On the Generation of Face Morphs by Inversion of Optimal Morph Embeddings

Hatef Otroshi Shahreza, Laurent Colbois, and Sébastien Marcel

***Abstract*—Automatic face recognition systems are widely used in different applications which require authentication. Among various types of attacks against face recognition systems, morphing attacks have become a major concern, where face images of two subjects are combined into a face morph image which is submitted for enrolment. In a successful attack, both contributing subjects can then authenticate against the morph reference. In this work, we propose a new method to generate face morphs based on inversion of the optimal morph embeddings. To this end, we first find the optimal morph embeddings using the face embeddings of two source face images and then use state-of-the-art template inversion techniques to generate the morph. We use three different template inversion methods: the first one exploits a fully self-contained embedding-to-image inversion model, while the second and third leverage the realistic image generation of a pretrained StyleGAN network and a foundation model based on diffusion models, respectively. Furthermore, we use optimization methods to improve the performance of template inversion methods in the generation of face morph images from optimal morph embeddings. In our experiments, we evaluate the performance of generated face morph images and compare them with state-of-the-art morph generation methods, showing the superiority of our method. We showcase that our method can outperform state-of-the-art deep-learning-based morph generation methods, both in white-box and black-box attack scenarios, and compete with state-of-the-art landmark-based morph generation methods. Moreover, we perform a practical print-scan attack to simulate a real-world scenario and compare our method with previous methods in the literature, demonstrating the effectiveness and superiority of our method. The source code of our proposed method and all experiments are publicly available.**

***Index Terms*—Face Recognition, Embedding, Generation, Morph Attack, Optimal Morph, Template Inversion.**

I. Introduction

FACE recognition (FR) systems have become a ubiquitous solution for automatic authentication in various applications, such as unlocking smartphones[1], e-banking[2], automated border control[3], etc. In spite of advancements in developing face recognition systems over the past decades, these systems are vulnerable to different types of attacks. Among those, morphing attacks in particular are a major concern. In morphing attacks, faces of two contributing subjects are mixed to form a so-called *morph*, which is then submitted as a reference for enrolment in a FR system (e.g., as a passport photo). Later, during the verification stage (e.g., while holding passport at border control), both contributing subjects can then be authenticated by the FR system in a successful attack,

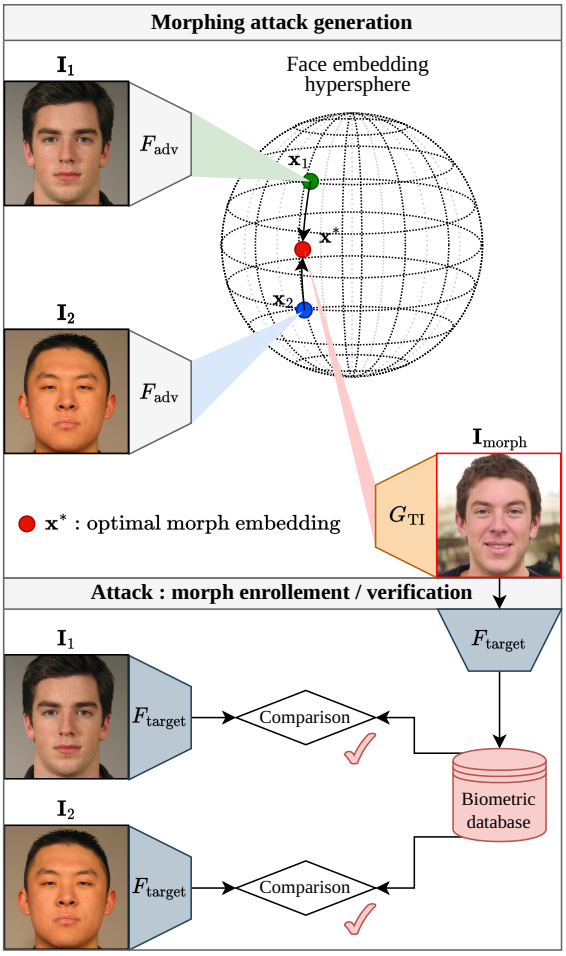


Fig. 1: Morph attack based on optimal morph embedding: 1) Morph Generation: face embeddings are extracted from the source face images using a face recognition network F_{adv} available to the adversary. Then, an optimal morph embedding is computed in the embedding space. Finally, the optimal morph embedding is fed to a template inversion model G_{TI} to generate a candidate morph image. 2) Attack: the generated morph is registered as biometric reference in a database using a distinct face recognition network F_{target}, the target of the attack. In successful attacks, both contributing subjects can authenticate against the stored reference e.g., share a passport.

This research is based upon work supported by the H2020 TReSPAsS-ETN Marie Skłodowska-Curie early training network (grant agreement 860813). This work was also supported by the Swiss Center for Biometrics Research & Testing and the Idiap Research Institute.

Authors are with the Biometrics Security and Privacy Group of Idiap Research Institute, Martigny, Switzerland. Hatef Otroshi Shahreza is also affiliated with École Polytechnique Fédérale de Lausanne (EPFL). Laurent Colbois and Sébastien Marcel are also affiliated with Université de Lausanne (UNIL).

Hatef Otroshi Shahreza and Laurent Colbois contributed equally.

[1]https://apple.co/3mLGCYV

[2]https://bloom.bg/3d2H8j2

[3]https://cnet.co/3sG8qSd

which poses a critical security issue to the FR system as it breaks the fundamental "one passport - one identity" principle.

Face morphs can be generated at the image-level by interpolating the facial landmarks and blending the texture information (landmark-based methods) [1]. Alternatively, several methods have been proposed using deep learning techniques, by interpolation or optimization in the latent space of a face generator network based on Generative Adversarial Networks (GANs) [2], [3] or diffusion models [4], [5] (latent-based methods). By leveraging the power of face generator networks, latent-based methods can generate high-quality face images. However, interpolation in the latent space of a face generator network is not theoretically guaranteed to smoothly interpolate the identity information in the resulting image. Potentially because of this, it is observed that latent-based methods have inferior performance compared to landmark-based methods.

Instead of using the latent space of a face generator network, what if we use could use a different representation space, in which the identity can really be smoothly interpolated, and from which it is still possible to generate face images? We could then interpolate an optimal morph representation in this space and subsequently generate more successful face morphs. As a matter of fact, the embedding space of a face recognition model, which is used for distinguishing the identity in FR systems, can be the best candidate to represent identities and compute optimal morph representations by interpolation. We can find a theoretical optimal morph embedding between face embeddings of two contributing subjects, which has the same distance with both embeddings. However, the generation of face images from a face embedding is not straightforward and requires further efforts. Nevertheless, recent works in the field of face template inversion have demonstrated the feasibility of reconstructing face images from face embeddings with significant identity preservation performance [6]–[23]. The template inversion methods can provide new opportunities for exploring the embedding space of face recognition models, and also enables one to perform desired arithmetic operations in the embedding space before going back to the image space.

In this paper, we propose a novel method to generate morph images, by calculating an optimal morph embedding and then reconstructing its corresponding face image using template inversion methods. Fig. 1 illustrates the generation of face morphs with our proposed method and the attack scenario. We use three state-of-the-art template inversion methods to generate morph images as three deep morphing attacks. However, these methods also have some errors caused by imperfections in the performance of the template inverter, which leads the generated morph images to not exactly map back to the same embedding as the input to the inverter (optimal morph embedding). To address this issue, we apply optimizations in each of the template inversion models to fine-tune the morph, such that the generated face morphs have more similar embeddings to the optimal morph embeddings. In our experiments, we evaluate the performance of generated morph images and compare with previous morph attacks on the Face Research Lab London dataset (FRLL) [24] and Face Recognition Grand Challenge (FRGC) [25] datasets. We also perform a more realistic evaluation of our method by printing and then scanning the morphs (similar to what would happen in a real world passport application scenario), and compare to previously proposed morphing attacks. Our experiments demonstrate superiority of our proposed morph generation method and also show the significant vulnerability of FR models.

In summary, our contributions are as follows:

- We propose a new method to generate morphs, by interpolating source embeddings into an optimal embedding in the face embedding space, and generating morph images using template inversion methods. To our knowledge this is the first work on generation of morph images using template inversion and interpolating in the embedding space of FR models.
- We consider for this purpose three different template inversion techniques. For each case, we also extend the process with an optimization step to further improve the resulting morph.
- We provide extensive experimental evaluation, demonstrating the superior effectiveness of our proposed method compared to state-of-the-art morph attacks. In particular, we perform a print-scan evaluation of generated morph images, which demonstrates the superiority of our method and shows the vulnerability of FR systems in practical scenarios.

The remainder of the paper is organized as follows. In Section II, we review related works about morph generation as well as face template inversion. Then, in Section III we explain in detail the concept of optimal morph embedding and describ our proposed method to generate morphs from optimal morph embeddings. In Section IV, we report our experimental results and discuss our findings.

II. Related Work

In this section, we review the related work in the literature on generation of face morphs (Section II-A) and template inversion methods to reconstruct face images from facial embeddings (Section II-B).

A. Morph Generation

The idea of morphing attacks has first been introduced in 2014 by [1], which suggests a scenario where a wanted criminal and an accomplice create a fake image (the morph) mixing their two faces. After the accomplice uses this morph to apply for a passport, both contributing sources can then share it, enabling the criminal to go unnoticed through automated border control (ABC) gates. The feasibility of this attack is demonstrated by manually generating morphs, enrolling them in a FR system and showcasing that both source identities can then successfully authenticate against the morph. The morphing approach introduced in [1] is nowadays known as **landmark-based** morphing. The general idea is to label specific face landmarks in both source images, warp the images to align those landmarks, and then average the pixels. The original process involved a lot of manual work and suffered from visible so-called *ghosting* artifacts on the non-aligned boundary of the face. Subsequent works proposed an

automated approach [26], and solved the ghosting artifacts issue by blending the morphed face back into one of the source images [27].

