9%	-13.6%	-13.6%	-11.3%	-12.7%	-10.9%	-8.9%	-9.7%	-10.7%	-10.7%	-10.5%
	15	-14.1%	-13.1%	-19.2%	-13.6%	-13.1%	-16.9%	-9.3%	-8.8%	-6.9%	-5.8%	-9.1%	-8.1%
	20	-8.5%	-11.3%	-9.8%	-16.4%	-5.6%	-8.5%	-5.3%	-7.4%	-5.6%	-6.3%	-6.6%	-5.3%
	25	-7.5%	-12.7%	-7.5%	-7.5%	-7.5%	-7.5%	-13.1%	-8.5%	+1.9%	-13.1%	-13.1%	-13.1%
	30	-8.5%	-10.8%	-10.3%	-8.5%	-8.5%	-8.5%	-5.9%	-6.4%	-4.7%	-5.9%	-5.9%	-5.9%
	35	-7.5%	-8.5%	-7.5%	-7.5%	-7.5%	-7.5%	-3.4%	-5.0%	-4.2%	-3.4%	-3.4%	-3.4%
	40	-11.3%	-8.0%	-4.2%	-11.3%	-11.3%	-11.3%	-4.9%	-4.7%	-2.9%	-4.9%	-4.9%	-4.9%
	45	-8.5%	-8.5%	-8.0%	-8.5%	-8.5%	-8.5%	-2.1%	-2.9%	-1.9%	-2.1%	-2.1%	-2.1%
	50	-7.5%	-4.2%	-5.6%	-7.5%	-7.5%	-7.5%	-0.1%	-0.4%	-0.2%	-0.1%	-0.1%	-0.1%

Overall, this suggests that, despite some variability, pose conditions are generally less critical than the overall visual quality of the frames.

The variability observed in terms of quality suggests that this factor might have a noticeable impact on the effectiveness of the updating process. Extensive experiments have therefore been performed to analyze this aspect. Table III reports the relative variation of the EER after model updating, for both the initial test set and the adaptation test set, as a function of the quality threshold (applied to the unified quality score) and the head pose-related score. Negative values indicate an improvement in performance (EER reduction). Both the unified quality and head pose scores follow a quality-based interpretation, where lower values correspond to poorer image quality.

Overall, the results reveal different behaviors depending on the considered quality factor. When using the unified quality score, the best performance is generally achieved without applying strong selection, suggesting that retaining a wide range of samples is beneficial. In this case, even lower-quality samples contribute positively to the update, likely because they introduce variability that is not well represented in the initial model (*e.g.,* uneven lighting conditions or blurring). In contrast, when considering the head pose-related score, applying a selection becomes more advantageous. Filtering out samples with extreme pose variations leads to more consistent improvements, indicating that pose represents a more critical and potentially disruptive factor for D-MAD. This suggests that not all sources of quality degradation should be treated equally during the selection process.

These observations highlight that effective online adaptation should rely on a factor-aware selection strategy. While preserving diversity is important, especially for general image quality, selectively filtering specific factors such as extreme pose can lead to more stable and effective updates. Overall, the results confirm that online adaptation improves performance not only on the adaptation test set but also on the initial test set, indicating that the updated model becomes more robust while maintaining its ability to reliably detect morphing attacks.

