Towards Federated Learning for Morphing Attack Detection

Towards Federated Learning for Morphing Attack Detection

Marta Robledo-Moreno[1], Guido Borghi[2], Nicolò Di Domenico[3],
Annalisa Franco[3], Kiran Raja[4], Davide Maltoni[3]
[1]Universidad Autónoma de Madrid, Spain
[2]University of Modena and Reggio Emilia, Italy
[3]University of Bologna, Italy
[4]Norwegian University of Science and Technology, Norway
marta.robledo@estudiante.uam.es, guido.borghi@unimore,
{name.surname}@unibo.it, kiran.raja@ntnu.no

Abstract

Through the Face Morphing attack is possible to use the same legal document by two different people, destroying the unique biometric link between the document and its owner. In other words, a morphed face image has the potential to bypass face verification-based security controls, then representing a severe security threat. Unfortunately, the lack of public, extensive and varied training datasets severely hampers the development of effective and robust Morphing Attack Detection (MAD) models, key tools in contrasting the Face Morphing attack since able to automatically detect the presence of morphing images. Indeed, privacy regulations limit the possibility of acquiring, storing, and transferring MAD-related data that contain personal information, such as faces. Therefore, in this paper, we investigate the use of Federated Learning to train a MAD model on local training samples across multiple sites, eliminating the need for a single centralized training dataset, as common in Machine Learning, and then overcoming privacy limitations. Experimental results suggest that FL is a viable solution that will need to be considered in future research works in MAD.

1. Introduction

Face Morphing, *i.e*. the technique to merge in a single face two different identities (see Fig. 1), has emerged as a serious threat to Face Recognition systems (FRS) [18]. Indeed, it has been proven [50, 53] that a morphed face can bypass face verification-based security controls, thus enabling a criminal to use the same legal document belonging to an accomplice to fool commercial FRSs located, for instance, in Automated Border Control (ABC) gates. Therefore, accurate and robust Morphing Attack Detection (MAD) [43] solutions are strongly needed to detect the traces of the morphing procedure on document images to improve the security of FRS systems.

(a) Subject 1 (b) Morphed (c) Subject 2

Figure 1. Example of a morphed face (Fig. 1b), a hybrid identity created starting from two subjects (Fig. 1a and 1c). Literature studies [50, 53] have revealed that morphed images can bypass face-based security controls.

Unfortunately, MAD development is often hampered by scarce data availability due to privacy regulations that limit the acquisition, storing, and sharing of datasets with faces and other personal details for improving MAD [5]. Both the lack of large-scale and varied training datasets contrasts the generalization capabilities of MAD algorithms, especially if based on data-hungry deep learning-based approaches, that therefore exhibit poor performance on unseen data [50]. In this scenario, the rise of new AI-based generation tools, such as Generative Adversarial Networks (GANs) [21], Variational Autoencoders (VAEs) [26], and Diffusion Models [46], represents a viable solution to create synthetic data for training purposes in agreement with privacy regulations. However, synthetic images are often characterized by limited quality and resolution and by the presence of generation artifacts that can compromise the efficacy and the generalization ability of solutions trained with them [8]. In particular, in [58] the feasibility of training a MAD system using only synthetic images has been investigated, reaching the conclusion that a model trained with only synthetic images is less robust than a model trained us-

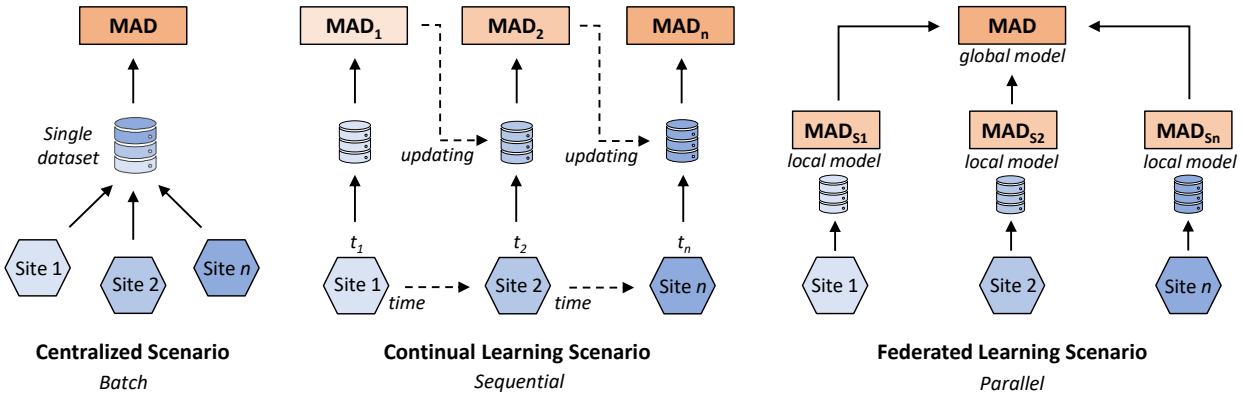

Figure 2. Different strategies to train a MAD model. In the centralized scenario, data belonging to different sites are collected in a single dataset and used to train the single MAD model in a batch-based manner. Differently, in the Continual Learning scenario, the same MAD model is sequentially updated each time new data become available on the same or different sites. Finally, in the Federated Learning scenario, several MAD models are locally trained on their own data and then a global model is build over the single local models.

ing also real images. However, the recent literature [23, 10] highlights the rising importance of synthetic data in training MAD models that achieve competitive performance.