More recently, a new family of morphing generators has emerged, based on deep learning approaches (**deep morphs**). At a high-level, all methods rely on a similar idea:

1) Define a meaningful latent space for representing face images, such that both image-to-latent encoding and latent-to-image decoding are available.
2) Encode the source images into their latent representation.
3) Generate a latent morph by interpolating between the latent sources.
4) Decode the latent morph into a face image.

In some cases, the latent morph can be fine-tuned by optimization, aiming to maximize the identity similarity of the decoded face morph to both sources' faces. Importantly, this additional step requires the introduction of a face recognition network as a part of the morph generation pipeline, to evaluate the loss in the optimization process.

The first proposed deep morphing method was introduced in [2]. They trained a generative adversarial network (GAN) for face generation, jointly with an encoder into the latent space of the GAN. They used the resulting latent space for face representations and the synthesis network of the GAN as the decoder. A main limitation of their approach is the low resolution of the resulting images, which are among others not ICAO-compliant. Higher resolution morphs were achieved in following works by exploiting the StyleGAN model [28], a powerful high-resolution face generator. Faces are again encoded in the latent space of the GAN, specifically in the $\mathcal{W}$ space [29] or the $\mathcal{W}+$ space [30]. In the absence of an image-to-latent encoder, the encoding is performed by optimization, looking for a latent able to accurately decode into the target face. Building on this, [3] first introduced the latent morph fine-tuning process (by optimizing the input of the generator using a biometric loss to further guarantee the resulting image is an effective morph), and showed its impact on the morph's effectiveness. Further developments in GANs for faces will naturally have applications in morph generation by using similar processes. For example, StyleNAT [31] proposed a transformer-based architecture for the generator, although to the best of our knowledge no work has yet evaluated the impact of this specific new design on morph effectiveness.

More recently, approaches relying on diffusion probabilistic models (DPM) as the main generative backbones have been proposed. While DPMs are typically not associated with a structured latent space, [32] introduces a diffusion autoencoder which enables encoding of real images into a semantically structured latent space. Using this new latent space for performing the latent morph interpolation [4] and [5] independently propose a similar diffusion-based deep morphing attack.

All those proposed methods (especially when not including a latent morph fine-tuning step) strongly rely on a property of linear perceptual continuity of the latent space: when moving regularly along a segment in the latent space, the perceptual changes in resulting decoded image also look regular. There is thus an assumption that the halfway point in between two source latents will decode into a satisfying morph, with roughly equivalent perceptual similarity to both source images. However, this perceptual similarity is occurring at the level of generic image features, not specifically of the identity. In other words, one can expect the morph to "look similar *overall*" to both sources, but not necessarily to "look similar *in identity*". While those two notions can in many cases roughly align (which is proven by the success of aforementioned deep morphing works), they are still conceptually different. This thus raises the question of finding a latent space with stronger conceptual guarantees that gradually moving along a segment actually will correspond to a gradual change in the identity of the generated image. As it happens, there exists a very natural space for this: the embedding space of a pretrained face recognition network. Encoding faces in this embedding space and performing the interpolation there provides the strongest theoretical guarantee that the resulting morph will be very effective [33], [34]. One remaining crucial ingredient to use this space for morph generation in practice is a decoder from face embeddings to face images, also known as a **template inversion** system.

B. Template Inversion

There are several methods in the literature for reconstructing face images from facial embeddings (also known as facial templates), which are mainly proposed for template inversion (TI) attacks against face recognition systems [6]–[23]. These methods can be categorized into *optimization-based* and *learning-based* methods. In the *optimization-based* methods, a separate optimization should be solved for generating a face image from each embedding, while in *learning-based* methods a neural network is trained which is later used for face reconstruction in the inference stage. Therefore, *learning-based* methods often have less execution time in the inference stage. Template inversion methods can also be categorized into *white-box* and *black-box* methods, based on the amount of knowledge available from the face feature extractor model of FR systems. In *white-box* methods, such as [6], [12], all the parameters and internal functioning of the face feature extractor model are known, and therefore the feature extractor model can be used in gradient-based optimization to reconstruct face images or for training a face reconstruction network. On the other hand, in *black-box* methods, such as [8], [9], [19], the internal functioning of the face feature extractor model is unknown. Consequently, the feature extractor model cannot be used in the training process of the face reconstruction network; however, it can be applied in non-gradient-based optimization approaches. Since in the *white-box* methods more knowledge of the feature extractor model is available, it is expected (and also shown in [7], [21]) that *white-box* methods achieve better reconstruction performance than *black-box* methods. While majority of methods are proposed for only *white-box* or *black-box* scenarios, there are few methods that can be used for both *white-box* and *black-box* template inversion [7], [9], [21].

We can also categorize template inversion methods by their output (i.e., generated face images), based on the resolution and quality of reconstructed face images. Specially, methods

TABLE I: Template Inversion methods in the literature.

Ref.	Method Basis	Reconstruction Quality	Reconstruction Resolution	White-box/ Black-box	Available source code
[6]	optim./learning	low	low	white-box	✗
[7]	learning	low	low	both	✗
[8]	learning	low	low	black-box	✓
[9]	learning	low	low	both	✗
[12]	learning	low	low	white-box	✓
[13]	learning	low	low	black-box	✓
[14]	learning + optim.	high	low	black-box	✗
[16]	learning	low	low	black-box	✗
[18]	learning	high	high	black-box	✓
[19]	optimization	high	high	black-box	✓
[20]	optimization	high	high	black-box	✗
[21]	learning	high	high	both	✓
[23]	learning	high	high	whitebox	✓

that are based on convolutional neural networks, such as [8], [12], often generate low-quality images which have blurriness or other artifacts. For instance, the method in [12] has the state-of-the-art reconstruction performance in terms of identity preserving, but the reconstructed face images manifest blurry and unrealistic images. In contrast, most GAN-based methods yield high-quality and realistic (i.e., *human-face-like*) images. Specially, some methods used StyleGAN to reconstruct face images from facial templates, which generate *high-resolution* and *realistic* face images. For instance, in [19], [20] authors solved an optimization on the input space of StyleGAN to find the latent code that can generate a face image which has a similar embedding. In [21], authors trained a network to map face embeddings to the intermediate latent space of StyleGAN, and then used the remaining network of StyleGAN to generate face image. The template inversion methods based on StyleGAN leverage the power of *high-resolution* face generation of StyleGAN, while other methods in the literature generate *low-resolution* face images. Recently, in [23], CLIP [35] and Stable Diffusion [36] models were fine-tuned on 42 million face images from WebFace260M dataset [37] to generate face images from embeddings of a face recognition model. Table I summarizes the template inversion methods proposed in the literature.

III. Methodology

We introduce a novel method for generating deep morphing attacks, which is grounded in the concept of optimal morph embedding. We first describe the threat model for the morph generation. We then present our definition of optimal morph embedding and the concept of generating face morphs from optimal morph embeddings in Section III-B. Finally, in Section III-C we describe our method to generate face morphs from optimal morph embeddings using template inversion methods.

A. Threat Model

We consider a morph attack against a face recognition system based on the following threat model (illustrated in Fig. 1):

- *Adversary's goal*: The adversary aims to create a face morph image $\boldsymbol{I}_{\text{morph}}$, mixing the identities from two source images $\boldsymbol{I}_1$, $\boldsymbol{I}_2$, which are for two different subjects; then enroll $\boldsymbol{I}_{\text{morph}}$ into a face recognition database (e.g., passport creation). Afterhand, the goal is for both contributing subjects to successfully authenticate against the stored reference (e.g., enabling them to share the passport to go through an automated border control gate).
- *Adversary's knowledge:* The adversary is assumed to have the following information:
  - The adversary has access to a face recognition network F_{adv} and a template inversion network G_{TI}, which is able to invert face embeddings extracted by F_{adv}.
  - The adversary may also have a *white-box* knowledge of the target face recognition system F_{target} (i.e., white-box scenario), and can use it in morph generation process (i.e., $F_{\text{adv}} = F_{\text{target}}$). Otherwise, if the adversary does not have a *white-box* knowledge of the target face recognition system F_{target} (i.e., black-box scenario), the target face recognition system F_{target} cannot be used by the adversary, and therefore the adversary uses an off-the-shelf face recognition model as F_{adv} (i.e., $F_{\text{adv}} \neq F_{\text{target}}$).
- *Adversary's capability:* The adversary can submit the generated morph for enrollment into a target face recognition system F_{target}. We consider two scenarios for the enrolment process:
  1) The adversary can submit $\boldsymbol{I}_{\text{morph}}$ as a digital image for enrollment.
  2) The adversary needs to print the image $\boldsymbol{I}_{\text{morph}}$, which will be then scanned for enrollment (print-scan).
- *Adversary's strategy:* The adversary's strategy is to compute face embeddings from both contributing subjects using F_{adv} and average them to obtain an optimal morph embedding, then invert this embedding back into $\boldsymbol{I}_{\text{morph}}$ using the template inverter G_{TI}.

B. Optimal Morph Embedding and Optimal Face Morph

Let us consider $\boldsymbol{I}_1$ and $\boldsymbol{I}_2$ as two source images whose morph we want to generate. Also, let $F_{\text{adv}}(.)$ denote a feature extractor used by the adversary, which extracts n-dimension face embeddings $\boldsymbol{x} = F_{\text{adv}}(\boldsymbol{I}) \in \mathcal{X} \subset \mathbb{R}^n$ from given face image $\boldsymbol{I}$. Given a distance function $d(.,.)$, the optimal morph embedding $\boldsymbol{x}^*$ can be defined as the face embedding whose distance to the embeddings of both source images is minimized. In other words, the optimal morph embedding can be defined as:

$$\boldsymbol{x}^* = \underset{\boldsymbol{x}\in\mathcal{X}}{\arg\min} \left[d(\boldsymbol{x}_1, \boldsymbol{x}) + d(\boldsymbol{x}_2, \boldsymbol{x})\right], \tag{1}$$

where $\boldsymbol{x}_1 = F_{\text{adv}}(\boldsymbol{I}_1)$ and $\boldsymbol{x}_2 = F_{\text{adv}}(\boldsymbol{I}_2)$ are face embeddings of source images $\boldsymbol{I}_1$ and $\boldsymbol{I}_2$, respectively. Without loss of generality, we can assume that facial embeddings are normalised[4], therefore the space of facial embeddings $\mathcal{X} \subset \mathbb{R}^n$ covers a unit hypersphere or n-ball, i.e., $||\boldsymbol{x}|| = 1, \forall \boldsymbol{x} \in \mathcal{X}$. Now, if we consider the cosine distance as our distance

[4]If embedding $\boldsymbol{x}$ extracted by $F_{\text{adv}}(.)$ is not normalised, we normalise it so that $||\boldsymbol{x}|| = 1$.

metric $d(\cdot,\cdot)$ for the normalized face embedddings, the optimal morph embedding has the same distance with both $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$. Therefore, based on the following lemma, for $n > 2$, there can be an infinite number of answers for optimal morph embeddings.