VI. Conclusions and Future Works

In this work, we proposed a memoryless adaptive framework for differential morphing attack detection designed for operational scenarios. The method enables continuous online adaptation of a pre-trained D-MAD model using only data acquired at the gate, without storing past samples or access to ground-truth labels. To mitigate the risk of performance degradation caused by bona fide-only updates, the approach integrates on-the-fly generation of synthetic morph samples, ensuring the preservation of discriminative capability against morphing attacks. The results show that adapting the model with operational data is highly beneficial, improving performance on both the target domain and the original test set. This indicates that the proposed strategy enhances the robustness of the model rather than overfitting to the new data. Among the evaluated update strategies, the LwF-based continual learning approach offers the best trade-off between adaptation and stability, effectively mitigating catastrophic forgetting while enabling progressive specialization to the operational domain. Additionally, the analysis of quality-based sample selection highlights that not all quality factors impact the adaptation process equally. While preserving variability in general image quality proves beneficial, selectively filtering samples based on pose leads to more stable and reliable updates. Furthermore, the comparison between real and synthetic galleries shows that, although real data still provides the best performance, synthetic identities represent a viable and privacy-compliant alternative, with only a limited degradation in effectiveness. Overall, the proposed framework demonstrates that memoryless online adaptation is an effective solution for deploying MAD systems in real-world scenarios, where data distributions are dynamic and constrained by privacy requirements.

Future work will explore several directions. Firstly, integrating a preliminary document verification stage could further improve the robustness of the adaptation process by filtering out potential morph outliers in the assumed bona fide data (e.g. by exploiting the D-MAD model itself). Finally, more advanced strategies for synthetic morph generation could be investigated to better approximate real attack distributions (e.g. by considering multiple morphing methods).

REFERENCES

[1] G. Borghi, N. Di Domenico, A. Franco, M. Ferrara, and D. Maltoni. Revelio: A modular and effective framework for reproducible training and evaluation of morphing attack detectors. *IEEE Access*, 11:120419–120437, 2023.

[2] G. Borghi, A. Franco, N. Di Domenico, M. Ferrara, and D. Maltoni. V-mad: Video-based morphing attack detection in operational scenarios. In *2024 IEEE International Joint Conference on Biometrics (IJCB)*, pages 1–10. IEEE, 2024.

[3] G. Borghi, A. Franco, G. Graffieti, and D. Maltoni. Automated artifact retouching in morphed images with attention maps. *IEEE Access*, 9:136561–136579, 2021.

[4] N. Damer, S. Zienert, Y. Wainakh, A. M. Saladie, F. Kirchbuchner, and A. Kuijper. A multi-detector solution towards an accurate and generalized detection of face morphing attacks. In *2019 22th International Conference on Information Fusion (FUSION)*, pages 1–8. IEEE, 2019.

[5] L. DeBruine and B. Jones. Face research lab london set. *Psychol. Methodol. Des. Anal*, 2017.

[6] J. Deng, J. Guo, N. Xue, and S. 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.

[7] N. Di Domenico, G. Borghi, A. Franco, and D. Maltoni. Face restoration for morphed images retouching. In *2024 12th International Workshop on Biometrics and Forensics (IWBF)*, pages 1–6. IEEE, 2024.

[8] N. Di Domenico, G. Borghi, A. Franco, and D. Maltoni. Onot: a high-quality icao-compliant synthetic mugshot dataset. In *2024 IEEE 18th International Conference on Automatic Face and Gesture Recognition (FG)*, pages 1–10. IEEE, 2024.

[9] P. K. Diederik and W. Max. An introduction to variational autoencoders. *Foundations and Trends® in Machine Learning*, 12(4):307–392, 2019.

[10] M. Ferrara and A. Franco. *Morph Creation and Vulnerability of Face Recognition Systems to Morphing*, pages 117–137. Springer International Publishing, Cham, 2022.

[11] M. Ferrara, A. Franco, and D. Maltoni. The magic passport. In *IEEE International Joint Conference on Biometrics*, pages 1–7, 2014.

[12] M. Ferrara, A. Franco, and D. Maltoni. Face demorphing. *IEEE Trans. Inf. Forensics Secur.*, 13(4), 2018.

[13] M. Ferrara, A. Franco, and D. Maltoni. Decoupling texture blending and shape warping in face morphing. In *International Conference of the Biometrics Special Interest Group (BIOSIG)*, September 2019.

[14] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio. Generative adversarial networks. *Communications of the ACM*, 63(11):139–144, 2020.

[15] R. Ismayilov, D. Sero, and L. Spreeuwers. Fluxsynid: A framework for identity-controlled synthetic face generation with document and live images. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 3698–3708, 2025.