Commonly, MAD learning models are developed on large dataset collections, often centralized in a single training site, following the common Machine Learning (ML) scenario. However, as mentioned, centralization is impeded by the possibility of transferring and sharing data, in addition to the risk of privacy breaches and potential misuse of personal information. In this context, Federated Learning (FL) [29] provides an alternative solution that enables model training without centralized data storage. By distributing the training process across multiple devices or data sources while preserving data privacy, FL aligns seamlessly with privacy compliance regulations and offers an opportunity to train MAD models on datasets that are not shared in their entirety.

Therefore, in this paper, we introduce and investigate the application of FL techniques for the face morphing attack detection task. While few existing works [5, 38] have made strides in this direction, these efforts predominantly rely on the Continual Learning (CL) [37] paradigm, an alternative approach based on the sequential updating of the same model over time and across multiple sites (see Fig. 2), using specific algorithms [30, 27] to prevent the so-called catastrophic forgetting [34].

In contrast, our work represents one of the first attempts to harness the power of Federated Learning for face morphing attack detection, paving the way for future research works. In summary, we validate the feasibility and effectiveness of the FL paradigm in the development of both Single-Image (S-MAD) and Differential (D-MAD) systems (see Sect. 2.2), comparing the centralized and distributed MAD training on images generated with 11 different morphing algorithms and available in public datasets.

2. Related Work

2.1. Face Morphing

Face morphing is an image manipulation technique to gradually transform one image into another. This method, originally detailed in [18] within the context of manipulating electronic machine readable travel (eMRTD) documents, enables the generation of human faces with a dual identity (see Fig. 1), that can bypass automated face verification systems and human controls [50, 43, 22]. Recently, this paradigm has been applied also on 3D data [55].

Face morphing is exacerbated by the proliferation of techniques employing generative AI for face morphing [62, 60, 9], simplifying the task for potential malefactors. Furthermore, the morphed images can undergo refinement through manual [19] or automated retouching processes [4, 15], effectively eliminating discernible and imperceptible artifacts and increasing the level of challenge in this task.

Hence, there is a strong need to develop new Morphing Attack Detection (MAD) [43, 52] systems, *i.e.* automated tools explicitly devised to detect the presence of morphing in input images, that are accurate and able to generalize on unseen images.

2.2. Morphing Attack Detection

The existing literature mainly presents two categories of MAD methods [43]: Single-image MAD (S-MAD) and Differential MAD (D-MAD).

S-MAD systems rely solely on single input images present in documents. Typically, S-MAD systems prioritize the detection of artifacts or traces left by the morphing procedure in the whole or part of the input image [31]. In [39] authors propose a system based on an additional external information, *i.e.* a watchlist. Generally, this task is considered challenging, since it is usually based exclusively on infor-

mation contained in a single image: this observation is confirmed by the results obtained on sequestered datasets [2].

Conversely, a D-MAD system receives two distinct images as input: a trusted live capture and an image under examination which may potentially be morphed. D-MAD systems operate under the assumption that one of the two inputs has been acquired through a trusted process, such as via the camera in the ABC gate or a procedure overseen by a law enforcement officer. In this case, D-MAD system can rely not only on the detection of morphing traces in input document images but also on the comparison of input identities [54, 25] or the analysis of the differences computed between the two images [44].

2.3. MAD training strategies

Regardless of the type of algorithm (S-MAD or D-MAD), a MAD approach can be trained using different training strategies, summarized in the visual overview of Figure 2. In the literature, the large majority of proposed MAD solutions are based on a centralized solution, in which data produced in different sites are collected in a single dataset and a batch-based training procedure creates a single MAD model. Currently, state-of-the-art MAD models are based on this kind of approach that presents issues related to the privacy of the training data (usually acquired in multiple sites and then transferred and stored in a single location). Conversely, only a few recent works focus on alternative training procedures, also investigating different operational scenarios.

In [5], a method based on Continual Learning (CL) to enable the incremental and sequential training of an S-MAD system on different sites is proposed. Specifically, the authors compare two different continual learning algorithms, *i.e.* Elastic Weight Consolidation (EWC) [27] and Learning without Forgetting (LwF) [30], and a fine-tuning strategy to investigate the performance of morphing detectors trained on different datasets, sequentially provided as input. Recently, in [38] a CL strategy for D-MAD algorithms is analyzed, focusing particularly on the input chunk size that influences the learning process.

In general, continual learning strategies are more focused on a sequential update of the same MAD model over time rather than the development of a single MAD solution based on multiple training sites, like in the FL scenario. Moreover, we observe that incremental learning strategies based on CL still present a gap in performance compared to traditional centralized training procedures.

2.4. Federated Learning

Federated Learning (FL) [29], a relatively recent paradigm in the field of machine learning, has garnered considerable attention due to its novel approach to data privacy protection and decentralized learning processes.

At its core, FL enables multiple client devices or servers to collaboratively learn a shared model while keeping all the training data localized, thus addressing significant concerns regarding data privacy and security (see Fig. 2). This new paradigm introduces also challenges to be addressed, regarding, for instance, the availability of unbalanced local data, not representative of the whole population, or the need for efficient communication systems to connect all the clients with a limited latency [35]. It has found applications in diverse fields, especially on those systems based on sensitive information, from healthcare [61] to finance [7], *i.e.* in domains where data sharing is heavily restricted [17]. The efficiency of federated learning, particularly in handling vast amounts of data across various nodes, is another important area of research [65], and represents an important requirement for real-world applications with large datasets and numerous clients.