Lemma 1. *Given two points $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$ on the unit n-sphere $S^{(n-1)}$ (i.e., the boundary of an n-ball) in $\mathbb{R}^n$ with $n > 2$, there exists an infinite number of points $\boldsymbol{x}^*$ on $S^{(n-1)}$ that have the same cosine distance to both $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$.*

Proof. The set of points $\boldsymbol{x}^* \in S^{(n-1)}$ lies on the intersection of n-sphere $S^{(n-1)}$ and the hyperplane in $\mathbb{R}^n$ that $\boldsymbol{x} \cdot (\boldsymbol{x}_1 - \boldsymbol{x}_2) = 0$ that has a same cosine distance to both $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$. The intersection of the n-sphere $S^{(n-1)}$ and the hyperplane forms an $(n-2)$-dimension sphere. For $n > 2$, an $(n-2)$-dimension sphere contains an infinite number of points, and therefore there are infinite points on the n-sphere $S^{(n-1)}$ that have the same cosine distance to both $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$. □

Corollary 1. *A particular answer for the optimal morph embedding is the normalised average embedding:*

$$\boldsymbol{x}_{avg} = \frac{\boldsymbol{x}_1 + \boldsymbol{x}_2}{||\boldsymbol{x}_1 + \boldsymbol{x}_2||}. \tag{2}$$

The normalised average embedding has the same cosine distance to both $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$ and also is on the unit n-sphere. It is also very easy to calculate, and therefore for simplicity, we consider the normalised average embedding $\boldsymbol{x}_{\text{avg}}$ as an optimal morph embedding in our experiments (i.e., $\boldsymbol{x}^* = \boldsymbol{x}_{\text{avg}}$).

An ideal morphing algorithm would generate face morphs whose embeddings match an optimal morph embeddings $\boldsymbol{x}^*$. Nevertheless, while the optimal embedding can theoretically be computed, it has been considered as only a theoretical construct [33], and transforming the optimal embedding back into the image space is a priori non trivial. However, considering recent advancements in template inversion techniques, we can invert embeddings and generate a face image which has the desired embedding. Therefore, we believe that generation of almost optimal face morphs from optimal embeddings can be feasible. Hence, we leverage state-of-the-art template inversion methods to generate face morph based on optimal morph embeddings. Let $G_{\text{TI}}(.)$ denote a template inversion model which reconstruct face image $\boldsymbol{I} = G_{\text{TI}}(\boldsymbol{x})$. Then, we can use the template inversion model $G_{\text{TI}}(.)$ to generate an approximation of the optimal face morph $\boldsymbol{I}_{\text{morph}}$ from the optimal morph embedding $\boldsymbol{x}^*$:

$$\boldsymbol{I}_{\text{morph}} = G_{\text{TI}}(\boldsymbol{x}^*) \tag{3}$$

Our hypothesis is that the resulting images are strong candidates for highly effective morphing attacks. To generate face morphs from optimal morph embeddings, we build upon our three different template inversion methods introduced in [12], [21], [23] as described in Section III-C.

C. Generation of Face Morphs using Face Template Inversion

To generate the face morphs from the optimal morph embeddings, we employ state-of-the-art white-box template inversion methods proposed in [12] (for low-resolution morph generation), [21] (for high-resolution GAN-based morph generation), and [23] (for high-resolution diffusion-based morph generation). The adoption of a white-box template inversion particularly has advantages in our problem of morph generation, because we initially have two face images and extract their embeddings with a known feature extractor model F_{adv}. Therefore, it is reasonable to consider a *white-box* template inversion method and utilize a feature extractor that the adversary has *white-box* knowledge of its model.

The low-resolution template inversion method, referred to as *base-inversion* in the rest of the paper, consists of a self-contained decoder that maps from the face embedding space back to the image space. While it is expected to be highly accurate, it may generates images of limited quality and resolution as elaborated in Section IV-B. The high-resolution template inversion method based on StyleGAN, referred to as *GAN-inversion* in the rest of the paper, learns a mapping from the face embedding space into the intermediate latent space of a pretrained StyleGAN model. This approach allows us to leverage the high resolution and realism of StyleGAN generated images, even though it might come at the cost of a less accurate inversion. The high-resolution template inversion method based on diffusion model, referred to as *diffusion-inversion* in the rest of the paper, projects FR embeddings to the latent space of a pretrained CLIP model and generates images with Stable Diffusion model, which have high image quality. Therefore, each of these approaches can have their merits depending on whether the primarily goal is to fool a FR system or a human operator.

To train the template inversion models, as a preprocessing step, we first normalize the facial embeddings to have them lie on the embeddings hypersphere (as explained in Section III-B), and then train our template inversion networks. After training our template inversion models, we can generate morph image $\boldsymbol{I}^*_{\text{morph}} = G_{\text{TI}}(\boldsymbol{x}^*)$ as described in Eq. 3. For each of our template inversion methods, we also propose different optimisation-based approaches to improve the performance of template inversion technique.

1) Low-resolution Template Inversion (Base-inversion): To train the low-resolution template inversion method based on [12], we use a convolutional neural network with skip connections on the convolution blocks as the network structure. After the template inversion model is trained, it can be used to reconstruct face image from face embeddings. However, similar to any other neural network, the trained template inversion model suffers from some errors in the output (generated face images), in the way that the generated face image does not have exact same embeddings as the input face embeddings. To reduce such errors, we consider the template inversion model as a face generator network and optimize the input embedding so that the reconstructed face image has embeddings closer to the optimal morph embeddings. To this end, we use an iterative gradient descent optimization based on Algorithm 1 to solve the following optimization on the embeddings space $\mathcal{X}$ to find new embedding $\boldsymbol{x}^*_{\text{opt}}$ that can generate the face image $\boldsymbol{I}_{\text{morph, opt}} = G_{\text{TI}}(\boldsymbol{x}^*_{\text{opt}})$ which has embedding closer to optimal morph embedding $\boldsymbol{x}^*$:

Algorithm 1 Optimization on the input (embedding) of template inversion network

Inputs:
  $\boldsymbol{x}^*$: target optimal morph embedding
  $G_{\text{TI}}(.)$: template inversion network
  $F_{\text{adv}}(.)$: face recognition network (used for morph generation)
  λ : learning rate
  n_{itr} : number of iterations
Output:
  $\boldsymbol{I}_{\text{morph}}$: generated face image (approximation of optimal morph)
Procedure:
  $\boldsymbol{x} \leftarrow \boldsymbol{x}^*$
  For $n = 1, .., n_{\text{itr}}$ **do**
    cost $\leftarrow ||\boldsymbol{x}^* - [F_{\text{adv}} \circ G_{\text{TI}}](\boldsymbol{x})||_2$
    $\boldsymbol{x} \leftarrow \boldsymbol{x} - \text{Adam}(\nabla\text{cost}, \lambda)$
  End For
  $\boldsymbol{I}_{\text{morph}} \leftarrow G_{\text{TI}}(\boldsymbol{x})$
End Procedure

Algorithm 2 Optimization on the $\mathcal{W}$ / $\mathcal{W}^+$ space of StyleGAN in morph generation using template inversion network based on StyleGAN (GAN-inversion)

Inputs:
  $\boldsymbol{x}^*$: target optimal morph embedding
  $M_{\text{TI}}(.)$: mapping network of template inversion method
  $S_{\text{StyleGAN}}(.)$: synthetic network of StyleGAN
  $F_{\text{adv}}(.)$: face recognition network (used for morph generation)
  $\mathcal{S}$: original ($\mathcal{W}$) or extended ($\mathcal{W}^+$) intermediate latent space of StyleGAN
  λ : learning rate
  n_{itr} : number of iterations
Output:
  $\boldsymbol{I}_{\text{morph}}$: generated face image (approximation of optimal morph)
Procedure:
  $\boldsymbol{w} \leftarrow M_{\text{TI}}(\boldsymbol{x}^*)$
  For $n = 1, .., n_{\text{itr}}$ **do**
    cost $\leftarrow ||\boldsymbol{x}^* - [F_{\text{adv}} \circ S_{\text{StyleGAN}}](\boldsymbol{w})||_2$
    $\boldsymbol{w} \leftarrow \boldsymbol{w} - \text{Adam}(\nabla\text{cost}, \lambda), \quad \boldsymbol{w} \in \mathcal{S}$
  End For
  $\boldsymbol{I}_{\text{morph}} \leftarrow S_{\text{StyleGAN}}(\boldsymbol{w})$
End Procedure

$$\boldsymbol{x}^*_{\text{opt}} = \arg\min_{\boldsymbol{x}} ||\boldsymbol{x}^* - [F_{\text{adv}} \circ G_{\text{TI}}](\boldsymbol{x})||_2 \tag{4}$$

We use the Adam [38] optimizer for 100 iterations with the learning rate of 2.5×10^{-3} to solve this optimisation and find a better approximation of the optimal face morph.