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

[17] T. Karras, S. Laine, M. Aittala, J. Hellsten, J. Lehtinen, and T. Aila. Analyzing and improving the image quality of stylegan. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 8110–8119, 2020.

[18] Z. Li and D. Hoiem. Learning without forgetting. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 275–283, 2017.

[19] D. S. Ma, J. Correll, and B. Wittenbrink. The chicago face database: A free stimulus set of faces and norming data. *Behavior research methods*, 47(4):1122–1135, 2015.

[20] T. Neubert, A. Makrushin, M. Hildebrandt, C. Kraetzer, and J. Dittmann. Extended stirtrace benchmarking of biometric and forensic qualities of morphed face images. *IET Biometrics*, 7(4):325–332, 2018.

[21] L. Pellegrini, G. Borghi, A. Franco, and D. Maltoni. Detecting morphing attacks via continual incremental training. In *2023 IEEE International Joint Conference on Biometrics (IJCB)*, pages 1–10. IEEE, 2023.

[22] P. J. Phillips, P. J. Flynn, T. Scruggs, K. W. Bowyer, J. Chang, K. Hoffman, J. Marques, J. 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)*, volume 1, pages 947–954. IEEE, 2005.

[23] P. J. Phillips, H. Wechsler, J. Huang, and P. J. Rauss. The FERET database and evaluation procedure for face-recognition algorithms. *Image and vision computing*, 16(5):295–306, 1998.

[24] A. Quek. FaceMorpher morphing algorithm. https://github.com/alyssaq/face_morpher. Accessed: 2022-11-30.

[25] K. Raja, M. Ferrara, A. Franco, L. Spreeuwers, I. Batskos, F. De Wit, M. Gomez-Barrero, U. Scherhag, D. Fischer, S. K. Venkatesh, et al. Morphing attack detection-database, evaluation platform, and benchmarking. *IEEE transactions on information forensics and security*, 16:4336–4351, 2020.

[26] R. Rombach, A. Blattmann, D. Lorenz, P. Esser, and B. Ommer. High-resolution image synthesis with latent diffusion models. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 10684–10695, 2022.

[27] E. Sarkar, P. Korshunov, L. Colbois, and S. Marcel. Vulnerability analysis of face morphing attacks from landmarks and generative adversarial networks. *arXiv preprint arXiv:2012.05344*, 2020.

[28] E. Sarkar, P. Korshunov, L. Colbois, and S. Marcel. Are gan-based morphs threatening face recognition? In *ICASSP 2022-2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 2959–2963. IEEE, 2022.

[29] Satya Mallick. "Face morph using opencv — c++ / python. https://learnopencv.com/face-morph-using-opencv-cpp-python/. Accessed: 2022-11-30.

[30] U. Scherhag, A. Nautsch, C. Rathgeb, M. Gomez-Barrero, R. N. Veldhuis, L. Spreeuwers, M. Schils, D. Maltoni, P. Grother, S. 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.

[31] U. Scherhag, C. Rathgeb, J. Merkle, and C. Busch. Deep face representations for differential morphing attack detection. *IEEE transactions on information forensics and security*, 15:3625–3639, 2020.

[32] J. M. Singh and R. Ramachandra. Fusion of deep features for differential face morphing attack detection at automatic border control gates. In *2022 10th European Workshop on Visual Information Processing (EUVIP)*, pages 1–5. IEEE, 2022.

[33] J. M. Singh, R. Ramachandra, K. B. Raja, and C. Busch. Robust morph-detection at automated border control gate using deep decomposed 3d shape & diffuse reflectance. In *2019 15th International Conference on Signal-Image Technology & Internet-Based Systems (SITIS)*, pages 106–112. IEEE, 2019.

[34] J. M. Singh, S. Venkatesh, and R. Ramachandra. Robust face morphing attack detection using fusion of multiple features and classification techniques. In *2023 26th International Conference on Information Fusion (FUSION)*, pages 1–8. IEEE, 2023.