It is worth noting that Federated Learning presents unique features that make this approach quite different from the CL paradigm. Indeed, from a general point of view, clients in FL are organized in an infrastructure that guarantees low latency, and they are simultaneously available at the same time. Moreover, differently from CL, it is important to retain a duplicate of the training dataset to facilitate potential model updates. We observe this procedure might raise issues concerning privacy – consider, for instance, the potential storage of photographs taken at airport gates [38] – and then the coexistence of CL and FL approaches is well motivated.

3. Federated Learning for MAD

3.1. Distributed learning

We model the FL approach on a central server and multiple clients (see Fig. 3): training is an iterative process where, at each epoch, the server transfers the current global model (w) to each client who subsequently updates the local model (w_i) on its dataset (d_i). After updating, each client transmits a weight update to the server, that consolidates the global model based on the aggregation weights (λ_i) assigned to each client:

$$w = \sum_{i}^{n} \lambda_i \, w_i \tag{1}$$

We implement a cross-site model evaluation to plot the accuracy of the global model distributed to each client on the local datasets. This distributed learning procedure has been implemented through the *Nvidia Federated Learning Application Runtime Environment* (FLARE) [47], exploiting its open-source and general-purpose paradigm. Specifically, we use PyTorch as deep learning framework. Each client has $\lambda = 1$ and it is equipped with a Nvidia GeForce GTX 1070 6GB GPU.

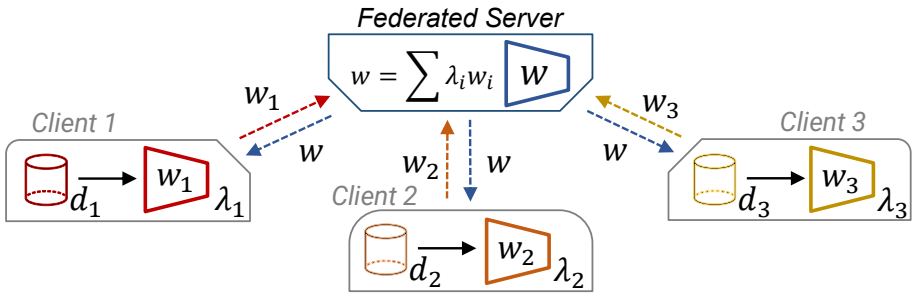

Figure 3. The Federated Learning training procedure implemented in this paper. Three different MAD models (w_i) are locally trained on their own private data (d_i). The weights of each model are then collected by a federated server (w) to create a single global model, as a result of a weighted summary of local weights. In this manner, the training procedure is privacy compliant since no data are transferred on different sites. Further details are reported in Section 3.1.

Aware of the high number of S-MAD solutions available in the literature, we implement the S-MAD model described in [3], due to the sota results obtained on sequestered datasets [2]. Therefore, we adopt the Inception-ResNet [57] architecture as a binary classifier (the output classes are "morphed" or "bona fide"), trained with the SGD optimizer with a momentum of 0.9 and an initial learning rate of 0.001. Input faces are detected and cropped through the MTCNN [63] face detector which has shown superior performance in the S-MAD task. No data augmentation procedures or additional pre-computed features (*e.g.* Fourier [64], wavelets [1], and PRNU [11]) are exploited, since their efficacy and generalizability in the MAD task seems to be limited [3].

Similarly, also for the D-MAD task, we take inspiration from one of the current state-of-the-art approaches described in [54]. Specifically, we extract two embeddings of size 512 from the input image through the ArcFace [13] model. Since the model is trained for the face recognition task, we assume these embeddings represent the identity of the input faces. Then, we combine the two embeddings through subtraction and then we classify the resulting vector with an MLP classifier, with layers of size $512, 250, 125, 64$, trained with the SGD optimizer with a learning rate of 0.001. The use of this deep learning architecture instead of the original one (SVM), is motivated by the investigation of federated learning with deep learning architectures through NVFLARE which is not possible with traditional machine learning approaches.

3.2. Datasets

Progressive Morphing Database (PMDB) [19]: it comprises 1108 morphed images, derived from three well-known datasets: AR [33], FRGC [40], and Color Feret [41], utilizing the public morphing algorithm delineated in [19]. For the creation of these morphed images, a total of 280 subjects were used, divided into 134 males and 146 females. Morphed images may exhibit visible artifacts, such as blurred regions or ghosting effects. However, background artifacts are removed through an automatic substitution procedure applied during the morphing process.

Idiap Morph database [48]: it constitutes a publicly accessible compilation of several datasets, encompassing five subsets generated using different morphing algorithms: OpenCV [49], FaceMorpher [42], StyleGAN [24], WebMorph [12], and AMSL [36]. These subsets leverage facial images from the Feret, FRGC, and Face Research Lab London Set [12] datasets. Notably, the morphed images produced using OpenCV and FaceMorpher algorithms exhibit a diminished visual quality due to the presence of artifacts in both the background and face region. With the StyleGAN-based method, the visual artifacts are less pronounced: however, typical GAN-related textures remain observable. The AMSL morphing algorithm has been utilized to generate 2175 morphed images. A compression technique is applied to all images to accommodate the limitation of the eMRTD used in official documents. Consequently, the resulting images are confined to a maximum size of 15 kB. This compression process notably increases the difficulty of the S-MAD task, as it tends to eliminate most artifacts potentially introduced during the morphing process.