2) High-resolution GAN-based Template Inversion (GAN-inversion): To train the high-resolution template inversion method based on [21], we use the same GAN training proposed in the original work. The method in [21], used StyleGAN3 [39], as a pre-trained face generation network, and employed the Wasserstein GAN (WGAN) [40] algorithm to learn a mapping $M_{\text{TI}}(.)$ from face embeddings to the intermediate latent space $\mathcal{W}$ of StyleGAN3. Then, the generate intermediate latent code $\boldsymbol{w} = M_{\text{TI}}(x) \in \mathcal{W} \subset \mathbb{R}^{16\times512}$ is fed to the synthesis network of StyleGAN $S_{\text{StyleGAN}}(.)$ to generate the reconstructed face image $\hat{\boldsymbol{I}} = G_{\text{TI}}(\boldsymbol{x}) = S_{\text{StyleGAN}}(\boldsymbol{w})$ using the synthesis network of StyleGAN3.

After the template inversion model is trained, it can be used to generate face morphs from optimal morph embedding $\boldsymbol{x}^*$ as follows:

$$\boldsymbol{I}_{\text{morph}} = G_{\text{TI}}(\boldsymbol{x}) = [S_{\text{StyleGAN}} \circ M_{\text{TI}}](\boldsymbol{x}^*) \tag{5}$$

However, similar to low-resolution template inversion, the trained model may have some errors in preserving the embedding in the generated face images. Therefore, we can optimize input to the template inversion network using the same optimization in Eq. 4 and Algorithm 1. Alternatively, we can optimize the generated intermediate latent code $\boldsymbol{w} = M(\boldsymbol{x})$ before feeding to the StyleGAN synthesis network S_{StyleGAN} and use the optimised intermediate latent code $\boldsymbol{w}_{\text{opt}}$ in the intermediate latent space $\mathcal{W}$ to generate face image $\boldsymbol{I}_{\text{morph, opt}} = S_{\text{StyleGAN}}(\boldsymbol{w}_{\text{opt}})$ which has embedding closer to optimal morph embedding $\boldsymbol{x}^*$ using the following equation and as presented in Algorithm 2:

$$\boldsymbol{w}_{\text{opt}} = \arg\min_{\boldsymbol{w}} ||\boldsymbol{x}^* - [F_{\text{adv}} \circ S_{\text{StyleGAN}}](\boldsymbol{w})||_2, \quad \boldsymbol{w} \in \mathcal{W} \tag{6}$$

To ensure that $\boldsymbol{w}_{\text{opt}}$ remains in the original intermediate latent space $\mathcal{W}$ of StyleGAN, we optimise one dimension of $\boldsymbol{w}$ latent code and repeat the optimised vector to have final vector in $\mathcal{W} \subset \mathbb{R}^{16\times512}$.

Alternatively, we can solve the same optimization as in Eq. 6, but in the extended intermediate latent $\mathcal{W}^+$ of StyleGAN. To this end, instead of optimizing one dimension and repeating it, we can optimize all the values in the latent code $\boldsymbol{w}$ independently. Therefore, we can find the optimised intermediate latent code $\boldsymbol{w}_{\text{opt}}$ using the following equation and as presented in Algorithm 2:

$$\boldsymbol{w}_{\text{opt}} = \arg\min_{\boldsymbol{w}} ||\boldsymbol{x}^* - [F_{\text{adv}} \circ S_{\text{StyleGAN}}](\boldsymbol{w})||_2, \quad \boldsymbol{w} \in \mathcal{W}^+ \tag{7}$$

To solve the optimization in Algorithm 2 for both Eq. 6 and Eq. 6, we similarly use the Adam [38] optimizer for 100 iterations with the learning rate of 2.5×10^{-3} to solve this optimisation and find a better approximation of the optimal face morph.

3) High-resolution Diffusion-based Template Inversion (diffusion-inversion): To generate high-resolution morph images using diffusion models, we use the pretrained model of identity-conditioned face generator proposed in [23]. The model is based on CLIP [35] and Stable Diffusion [36] models conditioned on the identity features extracted by a FR model (referred to as Insightface FR in the rest of the paper) and fine-tuned on 42 million face images from WebFace260M dataset [37]. Given a normalized face embedding $\boldsymbol{x} \in \mathcal{X}_{\text{Insightface}}$ and a random noise $\boldsymbol{n} \in \mathcal{N}$, the model can generate face image $\boldsymbol{I} = D(\boldsymbol{x}, \boldsymbol{n}, t)$ with similar FR embeddings, where t is the number of denoising iterations for the diffusion model. Therefore, to generate a morph image, we need to calculate the optimal morph embedding based on embeddings extracted from two source images using the Insightface FR as F_{adv}, and then generate a morph image with the diffusion model.

The generated images with this method are realistic and have high resolution. However, similar to previous methods,

the face embedding of the generated face image $\boldsymbol{I}$ has some differences with the initial embedding $\boldsymbol{x}$. While in previous methods (base-inversion and GAN-inversion), we propose to perform iterative optimizations to improve the inversion and face generation process, it is computationally difficult to perform similar optimization for the diffusion-based morph generation. Because, the generation of morph images from optimal morph embedding using the diffusion model, the generative model itself is based on an iterative denoising process and requires the use of the diffusion model multiple time-steps to denoise and generate each single image. Therefore, calculating gradients over different time-steps requires more computation, and also iterative optimization on input noise (such as Algorithm 1) will require much more runtime and resources. To mitigate this issue and to improve the quality of morph generation using diffusion model, we apply a greedy optimization and use k different noise vectors to generate morph image $\boldsymbol{I} = D(\boldsymbol{x}^*, \boldsymbol{n}, t)$ with each noise $\boldsymbol{n}$, and then select the generated image whose embedding $\boldsymbol{x} = F_{\text{adv}}(\boldsymbol{I})$ is the most similar to the optimal morph embedding $\boldsymbol{x}^*$. In our experiments, we generate each image with $t = 25$ iterations and generate $k = 10$ images for each given optimal morph embedding.

IV. Experiments

We aim to evaluate the effectiveness of our proposed morphing attacks. The effectiveness of a morphing attack depends on its ability to fool both a face recognition system (e.g., at an automated border control gate), and a human operator (e.g., the administrative employee receiving and processing the image submitted for a passport application). We compare the performance of our attack across those two aspects against previous attacks proposed in the literature. Code to reproduce our experiments is publicly released[5].

The ability of the attack to fool FR systems is evaluated through the mean of a vulnerability analysis, which simulates attacks by enrolling the morphs into a biometric verification system, then evaluating what percentage of them allow both contributing subjects to successfully authenticate. While in some cases, passport application photos can be submitted in digital format, they might sometimes have to rather be printed and physically sent to the processing office, where they will be scanned for redigitalization. Independent vulnerability analyses should thus be performed using the morphs either in their digital form, or after performing this print-scan operation for a more real-world setting. These analyses are performed in Section IV-A. We first compare several variants of our proposed attack to evaluate the importance of the input optimization step for the effectiveness of the attack. We then compare our proposed attack to previously proposed ones, both considering deep morphing and landmark-based morphing. Finally, we select a subset of attacks which are subjected to a print-scan process, and evaluate whether our conclusions change in this more realistic setting.

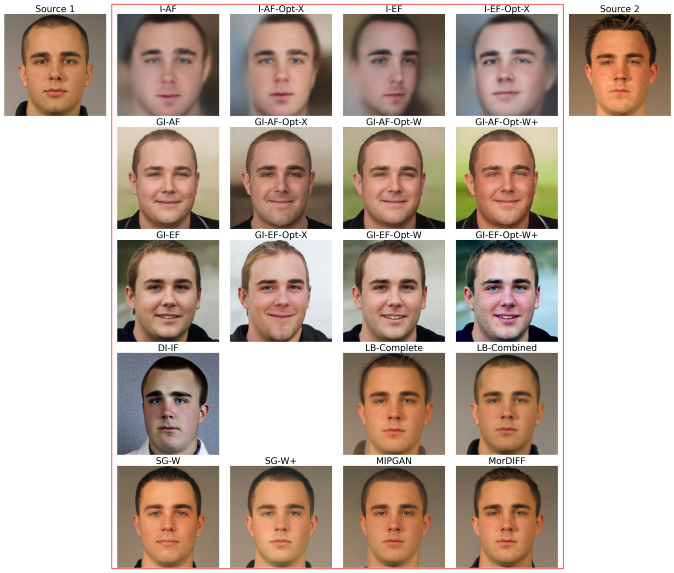
The ability of the attack to fool humans has two components. First, the morphed image should be sufficiently realistic looking to not raise suspicion. Secondly, it might also be necessary that the morph is good enough of a lookalike to both source identities (or at least of the accomplice's identity). While this is in principle already evaluated in the FR vulnerability analysis, there might be cases where human perception of identity differs from that of the automated system. Those aspects are discussed in Section IV-B, first with a qualitative discussion of the morphed faces, then with an quantitative evaluation of the morphs realism using the Fréchet Inception Distance (FID) metric.

For comparison with previous research, we include in particular in the analysis three deep morphing attacks methods, which all rely on exploiting the latent space of a pretrained StyleGAN model. The SG-W [29] and SG-W+ [30] methods do so by projecting the source images in the $\mathcal{W}$ (resp. $\mathcal{W}+$) latent space, then interpolating between the resulting latent and regenerating an image from the interpolated latent. The MIPGAN method [3] similarly finds a good latent for the morph, but does so by an optimization process including a biometric loss, thus providing better guarantee of the effectiveness of the attack. More recently proposed, the MorDIFF method [4] replaces the GAN-generator with a diffusion-based one, specifically a diffusion autoencoder [32]. Like for SG-W and SG-W+, the morphs are obtained by encoding the source images, this time in the latent space of the diffusion autoencoder, then interpolating between the encoded latents and decoding into an image. We also compare our methods to two landmarks-based (LB) methods, mainly *complete morphing* [26] (LB-Complete) which creates morphs by aligning face landmarks of the two sources through face warping then averaging the pixels, and *combined morphing* [27] (LB-Combined) which additionally blends back the morphed face into one of the source image, to remove unwanted ghosting artifacts. Those landmark-based morphing approaches can be considered state of the art in morphing attacks, in terms of effectiveness (and in particular have been evaluated to be more effective than deep morphing approaches). All morphing attacks considered in this work are summarized in Table II. We present in Fig. 2 examples of the resulting morphs using each of the considered methods.