ChiMo dataset [3]: it is a compilation generated from the images with neutral expressions of the Chicago Faces Database (CFD) [32] which includes photographs of 831 individuals from diverse ethnic backgrounds. For each subject, five other subjects sharing the same ethnicity and gender were selected for the morphing process. To identify the most similar subjects for each individual, the average face verification scores from three commercial SDKs — VeriLook[1], Cognitec[2], and Innovatrics[3] — were utilized. The morphing was executed using two different morphing factors (0.3 and 0.5) and three distinct morphing algorithms, *i.e.* FaceFusion [16], UTW [43], and NTNU [43], for a total

[1] www.neurotechnology.com/verilook.html
[2] www.cognitec.com
[3] www.innovatrics.com/

of 24k morphed images, with each algorithm contributing to 8k images.

FEI Morph dataset [14]: it is derived from the FEI Face Database [59] and encompasses images of 200 individuals, evenly distributed between male and female subjects. The facial representations predominantly feature individuals aged between 19 and 40, exhibiting diverse appearances, hairstyles, and accessories. Comprising 6000 morphed images, the dataset employs three distinct morphing algorithms, namely FaceFusion [16], UTW [43], and NTNU [43], utilizing two varying morphing factors (0.3 and 0.5).

3.3. Metrics

In assessing the efficacy of the MAD model after training, we employ the error-based metrics specific to the MAD task [43]. The Bona Fide Presentation Classification Error Rate (BPCER) quantifies the proportion of genuine images incorrectly identified as morphed:

$$\text{BPCER}(\tau) = \frac{1}{N}\sum_{i=1}^{N} H(b_i - \tau) \tag{2}$$

Conversely, the Morphing Attack Classification Error Rate (MACER) measures the proportion of morphed images inaccurately classified as bona fide:

$$\text{MACER}(\tau) = 1 - \left[\frac{1}{M}\sum_{i=1}^{M} H(m_i - \tau)\right] \tag{3}$$

In both equations, τ is the score threshold on which b_i, m_i, the detection scores, are compared; $H(x) = \{1 \text{ if } x > 0,\ 0 \text{ otherwise}\}$ is defined as a step function.

Commonly, BPCER is reported in relation to a predetermined MACER value: in our experiments, we examine $\text{BPCER}_{0.05}$ ($\mathbf{B}_{0.5}$), and $\text{BPCER}_{0.01}$ ($\mathbf{B}_{0.01}$), which correspond to the minimum BPCER attainable with an MACER of no more than 5%, and 1%, respectively. It is noteworthy that the last metric presents a particularly rigorous challenge and typically serves as the standard operational point for face verification systems in practical applications.

Finally, for the S-MAD task, we also report the Weighted Average Error across Datasets (WAED) [3]:

$$\text{WAED} = \sum_{E\in\mathcal{E}}\sum_{D\in\mathcal{D}} w_D w_E E(D) \tag{4}$$

where E is the error metric computed on a specific dataset D, $w_{D,E}$ are the weights for each dataset and metric. Specifically, we adopt $w_E = [0.3, 0.1, 0.2, 0.4]$ for the EER, $\text{BPCER}_{0.1}$, $\text{BPCER}_{0.05}$ and $\text{BPCER}_{0.01}$, respectively. Following [3], these weights reflect the importance of the most common real-world operating point (*i.e.* $\text{B}_{0.01}$), followed by the EER, a metric important for evaluating the performance of the system at a glance. As dataset weights w_D, we use $w_D = [1.0, 0.94, 0.88, 0.8, 0.78, 0.77]$ for the morphing algorithms FaceFusion, NTNU, UTW, WebMorph, Squirlz, and AMSL, respectively. In this case, these weights reveal the dataset complexity, measured through different face recognition models as reported in [3]. We observe the WAED metric is useful to simplify the comparison across different error metrics computed on several testing datasets into a single value.

Combination	Datasets			
	UBO	OpenCV	FaceMorpher	StyleGAN
C1	✔	✔	✔	
C2		✔	✔	✔
C3	✔		✔	✔
C4	✔	✔		✔

Table 1. Input data combination used in our federated learning evaluations. These 4 datasets are divided 80-20 for training and testing (first four lines of Tables 2 and 3), while in the other lines, a cross-dataset comparison is reported.

3.4. Experimental Protocol

For the S-MAD task, to facilitate comparative analysis of our findings, we adopt the protocol defined in [3]: we group training and testing morphed data relying on the morphing algorithm used to create their morphed images.

In the training set, we randomly include the 80% of the images created through UBO, OpenCV, FaceMorpher, and StyleGAN, and the remaining 20% is used for testing. The results obtained with this configuration are reported in the first four lines of the result tables. In this set, it is noteworthy that the morphed images exhibit a low visual quality: this degradation is attributed to various factors, such as the presence of morphing traces. In our implementation, the number of clients corresponds to the number of available GPUs ($n = 3$), and each client trains its model on a single training set. Therefore, there are 4 different combinations in input, as reported in Table 1.

A second testing set, is created through a cross-algorithm procedure, since morphing algorithms used to produce images in training and testing splits are different. The experimental results of this configuration are reported in the last six lines of the tables. We include all the images created using WebMorph, AMSL, Squirlz, FaceFusion, NTNU, and UTW morphing algorithms. Bona fide images are taken from the FRGC, Color Feret and AMSL (only neutral expression), and are divided into train and testing using the 80-20 split.

For the D-MAD task, we use in training the same 4 combinations of the S-MAD task (see Table 1), based on UBO, OpenCV, FaceMorpher and StyleGAN morphing al-

Morphing Alg.	C1			C2			C3			C4		
	EER	$\mathbf{B}_{0.05}$	$\mathbf{B}_{0.01}$	**EER**	$\mathbf{B}_{0.05}$	$\mathbf{B}_{0.01}$	**EER**	$\mathbf{B}_{0.05}$	$\mathbf{B}_{0.01}$	**EER**	$\mathbf{B}_{0.05}$	$\mathbf{B}_{0.01}$
UBO	.001	.000	.000	.063	.070	.090	.008	.000	.002	.003	.000	.000
OpenCV	.005	.000	.000	.002	.000	.000	.005	.002	.003	.000	.000	.000
FaceMorpher	.001	.000	.000	.001	.000	.000	.001	.000	.000	.003	.000	.000
StyleGAN	.094	.162	.507	.007	.000	.006	.060	.075	.378	.068	.082	.260
AMSL	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000
Webmorph	.025	.000	.350	.050	.050	.200	.011	.000	.150	.074	.100	.750
Sqirlz Morph	.080	.098	.239	.010	.003	.011	.033	.033	.056	.032	.026	.377
FaceFusion	.102	.249	.671	.079	.152	.537	.091	.190	.529	.113	.392	.826
NTNU	.153	.452	.823	.079	.124	.491	.097	.208	.558	.136	.509	.828
UTW	.266	.806	.960	.523	.978	1.00	.362	.858	.951	.509	.935	.981
WAED $\downarrow$		.3209			.2597			**.2504**			.3972	

Table 2. Performance with Federated Learning on the S-MAD task. Different combinations of input morphing algorithms are reported in each column. Input combinations are reported in Table 1. Further details about metrics are reported in Section 3.3.

Morphing Alg.	C1			C2			C3			C4		
	EER	$\mathbf{B}_{0.05}$	$\mathbf{B}_{0.01}$	**EER**	$\mathbf{B}_{0.05}$	$\mathbf{B}_{0.01}$	**EER**	$\mathbf{B}_{0.05}$	$\mathbf{B}_{0.01}$	**EER**	$\mathbf{B}_{0.05}$	$\mathbf{B}_{0.01}$
UBO	.000	.000	.000	.018	.007	.029	.000	.000	.000	.000	.000	.000
OpenCV	.002	.000	.003	.006	.000	.003	.005	.000	.003	.005	.000	.004
FaceMorpher	.001	.000	.000	.003	.000	.001	.002	.000	.001	.004	.000	.002
StyleGAN	.039	.027	.275	.006	.000	.006	.003	.000	.001	.003	.000	.000
AMSL	.002	.000	.000	.050	.050	.100	.000	.000	.000	.001	.000	.000
Webmorph	.158	.500	.800	.250	.450	.650	.121	.350	.650	.150	.450	.750
Sqirlz Morph	.010	.000	.011	.010	.001	.007	.003	.000	.003	.011	.000	.012
FaceFusion	.154	.362	.620	.094	.191	.444	.144	.351	.614	.153	.368	.631
NTNU	.120	.304	.628	.097	.213	.535	.119	.290	.599	.145	.384	.685
UTW	.231	.675	.922	.458	.929	.990	.261	.698	.901	.210	.598	.870
WAED $\downarrow$		.3267			.3212			**.3057**			.3231	

Table 3. Performance on the S-MAD task through the centralized training. Different combinations of input morphing algorithms are reported in each column. Input combinations are reported in Table 1. Further details about metrics are reported in Section 3.3.

gorithms, while in testing we use the FEI Morph dataset (ChiMo dataset does not include pairs and then its usage is limited to the S-MAD). Therefore, also in this case, a cross-dataset evaluation is carried out.

3.5. Experimental Results

Experimental results using the FL paradigm on the S-MAD task are reported in Table 2. From a general point of view, being aware that this is still a complex task in the literature, low error percentages suggest that FL is a promising approach for MAD. As expected, the performance is superior when the same morphing algorithm is present in both the training and testing set (first four lines of the table), while the error increases in the other cases, specifically with high-quality morphed images produced with NTNU and UTW, achieving the highest values. The low WAED values in combinations C1, C2 and C3 indicate that the FaceMorpher, which produces low-quality morphings, seems to have a positive impact on the training procedure.

The results obtained through FL are then compared with the performance of the same MAD model trained on the same training combinations but adopting the common centralized ML scenario, in which the available datasets are used together as a whole. Specifically, the training pipeline, including model architecture and face detector, data augmentation and learning parameters, are the same as the FL case, but implemented through the Revelio framework [3]. These results are reported in Table 3. Similar to the previous case, results confirm that the presence of the same morphing algorithms in the training set simplifies the detection of morphed images. Interestingly, the WAED metric reveals a generally higher error rate with respect to the

federated learning approach, even though the maximum error achieved is lower (for instance, C2 combination on the UTW produces EER $= 0.458$ and EER $= 0.523$ for centralized and FL solution, respectively).

The results of the D-MAD task are reported in Table 4: also in this case, we observe that the FL approach presents comparable, or even better, performance with respect to the centralized case. In particular, we observe that C4 is the best input combination, suggesting that, differently from S-MAD, the presence of artifacts in the trained data can negatively affect the accuracy of the face verification model used to extract identity-related embeddings.