A. Effectiveness on face recognition systems

We evaluate the ability of our proposed morphing attacks to fool face recognition systems through a **vulnerability analysis** study, whose point is to simulate morphing attacks on a FR system and evaluate the rate of successful attacks. To this end, first, a set of morphs are created from a list of pairs from a source dataset. Those morphs are enrolled as reference in a FR system, simulating a passport application. A specific operating threshold for the FR system is calibrated on a bonafide evaluation protocol, with a tolerance for a false match rate (FMR) of 10^{-3}, following the FRONTEX guideline [41]. For each morphs, probes from both contributing subjects are then presented to the system, trying to authenticate under the same registered identity represented by the morph. The morph is considered successful if both subjects manage to authenticate. This is usually evaluated through the Mated

[5]https://gitlab.idiap.ch/bob/bob.paper.tifs2025_inversion_morphing



(a) First pair of sources

(b) Second pair of sources

Fig. 2: All types of considered deep morphs for two different pairs of source identities.

Morph Presentation Match Rate, described in [42], for which we consider both the MinMax and ProdAvg variants. Those two variants differ in their handling of the case where several probes of each contributing subjects are presented to the system, with the possibility than only a portion of those probes are successfully matched to the morph reference.

In practice, we create attacks using two source datasets: the Face Research Lab London (FRLL) [24] dataset and the Face Recognition Grand Challenge (FRGC) [25] dataset. For FRLL, we select the same morphing pairs as in AMSL [27], an other morph dataset. For the vulnerability evaluation, we probe the system with all available frontal poses of the contributing subjects. When working with FRGC, we reuse both the morphing pairs and the probes from [3].

For the LB-Combined method which can generate two morphs per pair (depending on which source image the morphed face is blended back in), we arbitrarily select only one of those, to have a number of morphs comparable with other methods. Overall, this leads to a total of 1140 morphs for FRLL (per morphing method), and 2521 for FRGC. We make use of two open-source FR systems for the analysis: ArcFace (AF) [43], and ElasticFace (EF) [44]. Given inversion morphs necessitate a FR system for the generation itself, we generate independent morph sets using each of those models. We then use the same two models for simulating the considered attacks and evaluating their effectiveness. This approach in particular enables comparison between white-box attack scenarios (FR system available at morph generation time, i.e., when the same network is used for generation and evaluation), and black-box attack scenarios (when the attacked network differs from the one used for generation). On that aspect we also note that MIPGAN makes uses of the ArcFace network for computing the biometric component of the loss when fine tuning the latent morph; for this reason, we classify a MIPGAN attack on the ArcFace system as *white-box*. For calibrating the operating threshold we use the FRGC Experiment 2 protocol and operate at a tolerated FMR of 10^{-3}.

Finally, we strengthen our analysis by also including a closed-source commercial of-the-shelf (COTS) FR system against which we evaluate morphing attack effectiveness in the most realistic attack scenario, i.e., with print-scanned attacks. For this system, we use the operating threshold provided in the official documentation, again with a tolerated FMR of 10^{-3}.

1) Importance of fine-tuning the input to the inverter: Using this means of analysis, we first evaluate the importance of fine-tuning the latent morph representation on the attack effectiveness. As explained in Section III-C, this is done by optimizing the input to the inverter with the aim of better ensuring the generated morph maps back to the optimal morph embedding. Moreover, for GAN-inversion, this optimization

TABLE II: Overview of the considered morphing methods. The first block contains all considered inversion attacks, which are all our contribution. The second block lists previously proposed methods which we include in our work for comparative purposes.

Name	Approach
I-AF/EF	Base inversion of optimal template (FR-dependent)
+ Opt-X	+ optimization of the inverter's input
GI-AF/EF	GAN-inversion ($\mathcal{W}$) of optimal template (FR-dependent)
+ Opt-X/W/W+	+ optimization of the inverter's input (X) or the synthesis network input (W, W+)
DI-IF	Diffusion-inversion of optimal template
[29]SG-W	Encoding and interpolation in StyleGAN2 $\mathcal{W}$ space
[30]SG-W+	Encoding and interpolation in StyleGAN2 $\mathcal{W}+$ space
[3] MIPGAN	Optimization in StyleGAN2 $\mathcal{W}+$ space
[4] MorDIFF	Encoding and interpolation in the latent space of a Diffusion Autoencoder
[26]LB-Complete	Landmarks-based complete morph
[27]LB-Combined	Landmarks-based combined morph

TABLE III: Effect of optimizing the input to the inverter on the attack effectiveness. We distinguish white-box □ and black-box ■ attacks.

FRS	Attack	MinMax-MMPMR (%) FRLL	MinMax-MMPMR (%) FRGC	ProdAvg-MMPMR (%) FRLL	ProdAvg-MMPMR (%) FRGC
AF	□ I-AF	97.54	89.88	94.47	74.76
	□ I-AF-Opt-X	**100.00**	**99.80**	**99.91**	**98.07**
	□ GI-AF	52.02	41.81	42.46	21.95
	□ GI-AF-Opt-X	91.75	80.72	86.54	64.13
	□ GI-AF-Opt-W	88.77	79.06	83.09	61.57
	□ GI-AF-Opt-W+	**99.91**	**98.53**	**99.04**	**92.92**
	■ I-EF	90.70	79.65	85.57	61.46
	■ I-EF-Opt-X	**98.86**	**94.01**	**97.98**	**85.57**
	■ GI-EF	17.19	12.77	11.93	4.84
	■ GI-EF-Opt-X	62.46	47.32	53.29	30.81
	■ GI-EF-Opt-W	56.32	44.39	48.46	27.47
	■ GI-EF-Opt-W+	**95.35**	**83.74**	**92.57**	**71.44**
EF	□ I-EF	96.67	87.78	93.00	74.58
	□ I-EF-Opt-X	**100.00**	**99.52**	**99.87**	**95.72**
	□ GI-EF	33.68	27.45	26.47	13.69
	□ GI-EF-Opt-X	80.35	61.13	73.49	45.88
	□ GI-EF-Opt-W	75.79	61.96	69.10	45.38
	□ GI-EF-Opt-W+	**99.21**	**93.57**	**97.94**	**85.33**
	■ I-AF	91.75	75.09	86.80	58.25
	■ I-AF-Opt-X	**97.19**	**83.22**	**95.18**	**71.40**
	■ GI-AF	39.74	27.37	31.80	13.93
	■ GI-AF-Opt-X	68.51	47.56	63.57	34.68
	■ GI-AF-Opt-W	71.49	52.76	65.99	36.00
	■ GI-AF-Opt-W+	**92.46**	**77.07**	**90.13**	**62.48**

can actually be performed in intermediate spaces, either $\mathcal{W}$ or $\mathcal{W}+$, given that the optimal morph embedding is first mapped to the $\mathcal{W}$ space before being synthesized as an image. The results of the vulnerability evaluation across those different approaches are presented in Table III. We observe that in all cases the optimization process increases the effectiveness of the attack. For base inversion methods, while the effectiveness is already high without the optimization, including it further pushes the MMPMR close to the maximum. For GAN-Inversion, the effect is even more drastic, with around 3 to 6 times more attacks being successful with the optimization. We also observe in this second case that performing the optimization in $\mathcal{W}+$ achieves the best result by far.

One could also be worried that including the optimization step would start overfitting the attack to the used FR system, especially considering the fact it is already used to train the inverter. But, we observe that this is not the case: even in black-box attack scenarios, the optimization step leads to much higher attack effectiveness. This suggests that the image modifications brought by the optimization are actually able to generalize to other FR systems.

2) Comparison with other methods: We pick the best performing configurations of inversion morphs (with and without optimization), and compare their effectiveness to that of previous methods from the literature. The results are presented in Table IV. We are in particular interested in the performance of our methods compared against MIPGAN and MorDIFF, (which are considered state of the art prior to this work for deep morphing using respectively GANs and Diffusion models), and against the landmarks-based methods which have been shown in the literature to always perform better than deep morphing methods.

First focusing on the base inversion morphs, we observe that despite their limited visual realism, they perform extremely well against FR systems. In the white-box scenario, the ArcFace-based inversion with or even without fine-tuning beats MIPGAN. In the black-box scenario, base inversion without fine-tuning is already competitive with MIPGAN, and gets meaningfully stronger after fine-tuning, which makes it a leading method for deep morphing effectiveness. Moreover, base inversion actually reaches a performance somewhere inbetween LB-Complete and LB-Combined morphs. To the best of our knowledge, this is the first deep-learning based morphing method which actually manages to achieve competitive effectiveness with respect to landmark-based morphing.

Focusing now on GAN-Inversion, we observe that without fine-tuning the method is only mildly effective, both in the white-box and black-box scenario, with MMPMR values lying inbetween SG-W and SG-W+ morphing. However, applying the input optimization steps (specifically optimizing in the $\mathcal{W}+$ space) brings the performance at competitive levels. In the white-box scenario, it actually provides a performance close to be competitive with base inversion, but with the advantage of a much higher realism, as showcased in Section IV-B, and in particular beating MIPGAN and MorDIFF. In the black-box scenario, it reaches slightly better performance than MIPGAN and MorDIFF, but slightly below the landmark-based methods. Finally, the diffusion-inversion also performs quite competitively, notably more effective than the base-inversion and GAN-inversion without input optimization step in the black-box scenario, and only marginally weaker than the same method *with* input optimization. We can expect than performing input optimization for the diffusion-inversion method would further push the morph effectiveness, however as mentioned earlier, this process is non-trivial due to computational requirements.

Comparing with the visual aspects of the morphs (which is discussed more in depth in Section IV-B, we notice that the attack effectiveness seems somewhat uncorrelated to the visual realism of the morphs. Given the nature of inversion morphing (using a FR system at generation time), one can wonder whether it is more akin to an adversarial attack, i.e., the high effectiveness of the morph is not linked to actual face semantic attributes in the generated image, but rather to humanly imperceptible noise patterns introduced in the image which somehow manage to fool the FR system. This phenomenon might explain as mismatch between the effectiveness of those morphs on humans versus on systems.

The fact that the attack generalizes well in black-box scenarios is a first hint against this adversarial hypothesis, showing the morphs actually contain some meaningful high-level identity information able to fool a variety of FR networks. However, we can further test this hypothesis by doing a print-scan analysis. In this context, the morphs are printed and redigitalized before being enrolled in the system. This study has two main interests: first, it is a better simulation of a real-world scenario, in which a submitted passport picture can typically be subject to such print-scan process before enrolment. Secondly, the print-scan process has potential to degrade certain features of the image, causing a decrease in its vulnerability. With our inversion morphs in particular, if

TABLE IV: Comparison of the effectiveness of the best morph generation methods. We distinguish white-box □ and black-box ■ attack scenarios.

FRS	Attack	MinMax-MMPMR (%) FRLL	MinMax-MMPMR (%) FRGC	ProdAvg-MMPMR (%) FRLL	ProdAvg-MMPMR (%) FRGC
AF	□ I-AF	97.54	89.88	94.47	74.76
	□ I-AF-Opt-X	**100.00**	**99.80**	**99.91**	**98.07**
	□ GI-AF	52.02	41.81	42.46	21.95
	□ GI-AF-Opt-W+	99.91	98.53	99.04	92.92
	□ MIPGAN	-	73.22	-	54.77
	■ I-EF	90.70	79.65	85.57	61.46
	■ I-EF-Opt-X	**98.86**	**94.01**	**97.98**	**85.57**
	■ GI-EF	17.19	12.77	11.93	4.84
	■ GI-EF-Opt-W+	95.35	83.74	92.57	71.44
	■ DI-IF	97.11	84.69	94.56	71.09
	■ SG-W	1.05	4.32	0.64	1.44
	■ SG-W+	62.63	60.10	53.71	39.97
	■ MorDIFF	90.09	74.81	84.32	58.06
	■ LB-Complete	**99.21**	**95.48**	**98.16**	**87.51**
	■ LB-Combined	93.95	84.97	91.23	70.68
EF	□ I-EF	96.67	87.78	93.00	74.58
	□ I-EF-Opt-X	**100.00**	**99.52**	**99.87**	**95.72**
	□ GI-EF	33.68	27.45	26.47	13.69
	□ GI-EF-Opt-W+	99.21	93.57	97.94	85.33
	■ I-AF	91.75	75.09	86.80	58.25
	■ I-AF-Opt-X	**97.19**	**83.22**	**95.18**	**71.40**
	■ GI-AF	39.74	27.37	31.80	13.93
	■ GI-AF-Opt-W+	92.46	77.07	90.13	62.48
	■ DI-IF	92.11	72.19	87.85	56.25
	■ SG-W	3.07	10.19	2.06	3.56
	■ SG-W+	72.28	67.63	62.81	48.90
	■ MIPGAN	-	75.80	-	60.10
	■ MorDIFF	89.74	76.76	83.42	60.81
	■ LB-Complete	**99.47**	**96.63**	**98.51**	**89.33**
	■ LB-Combined	95.96	87.58	93.33	75.54

their effectiveness was due to some fine adversarial signal, a print-scan process could likely degrade such subtle patterns: hence, it is important to check how the attack performance varies in this setting, to evaluate how robust it is to minor degradations. We specifically select the best performing base inversion and GAN-Inversion attacks from Table IV, the best performing StyleGAN-based approach (MIPGAN), the diffusion-based approach (MorDIFF) and both landmark-based approaches, and restrict the analysis to FRGC morphs only. The morphs are printed as a grid of 35mm x 35mm then rescanned at a resolution of 300 DPI, using a *Kyocera TASKalfa 2554ci* (laser printer + scanner). Fig. 3 showcases the resulting morphs after print-scan processing. The probes for the vulnerability study are kept digitalized, to simulate a live capture and comparison at an automated border control gate. This experiment is a simulation of a real-world attack, and to make it even more realistic, as part of the analysis we also include a commercial of-the-shelf (COTS) system, where we have only API access[6]. The print-scan evaluation with COTS model provides the strongest estimate of morphing attack effectiveness by attacking a completely closed-source system. The results of the vulnerability analysis are presented in Table V.

We observe that while degraded, the performance of our morphing attacks is still preserved in this print-scan setting. The I-EF-Opt-X system performance degradation is even low enough that the method actually becomes stronger than

[6]Note that the API only provides us comparison scores, and we even do not have access to embeddings of the commercial model.

TABLE V: Vulnerability analysis on a subset of attacks using the FRGC source dataset, in a print-scan setting. We distinguish between white-box □ and black-box ■ attack scenarios.

FRS	Attack	MinMax-MMPMR (%)	ProdAvg-MMPMR (%)
AF	□ MIPGAN	62.16	43.65
	□ I-AF-Opt-X	**99.09**	**95.33**
	□ GI-AF-Opt-W+	97.26	90.19
	■ MorDIFF	66.56	48.48
	■ I-EF-Opt-X	**91.51**	**81.09**
	■ GI-EF-Opt-W+	81.63	68.49
	■ DI-IF	80.56	65.67
	■ LB-Complete	88.22	75.17
	■ LB-Combined	72.03	55.42
EF	□ I-EF-Opt-X	**100.00**	**99.11**
	□ GI-EF-Opt-W+	98.73	95.46
	■ MorDIFF	90.32	77.90
	■ MIPGAN	87.70	77.14
	■ I-AF-Opt-X	95.04	87.12
	■ GI-AF-Opt-W+	93.34	83.59
	■ DI-IF	88.22	77.46
	■ LB-Complete	**97.86**	**93.89**
	■ LB-Combined	93.45	84.52
COTS	■ MorDIFF	81.59	69.62
	■ MIPGAN	73.38	60.14
	■ I-AF-Opt-X	82.82	71.92
	■ I-EF-Opt-X	81.44	71.46
	■ GI-AF-Opt-W+	74.61	61.95
	■ GI-EF-Opt-W+	70.84	59.14
	■ DI-IF	91.39	83.03
	■ LB-Complete	**94.09**	**86.91**
	■ LB-Combined	77.23	64.33

landmark based morphing in this setting. It is non-trival to understand how this phenomenon can occur, but it at least suggests that the relevant face patterns in the morph are extremely robust even to print-scan degradation. In parallel, the GAN-Inversion morphs with fine-tuning stay more effective than MIPGAN and MorDIFF ones in this setting, and the GI-EF-Opt-W+ configuration in particular actually reaches a better performance than LB-Combined morphs in the black-box scenario.

The global effectiveness of all considered morphing attacks is further demonstrated by their success rate when attacking the COTS system. We observe that LB-Complete attacks are the most effective ones in this scenario, but inversion-based methods (specifically diffusion-inversion and base-inversion) are the next best ones, notably outperforming LB-Combined and MIPGAN attacks. Compared to MorDIFF, base-inversion achieves competitive performance but the diffusion-inversion outperforms MorDIFF by a large margin. This evaluation further demonstrates that even though white-box access to a singular face recognition system is necessary to generate inversion-based morphs, the resulting attacks keep their effectiveness even on a completely unrelated, closed-source system.

B. Perceptual analysis

We now aim to discuss perceptual aspects of the generated morphs, and in particular the 3 following points: whether the morphs can be perceptually perceived as a good lookalike to

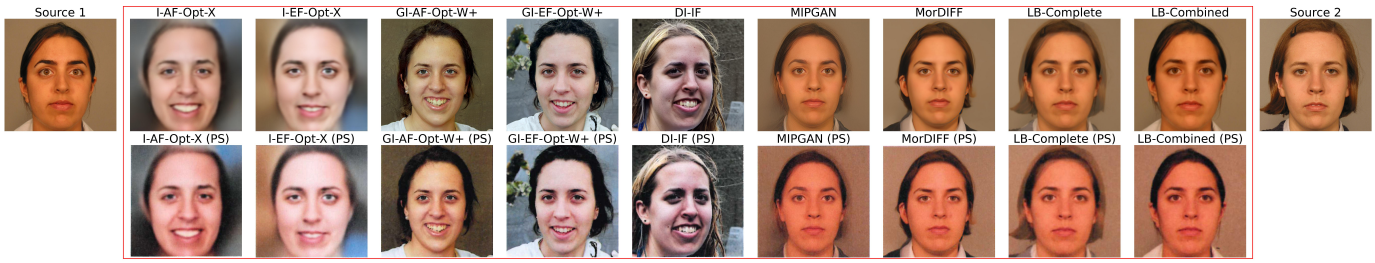


Fig. 3: Examples of the selected subset of morphing attacks before and after print-scan (PS).

both source identities, whether they are prone to be used in a real-world attack scenario, and finally whether they showcase sufficient overall realism.

On the aspect of assessing whether the morph can be considered a good lookalike to both identities, we observe that the inversion morphs behave somewhat differently than other methods. Indeed, as previous methods typically ensure a global image similarity of the morph to source identities, not only do they provide "intermediate" facial features, but also similar looking face shape, facial expression, and hair. In contrast, with inversion morphing, mainly the central portion of the face is an mixture of the facial features of both sources; however surrounding elements like the face shape or the hair, and certain covariates, like the expression or pose, can be relatively different. See for example Fig. 2a, in which GAN-inversion morphs showcase a somewhat rounder face shape than the source identities, and a smiling expression. This is not necessarily surprising, given the morph generation process. First, the inverted FR system typically only pays attention to a tight crop of the face, hence boundary features such as the exact cheeks shape or the hair should not matter. Moreover, an ideal FR system being independent from the face expression, there is no reason to invert a given embedding into specifically a neutral expression.