Finally, we test the best combinations of the FL and centralized approaches on the sequestered (*i.e.* data are not visible and are accessible only for the evaluation and not for the training) SOTAMD [43] datasets, through the BOEP platform [2]: experimental results obtained with the best previous input combination (C3) are reported in Table 5. Although the paper focuses on the federated learning scenario and does not aim to generate state-of-the-art results, we also report the results of the main competitors in the table for the reader's reference and for completeness.

For the S-MAD experiment, the FL solution achieves performance comparable to those of the centralized approach, with a slightly higher equal error rate and higher $\mathbf{B}_{0.1}$ ($\sim 17\%$). The comparison with the competitor highlights that both centralized and federated learning solutions achieve great accuracy; only the method described in [3], which shares the same deep learning-based architecture, overcomes both our solutions, revealing the importance of the augmentation techniques based on image compression – that are out of the scope of this paper – to prevent overfitting and achieve competitive results on the SOTAMD dataset.

For the D-MAD task, the FL strategy achieves lower performance with respect to some competitors but outperforms the centralized one. The last result is a bit counter-intuitive since the centralized approach should generally be able to better exploit the training data, having them all available at the learning stage. However, we believe that the FL approach can better deal with the problems raising from unbalanced training datasets. In particular, with the centralized approach small datasets will have a lower impact on the model training as compared to larger datasets, due to the reduced number of samples available. In the FL approach, if equal weights are assigned to all clients, such as in our case, local models are equally important and contribute to the global model to the same extent.

Considering the whole experimental evaluation and the error rates on the single datasets, results do not show a clear superiority of the centralized approach, thus suggesting the feasibility of the FL paradigm with respect to the centralized one. Finally, we observe the need for extensive research to better analyze the possible advantages deriving from FL and to improve the MAD performance for both S-MAD and D-MAD tasks.

Method	Federated Learning			Centralized		
	EER	$\mathbf{B}_{0.05}$	$\mathbf{B}_{0.01}$	EER	$\mathbf{B}_{0.05}$	$\mathbf{B}_{0.01}$
C1	.071	.097	.343	.106	.190	.808
C2	.070	.108	.423	.105	.438	.960
C3	.078	.123	.335	.093	.175	.615
C4	.068	.080	.333	.095	.203	.752

Table 4. Performance with Federated Learning on the D-MAD task. Different combinations of input morphing algorithms are in each row, as detailed in Table 1.

Method	S-MAD		Method	D-MAD	
	EER	$\mathbf{B}_{0.1}$		EER	$\mathbf{B}_{0.1}$
[3]	.103	.116	[54]	.045	.020
[45]	.318	.650	[14]	.102	.103
[20]	.389	1.00	[19]	.141	.172
[56]	.414	1.00	[6]	.234	.350
[28]	.423	.780	[51]	.335	.528
$\text{Ours}_{\text{CENT.}}$	.290	.550	$\text{Ours}_{\text{CENT.}}$	.226	.428
Ours_{FL}	.293	.726	Ours_{FL}	.182	.288

Table 5. Performance on the S-MAD and D-MAD tasks for the centralized, federated learning training strategies and competitors on the SOTAMD dataset [43] through the BOEP platform [2].

4. Conclusion

In this paper, we have presented one of the first investigations on the use of Federated Learning (FL) in the S-MAD and D-MAD tasks. Specifically, our goal has been to establish a foundation for understanding the potential benefits of leveraging FL in the context of biometrics, where privacy regulations often hamper the acquisition, distribution, and sharing of new personal data. In our experimental evaluation, the Nvidia Federated Learning Application Runtime Environment (FLARE) is exploited as a framework to implement the FL pipeline. Experimental results confirm the feasibility of this paradigm to obtain good MAD accuracy while respecting data privacy compliance. Further research work is needed in order to improve performance from a general point of view.

In future work, we plan to investigate the impact of FL in a real-world distributed MAD training scenario, to test the real impact of network communications, latency, and model transfer between multiple clients. In addition, experiments with more datasets, and then more clients, will be carried out to extend this version of the paper.

Acknowledgement

This project received funding from the European Union's Horizon 2020 research and innovation program under Grant Agreement No. 883356. Disclaimer: 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.

References

[1] P. Aghdaie, B. Chaudhary, S. Soleymani, J. Dawson, and N. M. Nasrabadi. Detection of morphed face images using discriminative wavelet sub-bands. In *2021 IEEE International Workshop on Biometrics and Forensics (IWBF)*, pages 1–6. IEEE, 2021. 4

[2] Biolab. FVC-onGoing. https://biolab.csr.unibo.it/fvcongoing/. Accessed: 2022-11-30. 3, 4, 7

[3] 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*, 2023. 4, 5, 6, 7

[4] G. Borghi, A. Franco, G. Graffieti, and D. Maltoni. Automated artifact retouching in morphed images with attention maps. *IEEE Access*, 9:136561–136579, 2021. 2

[5] G. Borghi, G. Graffieti, A. Franco, and D. Maltoni. Incremental training of face morphing detectors. In *2022 26th International Conference on Pattern Recognition (ICPR)*, pages 914–921. IEEE, 2022. 1, 2, 3

[6] G. Borghi, E. Pancisi, M. Ferrara, and D. Maltoni. A double siamese framework for differential morphing attack detection. *Sensors*, 21(10):3466, 2021. 7

[7] D. Byrd and A. Polychroniadou. Differentially private secure multi-party computation for federated learning in financial applications. In *Proceedings of the First ACM International Conference on AI in Finance*, pages 1–9, 2020. 3

[8] L. Chai, D. Bau, S.-N. Lim, and P. Isola. What makes fake images detectable? understanding properties that generalize. In *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part XXVI 16*, pages 103–120. Springer, 2020. 1