On the aspect of real-world application of inversion morphs in a real-world attack scenario, we can highlight two main limitations of the current methods. First, the non-uniform background in some cases (especially with GAN-inversion) might be problematic as it might not satisfy passport photo standards in many countries. This is only a small issue given some background post-processing should not be particularly difficult to achieve. Secondly, the non-neutral expression is also probably not compatible with photo passport standards. There, further work might be needed. One possibility would be to constrain the template inversion network to produce images with only neutral expression. For GAN-Inversion, another possibility would be to edit the $\mathcal{W}$ representation of the latent morph in to neutralize the expression, as proposed for example in [45]. Finally, an approach enabling for GAN-Inversion morphs to better resemble the source dataset would be to actually fine-tune the GAN generator on the source dataset, such as is done for MIPGAN. However, we hypothesize that this fine-tuning might be overall too destructive, due to catastrophic forgetting of the original weights, potentially leading to a decrease in realism of the generated image; as a matter of fact, we observe in many MIPGAN images blurry artifacts around the face which are normally absent from StyleGAN generated image, and might have been caused by a degradation of the generator's ability when fine-tuning it.

On aspects of realism, we observe that the GAN-inversion morphs are quite sharp an realistic looking, comparable to other StyleGAN-based approaches. They could even be considered to compare favorably against MIPGAN morphs, for which blurry contours are sometimes observed around the face. They also compare favorably to LB-Complete morphs given those latter ones contain ghosting artifacts. In contrast, base inversion morphs are of lower quality overall: while the center of the face itself is quite sharp, contours are relatively blurry. Some post processing, for example aiming to blend the face morph onto one of the source images, is probably necessary for those morphs to actually manage to fool a human evaluator.

An exact quantitative evaluation of the morphs' realism would require to run a perceptual study on human subjects, however we can get a first insight on the topic using the Fréchet Inception Distance metric (FID) [46] which maps human judgement relatively well and has been used in the literature for realism evaluation (e.g., [5], [28]). It estimates the perceptual distance between two sets of images by extracting their feature representations using a pretrained Inception network, fitting a Gaussian to both distributions, then computing the Fréchet distance between those. To compute the FID score, we use for each dataset the full set of real data (all source images used for morphing plus all the probes used in the vulnerability study, see Section IV-A) as the bonafide sets, and the generated morphs with each individual attack as the fake set. The results are presented in Table VI. The FID results confirm the relatively clear-cut observation that base inversion morphs are not of the greatest realism (high FID), but we actually also observe that the GAN-Inversion and diffusion-inversion morphs are not performing as well as StyleGAN or landmark-based morphs. One possible explanation is that while often used for that purpose, the FID technically does not measure exactly "realism", but rather differences between the bonafide and fake set: as we commented before, inversion morphs being less constrained, they do not preserve some image elements like background, expression or pose, which sets the morph distribution further away from the source distribution, in contrary to other morphing methods which constrain both the identities and other image factors.

TABLE VI: Fréchet Inception Distance (FID). A lower value indicate an estimated stronger perceptual realism, i.e., lower is better.

Attack	FRLL	FRGC
I-AF	270.14	264.60
I-EF	269.87	259.52
I-AF-Opt-X	271.19	265.14
I-EF-Opt-X	268.28	265.24
GI-AF	96.33	64.05
GI-EF	92.02	76.73
GI-AF-Opt-W+	87.99	62.99
GI-EF-Opt-W+	82.56	70.79
DI-IF	124.57	116.56
SG-W	25.99	21.31
SG-W+	**23.75**	**17.76**
MIPGAN	-	35.52
MorDIFF	58.79	81.65
LB-Complete	42.86	30.48
LB-Combined	**27.87**	**28.36**

TABLE VII: Detection Equal Error Rate (D-EER), in %, using publicly available morphing attack detectors. For each attack, the same set of bonafide is used and correspond to frontal samples extracted from the source dataset which is also used to create the morphs.

Attack	Model Dataset	MixFaceNet SMDD [47]	SPL-MAD [48]
I-AF-Opt-X	FRLL	0.09	0.00
	FRGC	0.92	0.08
I-EF-Opt-X	FRLL	0.49	0.00
	FRGC	0.86	0.08
GI-AF-Opt-W+	FRLL	49.57	70.63
	FRGC	53.57	67.76
GI-EF-Opt-W+	FRLL	55.46	75.53
	FRGC	67.88	78.18
DI-IF	FRLL	70.14	74.40
	FRGC	84.72	70.31
MIPGAN	FRGC	24.54	15.48
MorDIFF	FRLL	2.18	0.53
	FRGC	8.88	1.19
LB-Complete	FRLL	0.45	0.00
	FRGC	2.37	0.16
LB-Combined	FRLL	1.65	0.00
	FRGC	10.81	0.51

C. Detectability

We also evaluate the detectability of the generated morphs. This is done using publicly available morphing attack detectors and running them as is on our data. We consider MixFaceNet-SMDD [47], a detector trained on the SMDD dataset as a binary classifier , and SPL-MAD [48], which is a one class detector, i.e., it models the distribution of bonafide images and detect morphing attacks as out-of-distribution samples. Table VII presents the detection performance of those two models on each type of considered attacks (using always the same set of bonafide samples). We specifically report the Detection Equal Error Rate (D-EER), which corresponds to the error rate when at an operating threshold where both the proportion of attacks classified as bonafide and the proportion of bonafide classified as attacks are equal.

We first observed that base-inversion morphs are very effectively detected, which is likely related to their low resolution and subpar realism. However, both GAN-Inversion morphs and Diffusion-inversion morphs pose a high challenge for both detectors. This is notably despite the detectors performing decently on other GAN-based morphs (e.g., MIPGAN), and and other diffusion-based morphs (e.g., MorDIFF). We also observe that despite their high effectiveness, landmark-based morphs are typically well detected by available detectors. This is likely due to them appearing first historically, and thus being systematically considered in the development of morphing attack detectors.

As inversion-based morphs have been showcased to be simultaneously effective to fool face recognition systems, and not easily detected by existing detectors, it suggests that future improvements on the detectors might need to consider inversion morphs as an additional possible attack.

V. Conclusion

We introduced a new deep morphing method which works by approximating the optimal face morph using template inversion methods to reconstruct an image from the optimal morph embedding. We showcased in particular three distinct template inversion systems, one based on a fully trained embedding-to-image synthesis network, one other exploiting the latent space of a pretrained face GAN for the synthesis, helping to increase the realism of the morphs, and another based on probabilistic diffusion face generator model. In each case, we also proposed two generation settings, one where default output of the template inverter is used, and one where it is fine-tuned or optimised to better correspond to the optimal morph embedding, which we have shown to be extremely effective on the morphs capability to fool FR systems.

We have demonstrated both inversion and GAN-Inversion attacks are state of the art in terms of attack effectiveness against FR systems, not only beating previously introduced deep morphing approaches, but also reaching a competitive effectiveness with respect to landmark-based morphing, which is entirely novel for deep morphs, to the best of our knowledge. Moreover, we showcased that the attack generalizes well to black-box scenarios (where the FR system used for generation differs from the attacked one), and is robust to a print-scan degradation. These observations hold even when attacking a closed-source commercial off-the-shelf system. Finally, we observed that preexisting morphing attack detectors are performing poorly on the GAN and Diffusion variant of the proposed inverted morphs, which adds another aspect of concern.

We should note that inversion morphs however have some shortcomings, mainly the fact that they either showcase too low realism (for base inversion morphs), or too unconstrained faces (for GAN-Inversion morphs, the face shape, face expres-

sion, pose, and background in particular make them potentially difficult to actually use for a passport application). Some of those limitations might be relieved in the future through post-processing operations (for example, the background can be post-processed to make it uniform, or the expression of GAN-Inversion morphs can be edited by moving along a expression-neutralizing direction in the $\mathcal{W}$ latent space of the auxiliary StyleGAN). Moreover, we want to highlight the strong connection between template inversion research and morphing attack generation: each inversion system has a direct application to morphing using this same process of inverting the optimal morph embedding. Therefore, we can expect that as progress is done in template inversion, the resulting inversion morphs might become more and more realistic, constrained, and effective.

Overall, this work demonstrates such inversion-based morphs have strong potential to be used in actual real-world scenarios, and thus that research on morphing attack detection should bring particular care to this new proposed type of attack. In this paper, we considered a particular answer for optimal embedding (i.e., normalised average embedding), however as stated in Lemma 1, there can be an infinite number of answers for the optimal morph embeddings, which may lead to stronger morph attacks and require further research in future.

REFERENCES

[1] M. Ferrara, A. Franco, and D. Maltoni, "The magic passport," in *IEEE International Joint Conference on Biometrics*, Sep. 2014, pp. 1–7.

[2] N. Damer, A. M. Saladié, A. Braun, and A. Kuijper, "MorGAN: Recognition Vulnerability and Attack Detectability of Face Morphing Attacks Created by Generative Adversarial Network," in *2018 IEEE 9th International Conference on Biometrics Theory, Applications and Systems (BTAS)*, Oct. 2018, pp. 1–10.

[3] 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*, vol. 3, no. 3, pp. 365–383, Jul. 2021.

[4] 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)*. IEEE, 2023, pp. 1–6.

[5] Z. Blasingame and C. Liu, "Leveraging Diffusion For Strong and High Quality Face Morphing Attacks," *arXiv preprint arXiv:2301.04218*, 2023.

[6] A. Zhmoginov and M. Sandler, "Inverting face embeddings with convolutional neural networks," *arXiv preprint arXiv:1606.04189*, 2016.

[7] F. Cole, D. Belanger, D. Krishnan, A. Sarna, I. Mosseri, and W. T. Freeman, "Synthesizing normalized faces from facial identity features," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 3703–3712.

[8] G. Mai, K. Cao, P. C. Yuen, and A. K. Jain, "On the reconstruction of face images from deep face templates," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 41, no. 5, pp. 1188–1202, 2018.

[9] C. N. Duong, T.-D. Truong, K. Luu, K. G. Quach, H. Bui, and K. Roy, "Vec2face: Unveil human faces from their blackbox features in face recognition," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2020, pp. 6132–6141.

[10] H. O. Shahreza and S. Marcel, "Template inversion attack against face recognition systems using 3d face reconstruction," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2023, pp. 19 662–19 672.

[11] H. O. Shahreza, V. K. Hahn, and S. Marcel, "Face reconstruction from deep facial embeddings using a convolutional neural network," in *Proceedings of the IEEE International Conference on Image Processing (ICIP)*. IEEE, 2022, pp. 1211–1215.

[12] ——, "Vulnerability of state-of-the-art face recognition models to template inversion attack," *IEEE Transactions on Information Forensics and Security*, 2024.

[13] H. Otroshi Shahreza and S. Marcel, "Blackbox face reconstruction from deep facial embeddings using a different face recognition model," in *Proc. of the IEEE International Conference on Image Processing (ICIP)*. IEEE, 2023.

[14] M. Akasaka, S. Maeda, Y. Sato, M. Nishigaki, and T. Ohki, "Model-free template reconstruction attack with feature converter," in *2022 International Conference of the Biometrics Special Interest Group (BIOSIG)*. IEEE, 2022, pp. 1–5.

[15] H. O. Shahreza and S. Marcel, "Face reconstruction from partially leaked facial embeddings," in *ICASSP 2024-2024 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2024, pp. 4930–4934.

[16] S. Ahmad, K. Mahmood, and B. Fuller, "Inverting biometric models with fewer samples: Incorporating the output of multiple models," in *2022 IEEE International Joint Conference on Biometrics (IJCB)*. IEEE, 2022, pp. 1–11.

[17] H. O. Shahreza and S. Marcel, "Comprehensive vulnerability evaluation of face recognition systems to template inversion attacks via 3d face reconstruction," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2023.

[18] X. Dong, Z. Jin, Z. Guo, and A. B. J. Teoh, "Towards generating high definition face images from deep templates," in *Proceedings of the International Conference of the Biometrics Special Interest Group (BIOSIG)*. IEEE, 2021, pp. 1–11.

[19] E. Vendrow and J. Vendrow, "Realistic face reconstruction from deep embeddings," in *Proceedings of NeurIPS 2021 Workshop Privacy in Machine Learning*, 2021.

[20] X. Dong, Z. Miao, L. Ma, J. Shen, Z. Jin, Z. Guo, and A. B. J. Teoh, "Reconstruct face from features using gan generator as a distribution constraint," *arXiv preprint arXiv:2206.04295*, 2022.

[21] H. Otroshi Shahreza and S. Marcel, "Face reconstruction from facial templates by learning latent space of a generator network," *Advances in Neural Information Processing Systems*, vol. 36, pp. 12 703–12 720, 2023.

[22] H. O. Shahreza, A. George, and S. Marcel, "Face reconstruction from face embeddings using adapter to a face foundation model," in *Proceedings of the Computer Vision and Pattern Recognition Conference*, 2025, pp. 5584–5593.

[23] F. P. Papantoniou, A. Lattas, S. Moschoglou, J. Deng, B. Kainz, and S. Zafeiriou, "Arc2face: A foundation model of human faces," in *European Conference on Computer Vision*, 2024.

[24] L. DeBruine and B. Jones, "Face Research Lab London Set," May 2017.

[25] P. Phillips, P. Flynn, T. Scruggs, K. 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)*, vol. 1, Jun. 2005, pp. 947–954 vol. 1.

[26] A. Makrushin, T. Neubert, and J. Dittmann, "Automatic Generation and Detection of Visually Faultless Facial Morphs:," in *Proceedings of the 12th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications*. Porto, Portugal: SCITEPRESS - Science and Technology Publications, 2017, pp. 39–50.

[27] T. Neubert, A. Makrushin, M. Hildebrandt, C. Krätzer, and J. Dittmann, "Extended StirTrace benchmarking of biometric and forensic qualities of morphed face images," *IET Biom.*, 2018.

[28] T. Karras, S. Laine, and T. Aila, "A Style-Based Generator Architecture for Generative Adversarial Networks," in *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Jun. 2019, pp. 4396–4405.

[29] 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)*, May 2022, pp. 2959–2963.

[30] 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)*, Apr. 2020, pp. 1–6.

[31] S. Walton, A. Hassani, X. Xu, Z. Wang, and H. Shi, "Efficient image generation with variadic attention heads," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, June 2025, pp. 3264–3275.

[32] K. Preechakul, N. Chatthee, S. Wizadwongsa, and S. Suwajanakorn, "Diffusion Autoencoders: Toward a Meaningful and Decodable Repre-

sentation," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022, pp. 10 619–10 629.
[33] U. M. Kelly, L. Spreeuwers, and R. Veldhuis, "Worst-Case Morphs: A Theoretical and a Practical Approach," in *2022 International Conference of the Biometrics Special Interest Group (BIOSIG)*, Sep. 2022, pp. 1–5.
[34] L. Colbois, H. Otroshi Shahreza, and S. Marcel, "Approximating optimal morphing attacks using template inversion," in *IEEE International Joint Conference on Biometric*, Sep. 2023.
[35] A. Radford, J. W. Kim, C. Hallacy, A. Ramesh, G. Goh, S. Agarwal, G. Sastry, A. Askell, P. Mishkin, J. Clark *et al.*, "Learning transferable visual models from natural language supervision," in *International conference on machine learning*. PMLR, 2021, pp. 8748–8763.
[36] 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*, 2022, pp. 10 684–10 695.
[37] Z. Zhu, G. Huang, J. Deng, Y. Ye, J. Huang, X. Chen, J. Zhu, T. Yang, J. Lu, D. Du *et al.*, "Webface260m: A benchmark unveiling the power of million-scale deep face recognition," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2021, pp. 10 492–10 502.
[38] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," in *Proceedings of the International Conference on Learning Representations (ICLR)*, San Diego, California., USA, May 2015.
[39] T. Karras, M. Aittala, S. Laine, E. Härkönen, J. Hellsten, J. Lehtinen, and T. Aila, "Alias-free generative adversarial networks," *Advances in Neural Information Processing Systems*, vol. 34, 2021.
[40] M. Arjovsky, S. Chintala, and L. Bottou, "Wasserstein generative adversarial networks," in *International conference on machine learning*. PMLR, 2017, pp. 214–223.
[41] FRONTEX, "Best practice technical guidelines for automated border control ABC systems," 2015.
[42] U. Scherhag, A. Nautsch, C. Rathgeb, M. Gomez-Barrero, R. N. J. Veldhuis, L. Spreeuwers, M. Schils, D. Maltoni, P. Grother, S. Marcel, R. Breithaupt, R. Ramachandra, and C. Busch, "Biometric Systems under Morphing Attacks: Assessment of Morphing Techniques and Vulnerability Reporting," in *2017 International Conference of the Biometrics Special Interest Group (BIOSIG)*, Sep. 2017, pp. 1–7.
[43] 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*, 2019, pp. 4690–4699.
[44] F. Boutros, N. Damer, F. Kirchbuchner, and A. Kuijper, "ElasticFace: Elastic Margin Loss for Deep Face Recognition," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022, pp. 1578–1587.
[45] L. Colbois, T. de Freitas Pereira, and S. Marcel, "On the use of automatically generated synthetic image datasets for benchmarking face recognition," in *International Joint Conference on Biometrics (IJCB 2021)*, 2021, accepted for Publication in IJCB2021.
[46] M. Heusel, H. Ramsauer, T. Unterthiner, B. Nessler, and S. Hochreiter, "GANs Trained by a Two Time-Scale Update Rule Converge to a Local Nash Equilibrium," in *Advances in Neural Information Processing Systems*, vol. 30. Curran Associates, Inc., 2017.
[47] 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 *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, Jun. 2022, pp. 1605–1616.
[48] M. Fang, F. Boutros, and N. Damer, "Unsupervised Face Morphing Attack Detection via Self-paced Anomaly Detection," Aug. 2022.

Hatef Otroshi Shahreza received his Ph.D. from the École Polytechnique Fédérale de Lausanne (EPFL), Switzerland, in 2024, and was a Research Assistant with Idiap Research Institute, Switzerland, where he received Marie Skłodowska-Curie Fellowship for his doctoral program. Prior to his Ph.D., he received the B.Sc. degree (Hons.) in electrical engineering from the University of Kashan, Iran, in 2016, and the M.Sc. degree in electrical engineering from the Sharif University of Technology, Iran, in 2018. During his Ph.D., Hatef also spent 6 months as a Visiting Scholar at Hochschule Darmstadt, Germany. He is currently a Postdoctoral Researcher with the Biometrics Security and Privacy Group at Idiap Research Institute, Switzerland. He was the reception of the European Association for Biometrics (EAB) Research Award 2023 and the IEEE Biometrics Council Best Doctoral Dissertation Award 2025. His research interests include deep learning, computer vision, generative models, and biometrics.

Laurent Colbois graduated in 2020 from the École Polytechnique Fédérale de Lausanne (EPFL), Switzerland, obtaining his MSc. in Computational Science and Engineering, with a main focus on machine learning, signal processing and numerical simulations. He is working since July 2020 in the Biometrics Security & Privacy at the Idiap Research Institute in Martigny, Switzerland, where he specializes in deepfake detection and face recognition. He started in January 2021 a PhD jointly at Idiap and at the School of Criminal Sciences in University of Lausanne (UNIL), Switzerland, which he concluded in March 2025. He now is continuing his work at Idiap as a Postdoctoral Research.

Sébastien Marcel heads the Biometrics Security and Privacy group at Idiap Research Institute (Switzerland) and conducts research on face recognition, speaker recognition, vein recognition, attack detection (presentation attacks, morphing attacks, deepfakes) and template protection. He received his Ph.D. degree in signal processing from Université de Rennes I in France (2000) at CNET, the research center of France Telecom (now Orange Labs). He is Professor at the University of Lausanne (School of Criminal Justice) and a lecturer at the École Polytechnique Fédérale de Lausanne. He is also the Director of the Swiss Center for Biometrics Research and Testing, which conducts certifications of biometric products.