[9] N. Damer, M. Fang, P. Siebke, J. N. Kolf, M. Huber, and F. Boutros. Mordiff: Recognition vulnerability and attack detectability of face morphing attacks created by diffusion autoencoders. In *2023 11th International Workshop on Biometrics and Forensics (IWBF)*, pages 1–6. IEEE, 2023. 2

[10] N. Damer, C. A. F. López, M. Fang, N. Spiller, M. V. Pham, and F. Boutros. Privacy-friendly synthetic data for the development of face morphing attack detectors. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 1606–1617, 2022. 2

[11] L. Debiasi, U. Scherhag, C. Rathgeb, A. Uhl, and C. Busch. Prnu-based detection of morphed face images. In *2018 International Workshop on Biometrics and Forensics (IWBF)*. IEEE, 2018. 4

[12] L. DeBruine and B. Jones. Face research lab london set. *Psychol. Methodol. Des. Anal*, 2017. 4

[13] 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. 4

[14] N. Di Domenico, G. Borghi, A. Franco, and D. Maltoni. Combining identity features and artifact analysis for differential morphing attack detection. In *International Conference on Image Analysis and Processing*, pages 100–111. Springer, 2023. 5, 7

[15] N. Di Domenico, G. Borghi, A. Franco, and D. Maltoni. Face restoration for morphed images retouching. In *12th International Workshop on Biometrics and Forensics (IWBF)*, pages 1–6. IEEE, 2024. 2

[16] FaceFusion. Facefusion. http://www.wearemoment.com/FaceFusion/. Accessed: 2022-11-30. 4, 5

[17] T. Fedullo, D. Cassanelli, G. Gibertoni, F. Tramarin, L. Quaranta, I. Riva, L. Tanga, F. Oddone, and L. Rovati. Assessment of a vision-based technique for an automatic van herick measurement system. *IEEE Transactions on Instrumentation and Measurement*, 71:1–11, 2022. 3

[18] M. Ferrara, F. Annalisa, and M. Davide. The magic passport. In *IEEE International Joint Conference on Biometrics (IJCB'14)*, pages 1–7, 2014. 1, 2

[19] M. Ferrara, A. Franco, and D. Maltoni. Face demorphing. *IEEE Transactions on Information Forensics and Security*, 13(4):1008–1017, 2017. 2, 4, 7

[20] M. Ferrara, A. Franco, and D. Maltoni. Face morphing detection in the presence of printing/scanning and heterogeneous image sources. *IET Biometrics*, 10(3):290–303, 2021. 7

[21] 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. 1

[22] M. Hamza, S. Tehsin, M. Humayun, M. F. Almufareh, and M. Alfayad. A comprehensive review of face morph generation and detection of fraudulent identities. *Applied Sciences*, 12(24):12545, 2022. 2

[23] M. Huber, F. Boutros, A. T. Luu, K. Raja, R. Ramachandra, N. Damer, P. C. Neto, T. Gonçalves, A. F. Sequeira, J. S. Cardoso, et al. Syn-mad 2022: Competition on face morphing attack detection based on privacy-aware synthetic training data. In *2022 IEEE International Joint Conference on Biometrics (IJCB)*, pages 1–10. IEEE, 2022. 2

[24] 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. 4

[25] R. Kessler, K. Raja, J. Tapia, and C. Busch. Towards minimizing efforts for morphing attacks—deep embeddings for morphing pair selection and improved morphing attack detection. *Plos one*, 19(5):e0304610, 2024. 3

[26] D. P. Kingma, M. Welling, et al. An introduction to variational autoencoders. *Foundations and Trends® in Machine Learning*, 12(4):307–392, 2019. 1

[27] J. Kirkpatrick, R. Pascanu, N. Rabinowitz, J. Veness, G. Desjardins, A. A. Rusu, K. Milan, J. Quan, T. Ramalho,

A. Grabska-Barwinska, et al. Overcoming catastrophic forgetting in neural networks. *Proceedings of the national academy of sciences*, 114(13):3521–3526, 2017. 2, 3

[28] C. Kraetzer, A. Makrushin, T. Neubert, M. Hildebrandt, and J. Dittmann. Modeling attacks on photo-id documents and applying media forensics for the detection of facial morphing. In *Proceedings of the 5th ACM workshop on information hiding and multimedia security*, pages 21–32, 2017. 7

[29] T. Li, A. K. Sahu, A. Talwalkar, and V. Smith. Federated learning: Challenges, methods, and future directions. *IEEE signal processing magazine*, 37(3):50–60, 2020. 2, 3

[30] Z. Li and D. Hoiem. Learning without forgetting. *IEEE transactions on pattern analysis and machine intelligence*, 40(12):2935–2947, 2017. 2, 3

[31] M. Long, X. Zhao, L.-B. Zhang, and F. Peng. Detection of face morphing attacks based on patch-level features and lightweight networks. *Security and Communication Networks*, 2022(1):7460330, 2022. 2

[32] 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. 4

[33] A. Martinez and R. Benavente. The ar face database: Cvc technical report, 24. 1998. 4

[34] M. McCloskey and N. J. Cohen. Catastrophic interference in connectionist networks: The sequential learning problem. In *Psychology of learning and motivation*, volume 24, pages 109–165. Elsevier, 1989. 2

[35] B. McMahan, E. Moore, D. Ramage, S. Hampson, and B. A. y Arcas. Communication-efficient learning of deep networks from decentralized data. In *Artificial intelligence and statistics*, pages 1273–1282. PMLR, 2017. 3

[36] 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), 2018. 4

[37] G. I. Parisi, R. Kemker, J. L. Part, C. Kanan, and S. Wermter. Continual lifelong learning with neural networks: A review. *Neural networks*, 113:54–71, 2019. 2

[38] L. Pellegrini, G. Borghi, A. Franco, D. Maltoni, et al. Detecting morphing attacks via continual incremental training. In *Proceedings of 2023 IEEE International Joint Conference on Biometrics*, 2023. 2, 3

[39] F. Peng, L. Qin, and M. Long. Face morphing attack detection and attacker identification based on a watchlist. *Signal Processing: Image Communication*, 107:116748, 2022. 2

[40] 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. 4

[41] 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. 4

[42] A. Quek. FaceMorpher morphing algorithm. https://github.com/alyssaq/face_morpher. Accessed: 2022-11-30. 4

[43] 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. 1, 2, 4, 5, 7

[44] R. Ramachandra and G. Li. Residual colour scale-space gradients for reference-based face morphing attack detection. In *2022 25th International Conference on Information Fusion (FUSION)*, pages 1–8. IEEE, 2022. 3

[45] R. Ramachandra, S. Venkatesh, K. Raja, and C. Busch. Detecting face morphing attacks with collaborative representation of steerable features. In *Proceedings of 3rd International Conference on Computer Vision and Image Processing: CVIP 2018, Volume 1*, pages 255–265. Springer, 2019. 7

[46] 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 (CVPR)*, pages 10684–10695, June 2022. 1

[47] H. R. Roth, Y. Cheng, Y. Wen, I. Yang, Z. Xu, Y. Hsieh, K. Kersten, A. Harouni, C. Zhao, K. Lu, et al. Nvidia flare: Federated learning from simulation to real-world. In *Workshop on Federated Learning: Recent Advances and New Challenges (in Conjunction with NeurIPS 2022)*, 2022. 3

[48] 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. 4

[49] Satya Mallick. "Face morph using opencv — c++ / python. https://learnopencv.com/face-morph-using-opencv-cpp-python/. Accessed: 2022-11-30. 4

[50] 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)*, 2017. 1, 2

[51] U. Scherhag, R. Ramachandra, K. Raja, M. Gomez-Barrero, C. Rathgeb, and C. Busch. On the vulnerability and detection of digital morphed and scanned face images. In *Proc. Int. Workshop on Biometrics and Forensics (IWBF)*, volume 12, 2017. 7

[52] U. Scherhag, C. Rathgeb, and C. Busch. Face morphing attack detection methods. In *Handbook of Digital Face Manipulation and Detection: From DeepFakes to Morphing Attacks*, pages 331–349. Springer International Publishing Cham, 2022. 2

[53] U. Scherhag, C. Rathgeb, J. Merkle, R. Breithaupt, and C. Busch. Face recognition systems under morphing attacks: A survey. *IEEE Access*, 7:23012–23026, 2019. 1

[54] 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. 3, 4, 7

[55] J. M. Singh and R. Ramachandra. 3d face morphing attacks: Generation, vulnerability and detection. *IEEE Transactions on Biometrics, Behavior, and Identity Science*, 2023. 2

[56] L. Spreeuwers, M. Schils, and R. Veldhuis. Towards robust evaluation of face morphing detection. In *2018 26th European Signal Processing Conference (EUSIPCO)*, pages 1027–1031. IEEE, 2018. 7

[57] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich. Going deeper with convolutions. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1–9, 2015. 4

[58] J. Tapia and C. Busch. Impact of synthetic images on morphing attack detection using a siamese network. In *Iberoamerican Congress on Pattern Recognition*, pages 343–357. Springer, 2023. 1

[59] C. E. Thomaz and G. A. Giraldi. A new ranking method for principal components analysis and its application to face image analysis. *Image and vision computing*, 28(6):902–913, 2010. 5

[60] S. Venkatesh, H. Zhang, R. Ramachandra, K. Raja, N. Damer, and C. Busch. Can gan generated morphs threaten face recognition systems equally as landmark based morphs?-vulnerability and detection. In *2020 8th International Workshop on Biometrics and Forensics (IWBF)*, pages 1–6. IEEE, 2020. 2

[61] J. Xu, B. S. Glicksberg, C. Su, P. Walker, J. Bian, and F. Wang. Federated learning for healthcare informatics. *Journal of Healthcare Informatics Research*, 5:1–19, 2021. 3

[62] H. Zhang, S. Venkatesh, R. Ramachandra, K. Raja, N. Damer, and C. Busch. Mipgan—generating strong and high quality morphing attacks using identity prior driven gan. *IEEE Transactions on Biometrics, Behavior, and Identity Science*, 3(3):365–383, 2021. 2

[63] K. Zhang, Z. Zhang, Z. Li, and Y. Qiao. Joint face detection and alignment using multitask cascaded convolutional networks. *IEEE signal processing letters*, 23(10):1499–1503, 2016. 4

[64] L.-B. Zhang, F. Peng, and M. Long. Face morphing detection using fourier spectrum of sensor pattern noise. In *2018 IEEE international conference on multimedia and expo (ICME)*, pages 1–6. IEEE, 2018. 4

[65] M. Zhang, E. Wei, and R. Berry. Faithful edge federated learning: Scalability and privacy. *IEEE Journal on Selected Areas in Communications*, 39(12):3790–3804, 2021. 3